

T.C.
KARABÜK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI

VERİ MADENCİLİĞİNDE SINIFLAYICI TEKNİKLER
İLE DEMİR ÇELİK SEKTÖRÜNDE UZUN ÜRÜNLERİN
ÜRETİMİNE İLİŞKİN BİR TAHMİN MODELLEMESİ

YÜKSEK LİSANS TEZİ

Bülent ORDU

Karabük-2013

**T.C.
KARABÜK ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
İŞLETME ANABİLİM DALI**

**VERİ MADENCİLİĞİNDE SINIFLAYICI TEKNİKLER
İLE DEMİR ÇELİK SEKTÖRÜNDE UZUN ÜRÜNLERİN
ÜRETİMİNE İLİŞKİN BİR TAHMİN MODELLEMESİ**

Bülent ORDU

YÜKSEK LİSANS TEZİ

**Danışman
Doç. Dr. Filiz ERSÖZ**

Karabük-2013

YÜKSEK LİSANS TEZİ ONAY FORMU

İşletme Anabilim Dalı'nda Doç.Dr. Filiz ERSÖZ danışmanlığında, Bülent ORDU tarafından hazırlanan bu çalışma 03/01/2013 tarihinde jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir

Jüri Başkanı

Doç.Dr. Abdullah KARAKAYA

Jüri Üyesi- Danışman

Doç.Dr. Filiz ERSÖZ

F. Ersöz

Jüri Üyesi

Yrd.Doç.Dr. Mehmet PEKKAYA

M. Pekkaya

Karabük Üniversitesi Sosyal Bilimler Enstitüsü Yönetim Kurulu'nun/...../..... tarih ve/..... sayılı kararı ile onaylanmıştır.

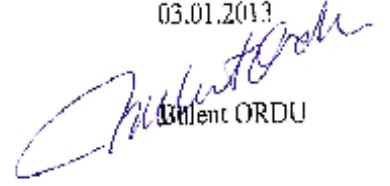
Enstitü Müdürü

Doç.Dr. Abdullah KARAKAYA

Karabük Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğine göre hazırlamış olduğum "**Veri Madenciliğinde Sınıflayıcı Teknikler İle Demir Çelik Sektöründe Uzun Ürünlerin Üretimine İlişkin Bir Tahmin Modellemesi**" adlı tezin tamamen kendi çalışmam olduğunu; hazırlanması, yürütülmesi, araştırmalarının yapılması ve bulgularının analizlerinde bilimsel etiğe ve akademik kurallara özenle riayet edildiğini; bu çalışmada kullanılan doğrudan kendime ait olmayan bulguların, verilerin ve materyallerin bilimsel etiğe uygun olarak kaynak gösterildiğini ve alıntı yapılan çalışmalara atfedildiğini beyan ederim.

Lisansüstü Eğitim-Öğretim yönetmeliğinin ilgili maddeleri uyarınca gereğinin yapılmasını arz ederim.

03.01.2013


Mehmet ÖRDÜ

İçindekiler

ÖNSÖZ	III
KISALTMALAR.....	IV
ÖZET	V
ABSTRACT.....	VI
ŞEKİLLER LİSTESİ	VII
TABLOLAR LİSTESİ.....	VIII
GİRİŞ	1
1. DEMİR - ÇELİK SEKTÖRÜ	2
1.1. DEMİR – ÇELİK TANIMI.....	2
1.2. DEMİR - ÇELİĞİN ÜRETİM AŞAMALARI	3
1.3. DEMİR - ÇELİK ÜRÜNLERİNİN SINIFLANDIRILMASI.....	6
1.4. DEMİR - ÇELİK SEKTÖRÜNÜN ÖZELLİKLERİ	11
1.4.1. Demir – Çelik Sektörünün Etkileşim Halinde Olduğu Diğer Sektörler....	12
1.4.2. Dünya’da Demir - Çelik Sektörü	12
1.4.3. Avrupa’da Demir - Çelik Sektörü	17
1.4.4. Türkiye’de Demir - Çelik Sektörü	20
2. VERİ MADENCİLİĞİ	31
2.1. VERİ MADENCİLİĞİNİN TANIMI	31
2.2. VERİ MADENCİLİĞİ GELİŞİM SÜRECİ	32
2.3. VERİ MADENCİLİĞİNİN KULLANIM ALANLARI.....	34
2.4. VERİ TABANLARINDA BİLGİ KEŞFİ SÜRECİ.....	36
2.5. VERİ MADENCİLİĞİ SÜRECİ.....	36
2.5.1. Problemin Tanımlanması	37
2.5.2. Verilerin Hazırlanması	38
2.5.2.1. Verilerin Toplanması.....	38
2.5.2.2. Verilerin Temizlenmesi	38
2.5.2.3. Verilerin Bütünleştirilmesi	38
2.5.2.4. Verilerin Dönüştürülmesi	39
2.5.2.5. Verilerin İndirgenmesi	39
2.5.3. Modelin Kurulması	39
2.5.4. Modelin Değerlendirilmesi	40

2.5.5. Modelin Uygulanması.....	40
2.6. VERİ AMBARLARI	41
2.6.1. Veri Ambarı Mimarisi.....	41
2.6.2. Veri Modelleme	42
2.7. VERİ MADENCİLİĞİNDE KULLANILAN YÖNTEMLER.....	43
2.7.1. Regresyon Analizi	44
2.7.1.1. Basit Doğrusal Regresyon Analizi	45
2.7.1.2. Çoklu Doğrusal Regresyon Analizi	47
2.7.1.3. Regresyon Varsayımları	49
2.7.1.4. Korelasyon Analizi	53
2.7.2. Karar Ağaçları	55
2.7.2.1. Karar Ağaçlarında Dallanma ve Budama Kriterleri.....	56
2.7.2.2. Bazı Karar Ağacı Algoritmaları	57
2.7.3. Yapay Sinir Ağları	61
2.7.4. Bayes Sınıflandırma Algoritması.....	62
2.7.5. Genetik Algoritmalar	63
2.8. VERİ MADENCİLİĞİ YAZILIMLARINDAN BAZILARI.....	64
2.8.1. IBM SPSS Statistics 19.0.....	64
2.8.2. IBM SPSS Modeler (Clementine 11.1).....	64
2.9. LİTERATÜR ÇALIŞMASI.....	67
3. UYGULAMA	75
3.1. ARAŞTIRMANIN YÖNTEMİ.....	75
3.2. UYGULAMA ADIMLARI	75
3.2.1. Problemin Tanımlanması	75
3.2.2. Verilerin Hazırlanması.....	76
3.2.3. Modelin Kurulması	78
3.2.3.1. Çoklu Regresyon Analizi	79
3.2.3.2. Chaid Algoritması Analizi	85
3.3. SINIFLAMA MODELLERİNİN KARŞILAŞTIRILMASI.....	96
4. SONUÇ ve ÖNERİLER	97
KAYNAKLAR	101
ÖZGEÇMİŞ.....	110

ÖNSÖZ

Araştırmanın amacı; veri madenciliği tekniklerini kullanarak Türkiye'nin sayılı entegre demir – çelik fabrikalarından biri olan A işletmesinde 1943 ve 1994 yılları arasında üretimi yapılan ve uzun ürünler olarak nitelendirilen ürün grubuyla ilgili olarak üretim temelli bir modelleme çalışması olarak belirlenmiştir.

Araştırmanın ilk evrelerinde uzun ürün grubu ile ilgili, özelleştirmeden sonraki döneme ait verilerin kullanılması hedeflenmiş olsa da, çeşitli nedenlerden dolayı veriler temin edilememiş ve çalışmanın Türkiye ve demir-çelik sektörü için önemi gözetilerek özelleştirmeden önceki döneme ait veriler Başbakanlık Yüksek Denetleme Kurulu (BYDK)' ndan temin edilerek çalışma yürütülmüştür.

Bu tez çalışmasının planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteğini esirgemeyen, bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren hocam sayın Doç. Dr. Filiz ERSÖZ'e, sayın Doç. Dr. Abdullah KARAKAYA'ya ve sayın Yrd. Doç. Dr. Mehmet PEKKAYA'ya teşekkürlerimi sunarım.

Çalışmanın temelini oluşturan uzun ürünler ile ilgili verilerin teminindeki desteklerinden dolayı Başbakanlık Yüksek Denetleme Kurulu (BYDK), Demir Çelik Üreticileri Derneği (DÇÜD) çalışanlarına teşekkürlerimi sunarım.

Sevgili eşim Öznur ile güzel kızımız Elif Berin'e, babam, annem ve kardeşlerime manevi hiçbir yardımı esirgemediğim yanımda oldukları için tüm kalbimle teşekkür ederim.

Bülent ORDU

KISALTMALAR

AB	Avrupa Birliđi
ABD	Amerika Birleşik Devletleri
AET	Avrupa Ekonomik Topluluđu
BAKKA	Batı Karadeniz Kalkınma Ajansı
CRISP-DM	Cross Industry Standard Process for Data Mining
CPS	Corporate & Public Strategy Advisory Group
EAO	Elektrikli Ark Ocakları
ERDEMİR	Eređli Demir-Çelik Fabrikaları
ETS	Emissions Trading Scheme (AB Emisyon Ticareti Planı)
GA	Genetik Algoritmalar
İMMİB	İstanbul Maden ve Metaller İhracatçı Birlikleri
İSDEMİR	İskenderun Demir Çelik Fabrikaları
KARDEMİR	Karabük Demir Çelik Fabrikaları
MKEK	Makine Kimya Endüstrisi Kurumu
MOLAP	Çok boyutlu Çevrimiçi Analitik İşleme
OLAP	Çevrimiçi Analitik İşleme
OLTP	Çevrimiçi Hareket İşleme
ROLAP	İlişkisel OLAP
SPSS	Statistical Package for the Social Sciences
TC	Türkiye Cumhuriyeti
TDÇİ	Türkiye Demir Çelik İşletmeleri
TOBB	Türkiye Odalar ve Borsalar Birliđi
TÜİK	Türkiye İstatistik Kurumu
VM	Veri Madenciliđi
VA	Veri Ambarı
VTBK	Veritabanlarında Bilgi Keşfi
WSA	World Steel Association (Dünya Çelik Birliđi)
YSA	Yapay Sinir Ağları

ÖZET**VERİ MADENCİLİĞİNDE SINIFLAYICI TEKNİKLER İLE DEMİR ÇELİK SEKTÖRÜNDE UZUN ÜRÜNLERİN ÜRETİMİNE İLİŞKİN BİR TAHMİN MODELLEMESİ**

ORDU Bülent

Yüksek lisans Tezi, İşletme Anabilim Dalı

Tez Danışmanı: Doç. Dr. Filiz ERSÖZ

03.01.2013, 140 sayfa

Bu çalışmada; demir çelik sektörü ile demir-çelik sektöründe üretimi yapılan ve uzun ürünler olarak nitelendirilen ürün grubu incelenmiştir. Ayrıca veri madenciliği tekniklerini kullanarak uzun ürün grubu ilgili olarak; uzun ürünlerin üretimine ilişkin değişkenler incelenmiş ve veri madenciliğinde sınıflandırma temelli teknikler kullanılarak yapılan analizler ile demir çelik üretim miktarına ilişkin tahmin modeli ve üretimi etkileyen en önemli değişkenler bulunmuştur. Çalışmada, öncelikle demir çelik sektörü hakkında bilgiler verilerek veri madenciliği (VM), veri madenciliğinin tarihsel gelişimi, bileşenleri, sınıflamaları, kullanım alanları, farklı alanlarda geliştirilen VM uygulamalarında kullanılan sınıflandırma temelli modeller incelenmiştir.

Çalışmada Başbakanlık Yüksek Denetleme Kurumu (BYDK)' ndan elde edilen, uzun hadde ürünlerinin 1943-1994 yılları arasındaki üretim verilerine, Veri Madenciliği yöntemleri uygulanmıştır. Elde edilen bilgiler doğrultusunda üretim temelli modeller oluşturularak, çalışma sonuçlarının ileriye dönük alınacak yatırım kararlarına ışık tutması hedeflenmiştir. Çalışmanın son bölümünde ise, karar ağacı algoritmaları ve çoklu regresyon analizi uygulanarak, elde edilen bulgular dahilinde Chaid karar ağacı algoritmasının uzun ürünlerin tahmininde çoklu regresyon modeline göre daha etkili ve belirleyici sonuçlar ürettiği görülmüştür.

Anahtar Sözcükler: Demir Çelik Sektörü, Uzun Hadde Ürünleri, Veri Madenciliği, Çoklu Regresyon, Karar Ağaçları

ABSTRACT

**A PREDICTIVE MODEL FOR THE MANUFACTURING of LONG-ROLLED
PRODUCTS in IRON-STEEL INDUSTRY WITH CLASSIFICATIONS
TECHNIQUES of DATA MINING**

ORDU Bülent

Master Thesis, Department of Business

Thesis Advisor: Assoc. Prof. Dr. Filiz ERSÖZ

03.01.2013, 140 pages

In this thesis, iron and steel industry together with long-rolled products manufactured in iron and steel industry have been studied. Moreover, using data mining techniques, variables related to long-rolled products manufacturing have been studied and a predictive model related to the amounts of iron and steel production and the most important variables that affect the production have been found by the analyses of techniques based on classification in data mining and by studying variables related to long rolled products manufacturing. In this study, by giving the information about iron and steel industry first, data mining (DM), the historical development, components, classification and usage of data mining, descriptive and predictive models of data mining that are used in different areas have been studied.

In this study, methods of data mining have been applied to the manufacturing data of long-rolled products that were obtained from BYDK in between 1943-1994. With these data, by creating models based on manufacturing, the results of the study are aimed to shed light on the investment decisions in future. In the last part of the thesis, by applying decision tree algorithms and multi regression analysis, it was concluded that Chaid decision tree algorithm produced more effective and determining results than multi-regression model.

Key Words: Iron and Steel Industries, Long - rolled Products, Data Mining, Multi-regression, Decision Trees.

ŞEKİLLER LİSTESİ

Şekil 1.1. Türkiye'nin Yıllar İtibariyle Uzun Ürünler Üretimi.....	9
Şekil 1.2. Dünya Çelik Üretiminin Coğrafik Dağılımı	15
Şekil 1.3. Dünya Nihai Çelik Ürünleri Tüketimin Coğrafi Dağılımı	16
Şekil 2.1. Veri Madenciliği ile İlişkili Alanlar	32
Şekil 2.2. Veritabanı Sistem Teknolojisi Gelişim Süreci	33
Şekil 2.3. Veri Madenciliği Süreci.....	37
Şekil 2.4. Budanmamış Karar Ağacı Örneği	57
Şekil 2.5. Clementine'de Kullanılan Veri Madenciliği Modellerinin Simgeleri	66
Şekil 2.6. Clementine 11.1 Kullanıcı Ara Yüzü	66
Şekil 3.1. Üretim Miktarı Tahmini İçin Oluşturulan Karar Ağacı.....	85
Şekil 3.2. Chaid Analizine İlişkin Kurallar Bütünü	86
Şekil 3.3. Karar Ağacının Başlangıç Düğümü.....	87
Şekil 3.4. Düğüm 0: Tahmini Üretim Miktarı	87
Şekil 3.5. Karar Ağacının Birinci Düğümü	88
Şekil 3.6. Düğüm 1: Tahmini Üretim Miktarı	88
Şekil 3.7. Karar Ağacının 2-11 Dğümleri.....	89
Şekil 3.8. Düğüm 2-11 Tahmini Üretim Miktarları	89
Şekil 3.9. Karar Ağacının 12-22 Dğümleri.....	90
Şekil 3.10. Düğüm 12-22 Tahmini Üretim Miktarları	91
Şekil 3.11. Karar Ağacının 23. Dğüümü.....	92
Şekil 3.12. Düğüm 23: Tahmini Üretim Miktarı	92
Şekil 3.13. Sınıflandırma Doğruluğu Grafiği	95

TABLOLAR LİSTESİ

Tablo 1.1. Demir Çelik Sektöründe Üretilen Ara ve Ana ile Yan Mamuller	6
Tablo 1.2. Dünyada Uzun Ürünlerin Üretim Miktarları	8
Tablo 1.3. Türkiye’de Uzun Ürünlerin Üretim Miktarları.....	9
Tablo 1.4. Dünya Ham Çelik Üretim Miktarları.....	13
Tablo 1.5. 2010 Yılı En Büyük Demir Çelik Üreticisi Ülkeler.	14
Tablo 1.6. 2011 Yılı Başlıca Çelik Üreticileri.	14
Tablo 1.7. Sektörlere Göre Demir Çelik Tüketimi.	17
Tablo 1.8. Avrupa Ülkelerindeki Ham Çelik Üretimi.	20
Tablo 1.9. Türkiye’nin Üretim Yöntemi Bazında Ham Çelik Üretimi.....	22
Tablo 1.10. Çelik Üretim Tesislerinin Bölgelere Göre Dağılımı.....	23
Tablo 1.11. Türkiye’de Demir Çelik Sektöründe İstihdam Değerleri.	23
Tablo 1.12. Türk Demir Çelik Sektörünün Yıllara Göre Ham Çelik Üretimi.	26
Tablo 1.13. Türkiye’nin Ürün Bazında Ham Çelik Üretimi.	27
Tablo 1.14. Türkiye’nin Ham Çelik Kapasitesi ve Kapasite Kullanım Oranları.....	27
Tablo 1.15. Türkiye’nin Toplam Çelik Üretimi.....	28
Tablo 1.16. Türkiye’nin Toplam Çelik Tüketimi.	28
Tablo 1.17. Türkiye’nin Çelikhane Kapasitesi ve Kapasite Kullanım Oranları.	29
Tablo 2.1. Korelasyon Analizi	55
Tablo 3.1. Regresyon Model Özeti	79
Tablo 3.2. Regresyon Modeli Anova Sonuçları.....	79
Tablo 3.3. Çoklu Regresyon Modeli/Bağımsız Değişkenler ve Katsayıları.....	80
Tablo 3.4. Değişkenlerin Tanımlayıcı İstatistikleri	81
Tablo 3.5. Korelasyon Tablosu	81
Tablo 3.6. Model Özeti Tablosundan Otokorelasyonun Tespiti.....	82
Tablo 3.7. Modelin Koşul İndeksi (CI) ve Özdeğer Vektör Değerleri	82
Tablo 3.8. Tolerans ve VIF Değerleri.....	83
Tablo 3.9. Artıkların İstatistiği.....	84
Tablo 3.10. Tahmini Üretim Miktarı ve Tahminin Güvenilirlik Değerleri	94
Tablo 3.11. Üretim Miktarı Analiz Sonuçlarının Karşılaştırması.....	95
Tablo 3.12. Üretim Miktarı Analiz Sonuçlarının Karşılaştırması.....	96

GİRİŞ

Türkiye ve dünya için önemli bir sektör olan demir çelik sektörü küresel ekonomide de oldukça önemli bir yere sahiptir. Çok eski zamanlardan beri çeşitli sanayilerin gelişmesinde ve toplumların kalkınmasında büyük önem arz eden demir çelik sektörü diğer sektörlerle öncülük yapmış ve yapmaya devam etmektedir. Üretilen ürünlerin çeşitliliği göz önüne alındığında ise geleneksel sanayilerin yanı sıra teknolojik alanlarda da demir çelik sektörünün öneminin giderek daha çok arttığı görülmektedir. Demir çelik üretiminde meydana gelen hızlı gelişmeler sonucunda endüstri devrimi gerçekleşmiş ve teknoloji alanında büyük gelişmeler yaşanmıştır.

Demir-çelik sektörünün ülkeler açısından önemi göz önüne alındığında, ihtiyaç duyulan miktarda üretimin yapılması büyük önem taşımaktadır. Bu noktada üretim miktarını etkileyen değişkenlerin belirlenerek, üretim miktarını tahmin edecek modellerin kurulması ve bu modeller içerisinde en iyisinin seçilmesi gerekir.

Çalışmada iki amaç belirlenmiş ve bu amaçları gerçekleştirmek için iki ayrı yöntem kullanılmıştır. İlk amaç, uzun ürün grubunun üretimini etkileyen değişkenler ile üretim miktarı arasındaki ilişkinin belirlenmesi, ikinci amaç ise uzun ürün grubunun geçmiş yıllardaki üretim miktarlarının değişimleri incelenerek, gelecek yıllardaki üretim miktarlarının tahmin edilmesi olarak belirlenmiş, bu amaca yönelik olarak da sırasıyla çoklu regresyon analizi ve Chaid karar ağacı uygulanmıştır. Bu analizler sonucunda; bağımlı değişken olan üretim miktarı, bağımsız değişkenlerin durumuna göre sınıflandırılarak, en uygun üretim miktarı tahmini için kurallar bulunması amaçlanmıştır. Elde edilen kurallar, demir çelik sektöründe uzun ürünler olarak nitelendirilen ürünlerin üretimlerine ilişkin önemli bilgiler vermekte ve yeni üretim faaliyetlerin planlanması noktasında karar desteği sağlamaktadır. Uygulamanın sonunda belirlenen modellerden elde edilen sonuçlar karşılaştırılarak yöntemlerin tahmin doğrulukları ortaya konmuştur.

Tez çalışması dört bölümden oluşmaktadır. Birinci bölümde, sektör ve ürünlerle ilgili, ikinci bölümde ise uygulamada kullanılan tekniklerle ilgili ayrıntılı bilgi verilerek, literatür çalışmaları anlatılmıştır. Üçüncü bölümde çalışmanın uygulama adımları anlatılmış, dördüncü bölümde ise sonuç ve önerilere yer verilmiştir.

1. DEMİR - ÇELİK SEKTÖRÜ

Dünyadaki gelişmelerden en fazla etkilenen sektörlerin başında kuşkusuz demir-çelik sektörü gelmektedir. Tüm endüstriyel alanlarla en fazla bağlantısı olan sektörlerden biri olan demir-çelik sektöründe tarih boyunca meydana gelen gelişmeler toplumların sosyo-ekonomik yapılarını şekillendirmede büyük rol oynamıştır. Demir - çelik sektörü, sanayi toplumu sürecinin çekirdeğini oluşturan bir sektör olarak, bilgi toplumuna geçişle birlikte önemini korumuş, tüm endüstriyel dallara girdi temin etmesinden dolayı, imalat sanayi, dayanıklı tüketim ve yatırım malı sanayileri üzerinde belirleyici bir unsur olmuştur (Atgür, 2006:5).

Dünya genelinde toplam ekonomik faaliyetlerin artması ile ülke ekonomilerindeki büyümenin bir yansıması olarak konuta olan ihtiyacın, otomobile olan talebin ve başta beyaz eşya olmak üzere diğer demir çelik ürünleri talebinin her geçen gün büyük bir hızla artması, dünya çelik üretimini de artırmış ve arttırmaya devam etmektedir (T.C. Sanayi ve Ticaret Bakanlığı, 2011:7).

1.1. DEMİR – ÇELİK TANIMI

Saf demir, dövülebilen, işlenebilen, kolayca tel ve levha haline getirilebilen orta sertlikte bir metaldir. Ham demir ve demir süngeri, çeliklerin ve demir döküm malzemelerin üretiminde kullanılan temel malzemelerdir. Demir doğada saf halde bulunmayıp genelde bileşik halindedir. Demir cevheri, demirin oksijen ve kükürtlü bileşiklerle oluşturduğu kimyasal bileşiklerdir (MEB Metal Meslek Bilgisi Ders Kitabı 1998:258).

En önemli demir cevherleri şunlardır:

- Manyetik demir taşı: % 60 – 70 demir içerir.
- Kırmızı demir taşı: % 40 – 60 oranında demir içerir.
- Kahverengi demir taşı: % 30 – 50 demir içerir.
- Siderit: % 30 – 45 demir içerir.

Demir cevherinden elde edilen ham demiri şekillendirmek ve mukavemet özelliğini arttırmak için ihtivasındaki karbon miktarının % 2'nin altına indirilerek elde edilen demir ve karbon bileşiğine ise çelik denir. Çelik içerisinde yaklaşık % 1,7

karbon, % 1 mangan, % 0,5 silisyum, % 0,05 kadar da kükürt ve fosfor içerir (İzmir 2011:30).

1.2. DEMİR - ÇELİĞİN ÜRETİM AŞAMALARI

Demir - Çelik ana sanayi, demir cevherinin arıtılmasıyla, demir ve çelik üretimini ve bunların şekillendirilmesiyle çubuk, levha, sac, profil, boru ve tel gibi ürünlerin elde edilmesi ve bunların koruyucu maddelerle kaplanması işlemlerini kapsamaktadır. Bu işlemler başlıca iki aşamada gerçekleştirilmektedir. Birinci aşamada demir cevheri arıtılıp, çelik üretimine elverişli ham demir haline getirilmekte, ikinci aşamada ise ham demir, çelik üretiminde kullanılmaktadır (<http://www.isdemir.com>).

Çelik üretimi için iki ana üretim tekniği vardır. Bunlardan birincisi konvansiyonel yöntem olarak da adlandırılan yüksek fırın ve bu fırında üretilen ham demir ile, diğeri de yüksek fırın dışında redüksiyon yöntemi ile sünger demir adı verilen yarı mamul ile çelik üretimidir. Ham demir, sünger demir ve hurda, çelik üretiminin en önemli hammaddeleridir. Çelik üretim süreci üç temel malzemenin işlenmesi ile başlar. Bu malzemeler kireçtaşı, kömür ve demir cevheridir. Öncelikle kömür, kok oluşturmak için kok fırınlarında ısıtılır. Bu sürece karbonizasyon adı verilir. Oluşan kok soğumaya bırakılır. Yüksek fırın sürecini hızlandırmak ve sinter oluşturmak için demir ve kireç taşı eşzamanlı olarak sinterleme işlemine alınır (Emel, 2004:209).

Çelik demir cevherinden veya hurdadan geri dönüşüm ile iki şekilde üretilmektedir. Sıvı çelik üretildikten sonra döküm ile ingot olarak veya sürekli döküm yöntemi ile kütük veya blum olarak şekillendirilir. Vasıflı çelikler alaşimsız, düşük alaşımli ve alaşımli çelikler olup, kitlesel olarak üretilen çeliklerden bazı noktalarda ayrılmaktadır. Bu noktalar; üretim yöntem ve araçları ile alt limitlerde bulunan çözünmüş gaz miktarlarıdır (<http://www.kardemir.com>).

Çeliklerin büyük çoğunluğu ısıl işlemlere karşı duyarlıdır. Kimyasal bileşimin yanı sıra uygulanan ısıl işlemler sonucunda istenen sertlik, mekanik ve fiziksel özellik, elektriksel özellik, korozyona ve yüksek sıcaklığa dayanım özelliklerine tam olarak kavuşturulabilir. Çelikler yapılarının gerektirdiği sıcaklıklara kadar ısıtıldıklarında şekillenme özelliğine kavuşur. Ayrıca kimyasal bileşim ve içyapı olarak uygun olan çelikler haddeleme, presleme gibi yöntemlerle soğuk olarak da şekillendirilebilir. Talaş kaldırıcı tezgâhlarda işlenerek, istenilen şekil ve yüzey düzgünlüğüne getirilebilir. Kimyasal bileşim olarak uygun olan çelikler kaynak işlemi ile birleştirilebilir. Çeliklerin

büyük bir bölümü çeşitli yöntemler ile metal ile kaplanmaya, emaye yapılmaya, boyanmaya ve plastik maddeler ile kaplanmaya elverişlidir (<http://www.kardemir.com>).

Demir - Çelik Üretiminde Kok Kömürünün Üretim Aşamaları

Taşkömürü, demir-çelik sektörü açısından stratejik öneme sahip bir hammaddedir. Türkiye’de taşkömürü varlığı olan tek bölge Zonguldak Havzası’dır. Demir-Çelik sektörü için taşkömürünün önemi oldukça fazla olup, son yıllarda dünyadaki değişen şartlar çerçevesinde, taşkömürü fiyatlarının sürekli yükseldiği görülmektedir (Taşkömürü Çalışma Grubu, 2006:153).

Yurtiçi ve yurtdışı kaynaklardan getirilen çeşitli özelliklerdeki taşkömürlerini stoklama, hazırlama, harmanlama, tartma ve kok fırınlarını beslemek amacıyla kömür hazırlama tesisleri inşa edilmiştir. Entegre tesislerde demir ve çelik üretimi büyük miktarda enerji gerektiren bir süreçtir. Bu enerjinin bir kısmı kok üretimi sırasında bataryalardan elde edilir. Kok kömürü, koklaşabilir kömür ya da kömür harmanlarının kok kamaralarında havasız ortamda damıtılması, gazlaştırılması, yani yüksek sıcaklıkta içindeki uçucunun alınması ile üretilir. İşletme pratiği ve ekonomisi açısından belli sayıda kok kamarasının bir araya gelmesi ile kok bataryaları oluşur. Kok bataryalarında kömürün koklaşması sırasında çıkan kok gazının soğutulması ile su buharı ve katran yoğunlaştırılarak ayrıştırılır. Kok bataryaları amonyağın, aromatik hidrokarbonların (benzol ve türevleri) çeşitli metotlarla temizlenmesi işlemlerinin yapılarak temiz kok gazının Demir-Çelik tesislerinde kullanıcı ünitelere sevk etmek ve kok gazını temizleme sırasında kazanılan kimyasal ürünleri işleyerek satışa hazır duruma getirmek amacıyla kurulmuştur. Bu tesislerde ham katran, amonyum sülfat (gübre), ham benzol üretilir (<http://www.isdemir.com>).

Yüksek sıcaklıkta kömür molekülü kapalı ve havasız bir ortamda ısıtılırsa kömür molekülü parçalanır. Böylece oluşan uçucu maddeler kömür bünyesinden çıktıktan sonra ortamda kalan sert, gözenekli, süngerimsi karbon yüzdesi fazla yapıya kok, yapılan işleme de koklaşma denir. Koklaşma işlemi, havasız ortamda yüksek sıcaklıkta (1100 °C) kömürün karbonizasyonunu içerir. Kok fırınları, silika tuğlalarla yapılmış dar ve uzun kamaralardır. Üflenen sıcak hava ile reaksiyona girip, yüksek fırın içerisindeki kimyasal reaksiyonların endotermik ısı gereksinimlerinin sağlanması, demir cevheri ve cürufun ısıtılması, eritilmesi ve demir cevherinin indirgenmesi için gerekli ısının sağlanması koklaşma ile mümkündür. (Çakır, 2006:2-5)

Son zamanlarda, özellikle de Amerika Birleşik Devletleri'nde, ürün maliyetlerini düşürebilmek amacıyla demir çelik üretiminde kok yerine doğalgaz tercih edilmeye başlanmıştır. Üretim bazı aşamalarında doğalgaz kullanımının daha pahalı olan kok kömürüne göre ton başına 6-7 \$ tasarruf sağlayacağı öngörülmüştür (Demir Çelik Store Dergisi, 2012:52).

Demir - Çelik Üretiminde Yüksek Fırın Prosesi

Demir cevheri, pelet, sinter gibi demirli malzemeler ile kireçtaşı, kuvarsit, dolomit gibi cüruf yapıcı malzemelerin metalürjik kokun yanması ile eritilerek sıvı ham demirin elde edildiği üniteye yüksek fırın denir. Yüksek fırınların kurulma amacı çelikhanenin ihtiyacı olan sıvı ham demirin üretilmesidir (<http://www.isdemir.com>).

“Toz halindeki demir cevherleri ile diğer demirli malzemeler ve katkı malzemelerinden (Kireçtaşı, dolomit taşı gibi) meydana gelen sinter harmanının kok kömürü ile yakılarak, yüksek fırınlarda kullanılır hale getirilmesi işlemine sinterleme denir” (<http://www.isdemir.com>).

Cevher ve yakıt, fırının tepesinden eklenirken, fırının alt kısmından da hava üflenerek kimyasal tepkime başlatılır. Bu reaksiyon sonucu metal oksitler indirgenerek sıvı ham demire dönüşür, indirgenemeyen diğer oksitler ise cüruf oluştururlar. Ergimiş metal ile cüruf fırının alt kısmına doğru ilerlerken oluşan gazlar da fırının üst kısmına hareket ederler (İzmir 2011:33) .

Çeliklerin şekillendirilmesinde kullanılan başlıca metot plastik şekillendirmedir. Haddeme ise çeliklerin plastik şekillendirilmesi için kullanılan en iyi yöntemdir. Çelikler sıcak ve soğuk haddeme olmak üzere iki şekilde haddelenirler (Akyalçın ve Yılmaz, 2007:231).

Haddehane, çelikhanede üretilen ingotların veya satın alınan blumların işlenmesiyle çeşitli profilde, uzun hadde ürünleri, inşaat ve yapı çelikleri, maden direği ve ray gibi ürünlerin üretildiği ünite dir.

Demir çelik sektöründe üretilen ürünler ara mamuller, ana mamuller ve yan mamuller olmak üzere üç kısımda incelenir. Ara mamul başka bir malın üretiminde girdi olarak kullanılmak üzere üretilen ürünlerdir. Ana mamul bir başka işlem gerektirmeyen, nihai kullanım için üretilen ürünler olup yan mamuller ise üretim süreci sonunda teknik nedenlerle ana ürün ile beraber ortaya çıkan ürünlerdir.

Yarı mamul, yan mamul ve ana mamul çeşitleri Tablo 1.1’de gösterilmiştir:

Tablo 1.1. Demir Çelik Sektöründe Üretilen Ara ve Ana Mamuller ile Yan Mamuller

Kok, Kok Tozu, Kok Gazı, Oksijen, Ham Benzol, Pres Naftalin, Granüle Cüruf, Ham Katran, Kreozot, Zift ve Katran Boyası, Amonyum Sülfat	Yan mamul
Sıvı Ham Demir, Pik, Demir Pik (Kanal + Karışık Pik), Sıvı Çelik, Kütük, İngot, Blum	Ara mamul
Yuvarlaklar, Demir Yolları Malzemeleri (Ray), Maden Direği, Profil, Köşebent, Platina	Ana mamul

1.3. DEMİR - ÇELİK ÜRÜNLERİNİN SINIFLANDIRILMASI

Demir-çelik ürünleri, yassı ürünler, uzun ürünler ve vasıflı çelikler olmak üzere üç ürün grubu içinde sınıflandırılmaktadır.

Uzun hadde ürünleri alt sektörü, demir cevheri veya hurdadan başlamak suretiyle, sıvı çeliği çeşitli yöntemlerle dökerek, ingot (külçe), blum ve kütük haline dönüştüren çelikhaneler ile, ingot blum ve kütüğü haddelemek suretiyle blum, kütük, demiryolu malzemesi, ağır, orta ve hafif profil, nervürlü veya düz betonarme çelik çubuklar, tel ve kangal (filmaşın) üretimi yapan haddehaneleri kapsamaktadır.

Uzun hadde ürünleri alt sektörü ağırlıklı olarak inşaat sektörüne, kaliteli filmaşın üretimi ile otomotiv ve otomotiv yan sanayisine, profil üretimi ile yapısal çelik sanayisine, ray üretimi ile de demiryollarına girdi sağlamaktadır. Demir çelik sektöründe yassı çelikler; sıcak haddelenmiş ürünler, soğuk haddelenmiş ürünler, levhalar, tenekeler, galvanizli saclar ve diğer hadde mamulleri olmak üzere altı ana ürün grubuna ayrılır. Ülkemizde yassı ürün üretimi yapan firmalar Erdemir, İsdemir, Borçelik ve Tezcan demir çelik tesisleridir (Metal Sektörü Yeterlilikler Kılavuzu, 2011:15).

Demir-çelik sektörü içinde kalkınmanın önemli bir ölçütü olarak görülen ve önemi giderek artan vasıflı çelikler ise, özel yapı çelikleri, paslanmaz çelikler ve ısıya dayanıklı çeliklerden oluşan orta ve yüksek alaşımli çelikler olup otomotiv ve otomotiv

yan sanayi, savunma sanayi, makine ve yay imalat sanayinde kullanılmaktadır (Atgür, 2006:5).

Dünya’da teknolojik yönden gelişmiş ülkeler, demir-çelik üretiminde, miktar olarak fazla üretim yapmaktansa; yassı, vasıflı, paslanmaz, kaplanmış veya daha özel, katma değeri yüksek çelik ürünlerin üretimine yönelmeye başlamışlardır. Buna karşılık, Türkiye de dahil olmak üzere, gelişmekte olan ülkeler, miktar olarak fazla üretim yapma eğiliminde kalmışlardır. Bu nedenle, vasıflı çelik, paslanmaz çelik ve yassı ürünlerde üretim yapmak önem arz etmektedir. 2011 yılında, Demir-Çelik Sektörünün ülkemizin toplam ihracatı içerisindeki payının % 10 olması sektörün ihracat potansiyelini göstermektedir (T.C. Bilim, Sanayi Ve Teknoloji Bakanlığı Demir Çelik Sektörü Raporu, 2012:12).

Yassı ürünler ve vasıflı çelik ürünlerin entegre tesislerde üretilebilmesi, uzun ürünlerin ise elektrik ark ocaklı tesislerde üretilebilmesi dolayısıyla, entegre tesislerin ağırlıkta bulunduğu ülkeler katma değeri yüksek olan yassı ürünlerin ve vasıflı çelik ürünlerinin ihracatçısı konumunda iken, üretimin çoğunluğunu elektrik ark ocaklı tesislerde gerçekleştiren ülkeler ise bu ürünlerin ithalatçısı konumundadır (Eruz, 2003:7).

Ülkelerin sanayi toplumu olmasında önemli bir yeri olan uzun ürünler sektörü hem yaşanan ekonomik gelişmelerden etkilenmiş hem de bu gelişmeleri etkileyen önemli bir unsur olmuştur. Uzun ürünler, sektörünün günümüzde gelişmiş ülkeler açısından nispi önemi azalmasına karşılık; gelişmekte olan ülkeler açısından önemini korumaktadır. Uzun ürünler üretim sektörü, inşaat yapı sanayiden demiryollarına kadar önemli girdiler temin ettiğinden sektörlerin maliyet yapılarını yakından etkilemektedir.

Uzun ürünler daha çok inşaat sektöründe kullanılan ürünlerdir. Gelişmiş ülkelerde altyapı yatırımları büyük ölçüde tamamlanmış olduğundan uzun ürün üretimlerinde yıllar bazında bir düşüş olduğu görülmektedir. Başta Çin ve Hindistan olmak üzere gelişmekte olan ülkelerde ise uzun ürünlerin üretiminde hızlı bir artış olduğu gözlemlenmektedir.

Tablo 1.2’de Dünyada uzun ürünlerin üretim miktarları, yıllar itibarıyla gösterilmiştir:

Tablo 1.2. Dünyada Uzun Ürünlerin Üretim Miktarları (Bin Ton)

	2001	2002	2003	2004	2005	2006	2007	2008	2009
Avusturya	1.171	1.208	1.148	1.241	1.244	1.223	1.395	1.405	1.164
Fransa	4.753	4.611	4.747	4.768	4.363	4.624	4.648	4.535	3.523
Almanya	12.527	12.640	12.608	13.619	12.724	14.000	14.333	13.716	10.229
İspanya	9.542	10.448	10.879	11.593	11.629	12.364	12.547	12.378	10.233
İngiltere	5.671	5.024	4.352	4.926	4.360	4.548	4.451	4.439	3.111
Polonya	...	4.109	4.492	4.678	3.906	4.756	4.919	4.779	4.083
AB(27)	60.028	65.434	67.171	71.133	67.345	74.739	76.071	73.658	50.713
Türkiye*	8.619	8.831	9.910	11.369	14.241	17.871	20.663	20.918	19.015
ABD	25.356	25.141	26.832	27.619	26.115	27.665	28.139	25.824	16.081
Brezilya	6.963	7.182	7.422	8.419	7.059	9.050	10.159	10.361	8.371
İran	3.151	3.711	4.469	4.176	4.687	4.464	5.183	4.994	...
Çin	93.835	112.225	134.849	162.888	190.916	232.970	267.254	267.153	329.536
Hindistan	13.148	13.425	15.299	15.320	16.446	17.993	24.348	27.794	27.865
Japonya	33.835	34.533	33.992	34.741	33.761	35.638	36.394	34.393	23.322
Güney Kore	18.009	19.910	20.396	20.116	18.799	19.582	20.406	19.983	18.669
Asya	179.966	204.797	229.517	260.325	287.686	334.218	375.122	376.166	424.600
Dünya	299.116	330.636	361.685	398.263	419.228	482.840	529.584	525.791	528.497

Kaynak: World Steel Association, World Steel in Figures, 2011.

* World Steel Türkiye’de uzun ürün üretimini 2000 yılından sonrası için rapor etmediği için 2000 yılı sonrası uzun ürün üretimi TÜİK verilerinden düzenlenmiştir.

2010 yılı itibariyle Türkiye’nin uzun ürün üretimi 13,4 milyon tona ulaşmıştır. Üretimin 10 milyon ton gibi büyük bir kısmını nervürlü inşaat demiri oluşturmaktadır (Belen, Yıldırım ve Çakmak, 2011:44). Uzun ürünlerin toplam sıvı çelik üretimine oranı % 75 olup % 25’lik kısmını ise yassı ürünler oluşturmaktadır. 2000-2009 yıllarına ait ortalama büyüme hızı ise % 9,26’dır. Bu büyüme hızı dünyadaki önemli çelik üreticisi ülkeler arasında, Çin’den sonraki en büyük büyüme hızıdır (Belen, Yıldırım ve Çakmak, 2011:32).

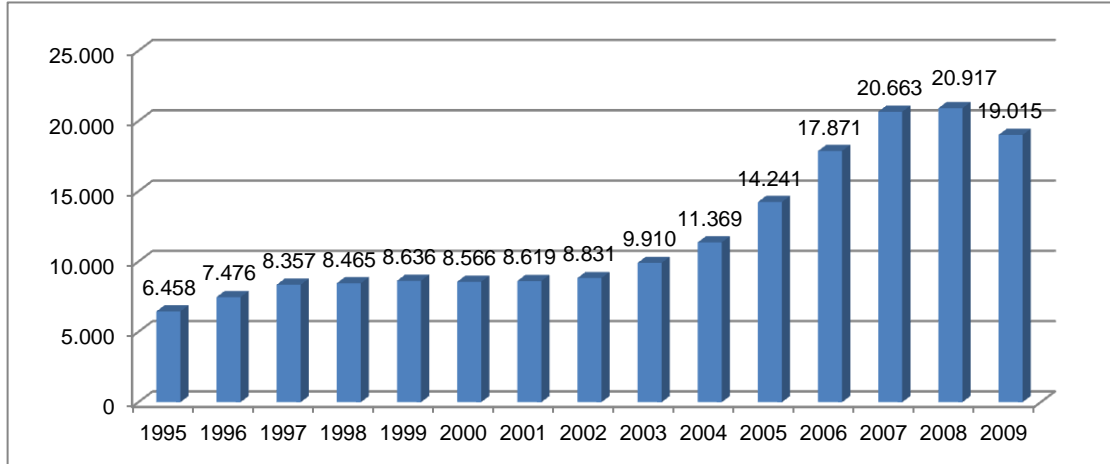
Türkiye’de uzun ürünlerin üretim miktarları Tablo 1.3’de gösterilmiştir:

Tablo 1.3. Türkiye’de Uzun Ürünlerin Üretim Miktarları (Ton)

Yıl	Ham Demir	Sıvı Çelik	Sac ve Levhalar	Profil	Yuvarlaklar	Uzun Ürünler
1995	4.363.488	12.797.709	2.081.140	202.798	6.255.113	6.457.911
1996	5.263.094	13.382.221	3.077.387	239.322	7.236.351	7.475.673
1997	5.566.864	13.644.093	3.549.440	475.576	7.881.888	8.357.464
1998	5.286.554	13.351.079	3.428.310	538.500	7.926.221	8.464.721
1999	5.181.328	13.670.226	3.596.335	620.329	8.015.233	8.635.562
2000	5.332.905	13.596.411	4.041.879	761.992	7.803.939	8.565.931
2001	5.289.167	14.382.257	3.726.692	576.919	8.042.140	8.619.059
2002	5.012.133	16.045.522	4.218.650	726.295	8.105.153	8.831.448
2003	5.693.764	17.644.020	4.719.059	876.129	9.033.826	9.909.955
2004	5.835.889	19.868.118	5.269.141	1.142.314	10.226.603	11.368.917
2005	6.344.200	20.961.240	5.753.199	1.245.757	12.995.111	14.240.868
2006	5.952.106	23.307.523	6.190.624	1.504.978	16.365.582	17.870.560
2007	6.235.537	25.760.889	6.771.352	1.979.643	18.683.442	20.663.085
2008	6.697.372	26.809.050	6.358.504	1.863.779	19.053.687	20.917.466
2009	6.913.325	25.303.741	6.437.623	2.019.682	16.994.836	19.014.518

Kaynak: İstatistikî Göstergeler 1923-2009, TÜİK, 2010, s.284.

Türkiye’nin yıllar itibariyle uzun ürünler üretimi Şekil 1.1’de gösterilmektedir:

**Şekil 1.1.** Türkiye’nin Yıllar İtibariyle Uzun Ürünler Üretimi (Bin Ton)

Uyarıldığı Kaynak: İstatistikî Göstergeler 1923-2009, TÜİK, 2010, s. 284.

Türkiye’de uzun ürünlerin üretildiği entegre tesisler olarak Kardemir ve İsdemir bulunmaktadır. Elektrik ark ocaklı tesislerde yıllar boyunca uzun ürün üretilmesine rağmen son yıllarda yassı ürün üretimine de başlanmıştır. Uzun ürün üreten tesisler

MKEK, Çolakoğlu, Kroman, İçdaş, Çemtaş, Asil Çelik, Ege Çelik, Diler, Habaş, İzmir Demir Çelik, Çebitaş, Ekinciler, Yazıcı, Yeşilyurt, Kaptan, Nursan, Sider, İlhan, Bilecik, Mega, Ede, Özkan, Yolbulan-Baştuğ' dur. Asil Çelik ve Çemtaş ayrıca vasıflı çelik ürünleri de üretmektedir.

Demir- çelik tesislerinde üretilen uzun ürünler kütük, blum, ray malzemesi, inşaat demiri, profil, filmaşın olup tanımları ve özellikleri şu şekildedir:

Kütük: Demir çelik fabrikalarında, izabe tesislerinde maden cevherinden veya hurdadan döküm sonu elde edilen ham kütleyle kütük demir denir. Kütük, bazik oksijen fırınlarından elde edilen sıvı çeliğin sürekli döküm makinelerinde dökülmesiyle üretilir. Kütük yarı mamul olarak filmaşın, inşaat çeliği, profil ve dikişsiz boru imalatında kullanılmaktadır (İzmir 2011:49).

Blum: 20 cm genişliğinden fazla kare veya dikdörtgen kesitli sürekli döküm çelikten üretilen yarı mamullere blum denir. Blum; ray, çubuk ve profil üretiminde kullanılan daha büyük ölçekli bir kütüktür (İzmir 2011:49).

İnşaat Demiri: Betonarme yapılarda kayma ve çekme gerilmelerini karşılamak amacı ile beton içine konulan, özel şekillendirilmiş çeliğe inşaat demiri denir. İnşaat demiri, çelik hasır tüm yapıların iskeletidir. Nervürlü inşaat demiri, çelik hasır dünyada, değişik ülkelere ait standartlara göre üretilmektedir. Bu standartların gelişimi ile nervürlü inşaat demirin mekanik özellikleri yıllar içinde gelişmiş ve en kaliteli seviyeye gelmiştir. Maliyet ve üretim süresi düşmüş, kalite ve üretim tonajı yükselmiştir. İnşaat demiri ve çelik hasırın analiz özelliklerinin değişmesinin sebebi ise ihtiyacın fazlaşması ve daha ekonomik demirlere ihtiyaç duyulmasıdır.

İnşaatların, statik yüke göre veya dinamik yüke göre göz önünde tutulan en önemli noktası inşaat demiri akma dayanıklılığıdır. Buna ek olarak, son yıllarda kullanılan inşaat demiri uzaması ve çekme dayanımı da göz önüne alınmaktadır. İnşaat demiri analiz özelliklerinin artırılması neticesinde inşaatlarda kullanılacak inşaat demiri kesiti düşürüleceğinden, daha az inşaat demiri kullanılmakta ve ağırlıktan, nakliyeden ve hacimden kazanılmaktadır. Kısaca maliyetler düşmektedir.

Filmaşın (Kangal): Sıcak haddelemeyle üretilen, genellikle yuvarlak veya yarı yuvarlak, kare, dikdörtgen ve altıgen kesitinde soğuk çekme inşaat çeliği donatısında (etriye, gönye, pilye, çiroz), fore kazık donatılarında ve sanayide (tel, çivi, çelik hasır,

cıvata, somun, kaynak elektrotu) kullanılan kangal halindeki yarı mamul metal çubuklara filmaşın denir.

Profiller: Profiller; standart profiller, kapalı profiller ve açık yapı profiller olmak üzere üç kısımda üretimi yapılan yapı elemanlarıdır. Standart profiller tek başlarına, perçinli, cıvalı ve kaynak bağlantılarıyla birleştirilerek kullanılan yapı elemanlarıdır (İzmir, 2011:51-52).

Ray: Demiryolu rayı; değişken yükler altında, yüksek zorlamaları, üzerinde oturduğu malzemeye intikal ettiren bir yapı elemanıdır. Demiryolu taşımacılığındaki gelişmeler; yüksek hız, uzun vagon boylarında daha büyük tekerlek yükleri, kaynaklı rayların kullanımındaki gelişmeler, daha büyük ivmeler ve daha kısa fren mesafeleri olarak sıralanabilmektedir. Performans artışını sağlamak için raylarda ihtiyaç duyulan ana özellikler; aşınma direnci, yorulma direnci, plastik deformasyon direnci, kalıntı gerilmelerin giderilmesi ve kaynaklanabilirliktir (Kalaycıoğlu, 2006:29).

Karabük Demir Çelik Fabrikaları'nın faaliyete geçmesinin ardından, 1950'li yıllarda küçük partiler halinde ray üretimi yapılmış ve ülke ihtiyaçları karşılanmaya çalışılmıştır. Daha sonraki dönemlerde uzun bir süre ray üretimine ara verilmiştir. 2001 yılında gerekli düzenlemeler yapılarak tekrar ray üretimine başlanmıştır. Kardemir A.Ş. Haddehaneler Müdürlüğü'nde üretimi hali hazırda devam eden rayın üretim aşamalarında sistemin optimizasyonu için yapılan çalışmalarla kalite en üst seviyede tutulmaya çalışılmaktadır (Kalaycıoğlu, 2006:39).

1.4. DEMİR - ÇELİK SEKTÖRÜNÜN ÖZELLİKLERİ

Demir çelik sektörü, demir cevherinin yer altından çıkartılmasını takiben, yoğunlaştırılmasından başlamak üzere dökme, dövme, haddeleme, çekme ve benzeri yöntemler ile üretiminin gerçekleştirildiği bir sektördür (TOBB, 2010:11).

Sektör, çelik ürünlerin kullanım alanının yaygınlaşması, her geçen gün tüketiminin artması, imalat sanayine ara mal üretilmesi ve ihracat potansiyeli gibi niteliklerinden dolayı ülke ekonomisi açısından büyük önem taşımaktadır (T.C. Bilim, Sanayi ve Teknoloji Bakanlığı Demir Çelik Sektörü Raporu, 2012:4).

Metal sektörünün ana sanayisini oluşturan demir-çelik sektörü, birçok sektöre girdi sağladığı için sanayinin lokomotif sektörü olarak bilinir. İnşaattan otomotive, savunma sanayisinden elektroniğe kadar birçok sektörün demir-çelik sektörüne bağlı

olması sektörü stratejik açıdan önemli kılmaktadır. Demir çelik sektörünün temel özellikleri şunlardır:

- Sermaye ve teknoloji yoğunluklu yatırımlar gerektirmesi,
- Sektör içinde yer alan alt sektörlerin ekonomik ve teknolojik açıdan birbirlerine bağlı olmaları,
- Tekelleşme oranının diğer sektörlerle oranla düşük olması,
- Üretim yapılan yerleşim yerleri arasında bölgesel yoğunlaşmanın olması,
- Sektörün kamu öncülüğünde gelişmiş olmasıdır (Bakka, 2010:6-7).

1.4.1. Demir – Çelik Sektörünün Etkileşim Halinde Olduğu Diğer Sektörler

Demir çelik sektöründe, başta inşaat ve otomotiv olmak üzere, boru, profil, dayanıklı tüketim eşyası, yakıt araç ve gereçleri imalatı, tarım araçları imalatı ve gemi inşa sektörüne yönelik üretim yapılmaktadır. İnşaat, otomotiv, makine ve metal eşya sektörlerinin, toplam dünya çelik tüketiminin % 94'ünü gerçekleştirmesine rağmen, Çin gibi gelişmekte olan ülkelerde bu oranın AB gibi gelişmiş ülkelere kıyasla daha yüksek seviyelere çıkabildiği, benzer şekilde otomotiv endüstrisinin Almanya ve ABD'deki toplam çelik tüketiminin % 20'sini gerçekleştirdiği, bu oranın Çin'de % 3 seviyesinde kaldığı belirtilmektedir (T.C. Sanayi ve Ticaret Bakanlığı Demir Çelik Sektörü Raporu, 2011:8).

1.4.2. Dünya’da Demir - Çelik Sektörü

Dünya’da demir cevheri rezervleri, birçok bölge ve ülkelere dağılmış olarak büyüklü küçüklü rezervler halinde bulunur. Dünya’da ekonomik değeri olan demir cevheri rezervleri hakkında bilgiler "Mineral Facts and Problems" dergisinin 1995 yılı baskısında belirtilmiştir. Toplam dünya rezervi 357 milyar ton civarındadır. Bu miktar, rezervi 1,02 milyar ton civarında olan dünya demir cevheri üretimi göz önüne alındığında yaklaşık 350 yıllık üretim ömrüne karşılık gelmektedir (TOBB, 2010:11).

Dünya demir-çelik üretim ve tüketimi, 19. yüzyılın ikinci yarısından itibaren ABD ve Ortadoğu olmak üzere pek çok ülkede önemli bir yer tutmuştur. 20. Yüzyılda endüstriyel üretimdeki gelişmelerin başlaması nedeniyle makine, imalat, kimya, enerji ve madeni eşya üretiminde önemli ilerlemeler sağlanmıştır (Arıkan, 2000:34).

1995-2010 yılları arasında dünyada toplam ham çelik üretim miktarları ve sektörün büyüme oranları Tablo 1.4’te gösterilmiştir:

Tablo 1.4. Dünya Ham Çelik Üretim Miktarları (Milyon Ton)

Yıllar	Üretim	Yıllık Büyüme(%)	Yıllar	Üretim	Yıllık Büyüme(%)
1995	752		2003	970	7,30
1996	750	-0,27	2004	1.071	10,41
1997	799	6,53	2005	1.144	6,82
1998	778	-2,63	2006	1.247	9,00
1999	789	1,41	2007	1.347	8,02
2000	849	7,60	2008	1.329	-1,34
2001	851	0,24	2009	1.231	-7,37
2002	904	6,23	2010	1.414	14,87

Kaynak: World Steel Association, World Steel in Figures, 2011, s.7.

Demir çelik üretiminde 2000’li yıllarla birlikte önemli bir aktör konumuna gelen Çin’in sahip olduğu demir çelik üretim ve tüketim değerleri, dünya çelik fiyatları üzerinde belirleyici bir rol oynamaktadır. Çin’deki kapasite artışları dünya çelik fiyatlarını düşürmekte ve bu durumdan çelik üreticisi ülkeler olumsuz etkilenmektedir. Buna karşılık AB ülkeleri ve önemli çelik üreticisi diğer ülkeler kapasitelerini azaltmak yoluyla sektörü küçültüp, katma değeri arttırarak Çin’in etkilerini minimum düzeye indirme çabası içine girmişlerdir ve dönemsel dalgalanmalara sebebiyet vermişlerdir (İMMİB, 2005:30).

1998 - 2008 yılları arasındaki 10 yıllık dönemde istikrarlı bir şekilde artış gösteren, 2008 ve 2009 yıllarında ise gerileyen dünya çelik üretimi, 2010 yılında % 15 oranında artışla, tüm zamanların en yüksek seviyesi olan, 1,41 milyar ton seviyesine ulaşmıştır (OECD, 2011:17).

2010 yılının en büyük demir çelik üreticisi ülkeleri ve üretim miktarları Tablo 1.5’de gösterilmiştir:

Tablo 1.5. 2010 Yılı En Büyük Demir Çelik Üreticisi Ülkeler (Milyon Ton)

Sıra	Ülke	Üretim	Sıra	Ülke	Üretim
1	Çin	627	13	Meksika	17
2	Japonya	110	14	İspanya	16
3	ABD	81	15	Fransa	15
4	Hindistan	68	16	Kanada	13
5	Rusya	67	17	İran	12
6	Güney Kore	58	18	İngiltere	10
7	Almanya	44	19	Polonya	8
8	Ukrayna	33	20	Belçika	8
9	Brezilya	33	21	Güney Afrika	8
10	Türkiye	29	22	Avustralya	7
11	İtalya	26	23	Avusturya	7
12	Tayvan	20			

Kaynak: World Steel Association, World Steel in Figures, 2011, s.9.

Dünya demir - çelik üretimi, imalat yöntemleri açısından incelendiğinde, ilk sırada yer alan ülkelerin üretimlerinin büyük bir kısmının entegre tesislerde gerçekleştirildiği görülmektedir (Belen, Yıldırım ve Çakmak, 2011:26). Çin toplam üretiminin % 90,2'sini entegre tesislerde, % 9,8'ni ise elektrik ark ocaklı tesislerde üretirken bir diğer önemli üretici olan Japonya ise toplam üretiminin % 78,2'sini entegre tesislerde, % 21,8'ini ise elektrik ark ocaklı tesislerde üretmektedir.

Dünyadaki en büyük demir - çelik üreticisi firmalar Tablo 1.6'da gösterilmiştir:

Tablo 1.6. 2011 Yılı Başlıca Çelik Üreticileri (Milyon Ton)

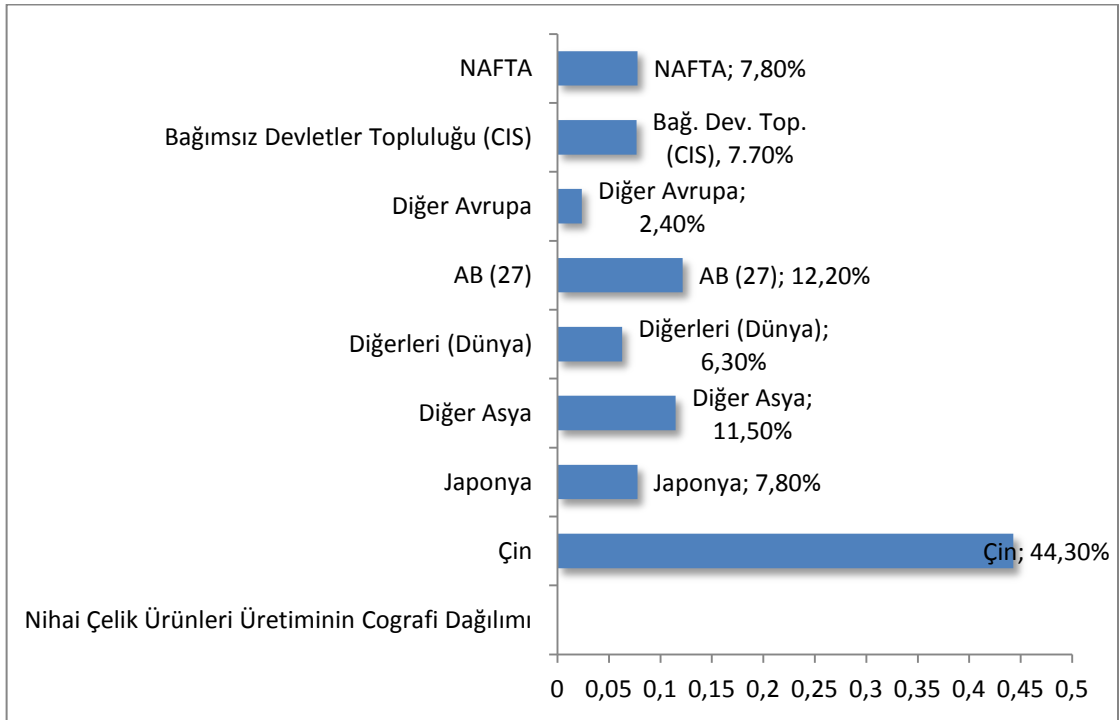
Sıra	Firma Adı	Üretim
1	ArcelorMittal	97
2	Hebei Group	44
3	Baosteel Group	43
4	POSCO	39
5	Wuhan Group	38
6	Nippon Steel	33
7	Shagang Group	32
8	Shougang Group	30
9	JFE	30
10	Ansteel Group	30

Kaynak: World Steel Association, World Steel in Figures, 2012,s.23.

Dünya Çelik Birliği'nin (World Steel Association) son verilerine göre AB'nin 27 üye ülkesi 2008 yılında toplamda yaklaşık olarak 1.33 milyar ton civarında olan küresel ham çelik üretiminin % 15'ine denk gelen miktarda (200 milyon ton) ham çelik üretmiştir. AB 2010 yılında 173 milyon ton ham çelik üreterek dünyadaki ikinci en büyük çelik üreticisi olmuştur (CPS Corporate & Public Strategy Advisory Group, 2011:6).

Çelik üretiminin coğrafik dağılımına baktığımızda 1 milyar 414 milyon ton olan dünya ham çelik üretiminin çok büyük bir kısmının (% 63,6) Asya kıtasında üretildiği görülmektedir. Asya'dan sonra en büyük pay % 14,60 ile Avrupa'ya ait olup, bu rakamın % 2'si ise Türkiye'ye aittir.

Dünya nihai çelik ürünleri üretiminin coğrafi dağılımı Şekil 1.2'de gösterilmiştir.

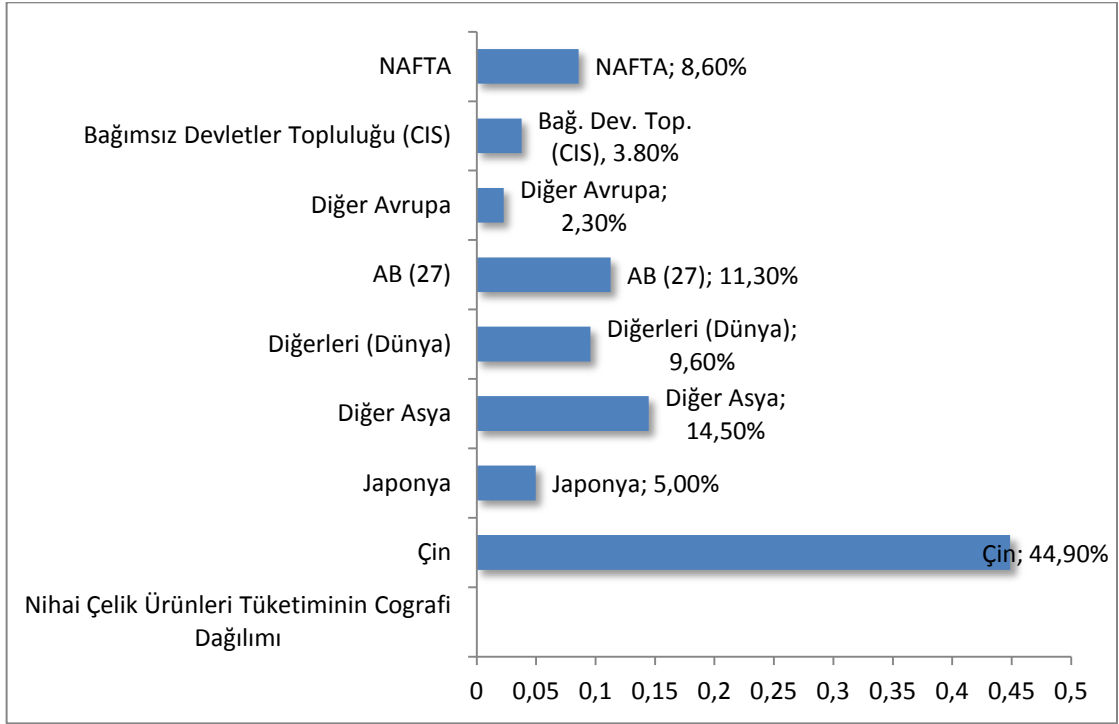


Şekil 1.2. Dünya Çelik Üretiminin Coğrafik Dağılımı

Uyarlandığı Kaynak: World Steel in Figures, World Steel Association, 2011, s.15.

Nihai çelik ürünleri tüketiminin coğrafik dağılımı incelendiğinde benzer bir grafikte karşılaşılmaktadır. 2010 yılı itibariyle 1 milyar 282 milyon tonluk dünya nihai çelik ürünleri kullanımının % 64,4'ü yine Asya ülkelerinde gerçekleştirilmektedir. Rakamlar dünya çelik endüstrisinin çekim merkezinin hem üretim hem de kullanım açısından gelişmiş ülkelere doğru kaydığını göstermektedir.

Dünya nihai çelik ürünleri tüketiminin coğrafi dağılımı Şekil 1.3'de gösterilmiştir.



Şekil 1.3. Dünya Nihai Çelik Ürünleri Tüketimin Coğrafi Dağılımı

Uyarlandığı Kaynak: World Steel Association, World Steel in Figures, 2011, s.15.

Demir çelik sektörünü ihracat ve ithalat bazında değerlendirdiğimizde ise ortaya şöyle bir tablo çıkmıştır:

2010 yılındaki en büyük çelik ihracatçısı ülkeler arasında;

- Japonya, 42,7 milyon ton ve % 11,0 pay ile ilk sırada,
- Çin Halk Cumhuriyeti 41,6 milyon ton ve % 10,8 pay ile ikinci sırada,
- Rusya 27,4 milyon ton ve % 7,1 pay ile üçüncü sırada,
- Türkiye 17,6 milyon ton ve % 4,6 pay ile 7. sırada yer almıştır.

Dünya çelik ürünleri toplam ithalat hacmi ise 2000 yılından 2009 yılına kadar sürekli bir artış göstermiştir. 2009 yılında dünya yarı ve nihai çelik ürünleri ithalatı, 2008 yılına göre % 24 oranında azalışla, 321 milyon ton olarak gerçekleşmiştir. Bu değer 2003 yılındaki 332 milyon tonluk ithalat hacminin dahi altındadır (TOBB, 2011:27).

2010 yılında, demir çelik ürünleri ithalatında, Güney Kore, 24,8 milyon ton ve % 6,6 pay ile ilk sırada, Almanya 22,7 milyon ton ve % 6,0 pay ile ikinci sırada, ABD 22,5 milyon ton ve % 6,0 pay ile üçüncü sırada yer alırken, Türkiye 11 milyon ton ve % 2,9 pay ile 9. sırada yer almıştır (World Steel Association, 2011:14).

Hemen hemen tüm ülkelerde inşaat sektörü, demir çelik sektörünün en büyük müşterisidir. Daha sonra ise otomotiv ve makine sanayi gelmektedir. Bu sektörlerin önümüzdeki yıllardaki performansları demir çelik sektörünün performansını çok ciddi şekilde etkileyecektir (Eurofer, 2011:13).

Demir-çelik tüketiminin sektörlere göre dağılımı, ülkeler bazında Tablo 1.7’de gösterilmiştir:

Tablo 1.7. Sektörlere Göre Demir - Çelik Tüketimi (%)

	İnşaat	Elektrikli Ev Aletleri	Makine Üretimi	Metal Ürünler	Otomotiv	Gemicilik
Türkiye	40,5	7,6	26,6	12,2	9,9	3,2
Almanya	31,0	5,0	15,0	11,0	33,0	5,0
İtalya	39,3	5,2	14,8	25,4	12,0	3,2
Güney Kore	45,1	7,5	8,0	14,4	13,3	11,7
Japonya	38,0	7,0	11,7	6,2	27,4	9,7

Kaynak: Eurofer Ekonomi ve Çelik Piyasalarının Görünümü 2011-2012 Raporu, Demir - Çelik Sektörü / Rekabet Gücü Raporu / 2011 Sektörel Rekabet Gücü Raporları Dizisi, Ekim 2011.

1.4.3. Avrupa’da Demir - Çelik Sektörü

Son yıllarda gelişmekte olan ülkelerde demir çelik sektörüne yapılan yatırımlar dünya çelik üreticisi ülkeleri arasındaki rekabeti fiyat açısından üst seviyelere çıkarmıştır. Fiyat açısından rekabet gücü zayıflayan Avrupa ülkeleri de pazar paylarını koruyabilmek için daha yassı ürünler, vasıflı ve paslanmaz çelik ürünleri gibi yüksek katma değerli ürünler üretip ihraç etme ve uzun ürünler gibi katma değeri düşük ürünleri diğer ülkelere ithal etme yolunu seçmişlerdir (Günay, 2008:18-24).

2011 yılının ilk çeyreğinde % 7,6 oranında gerçekleşen AB inşaat sektöründeki büyüme, ikinci çeyrekte % 2 seviyesine gerilemiştir. Ancak AB genelinde, ülke bazında

büyüme oranlarında farklılaşmalar gözlenmektedir. Polonya, Almanya, İsveç ve Fransa'da yılın ilk yarısında güçlü büyüme eğilimi devam ederken, İspanya ve Macaristan'da çift haneli daralmalar yaşanmıştır. 2011 yılının ilk çeyreğindeki % 16'lık artışın ardından, yılın ikinci çeyreğinde makine sektörünün faaliyetleri % 12 oranında büyümüştür. 2011 yılında % 4 oranında artış gösteren makine sektörüne yönelik yatırımların, 2012 yılının tamamında % 3 oranında artacağı tahmin edilmektedir. 2011 yılının ilk 8 aylık döneminde, AB genelinde otomobil satışları % 1,3 oranında azalmıştır. 2011 yılında özellikle Almanya ve Orta Avrupa ülkelerinde, otomobil sektöründe çift haneli üretim artışlarının meydana gelmesi Avrupa genelinde de ortalama % 10'luk bir artışa sebebiyet vermiştir (Eurofer, 2011:42).

İmalat sanayinin yüksek performansı sayesinde, 2011 yılının ikinci çeyreğinde, AB'nin reel çelik tüketimi % 7,6 oranında artış göstermiştir. Yılın ilk çeyreğinde reel tüketim artışının % 13,8 seviyesinde bulunması, tüketimin geçici faktörlerin etkisini kaybetmesi ile daha sürdürülebilir bir seviyeye oturmasına imkân sağlamıştır. 2012 yılının tamamında ise, reel tüketimin % 2 oranında artış göstereceği tahmin edilmektedir. 2011 yılının ilk çeyreğinde % 16,5 oranında artış gösteren AB'nin çelik tüketimi, ikinci çeyrekte % 11,4 oranında artmıştır. 2011 yılının ikinci çeyreğinde AB'nin çelik ithalatı, 2010 yılının aynı dönemine kıyasla % 45 oranında; bir önceki çeyreğe kıyasla % 20 oranında artış göstermiştir. İthalattaki artış, AB'li üreticilerin iç piyasaya sevkiyatlarının azalmasına ve ithal ürünlerin AB çelik piyasası içerisindeki payının % 21 seviyesine yükselmesine neden olmuştur (Eurofer, 2011:47).

2010 yılı itibariyle tüm Avrupa ülkelerinin dünya ham çelik üretimindeki pazar payı % 14,6'dır. Türkiye, % 2,1'lik pazar payı ile Almanya'nın ardından Avrupa'daki en büyük ikinci ham çelik üreticisi konumundadır. Almanya 2009 krizi sebebiyle yaklaşık 11 milyon ton üretim daralması yaşamıştır ve 2010 yılında kriz öncesi dönemdeki üretim miktarının altında kalmıştır. Türkiye ise kriz döneminde yaklaşık 1,5 milyon ton üretim daralması yaşamış ancak 2010 yılında hemen toparlanarak üretimini 2007 yılına göre yaklaşık 3,5 milyon ton artırmıştır. Avrupa'da Türkiye'den sonraki en büyük çelik üreticisi % 1,8'lik 2010 yılı pazar payı ile İtalya'dır (DÇÜD, Çelik Dergisi, Mayıs 2011:28).

Son veriler, demir çelik ürünlerinde ithalat baskısının yılın ikinci çeyreğinde de devam ettiğini göstermektedir. Yılın ikinci çeyreğinde, aylık ortalama 3,1 milyon ton

seviyesinde gerçekleşen üçüncü ülkelerden yapılan ithalat, son üç yıldan bu yana en yüksek seviyesine ulaşmıştır. 2010 yılının tamamında % 16 seviyesinde bulunan ithal ürünlerin AB çelik piyasasındaki payı, 2011 yılının ikinci çeyreğinde % 21 seviyesine yükselmiştir. 2011 yılının Ocak - Ağustos döneminde, özellikle yassı ürünlerin ithalatında büyük artışlar yaşanmıştır. 2011 yılının ilk sekiz aylık döneminde, toplam yassı ürün ithalatında % 51, toplam uzun ürün ithalatında % 21 oranında artış yaşanmıştır. Uzun ürünlerde, filmaşın ve nervürlü demir ithalatında düşüş yaşanırken, ticari çubuk ithalatında büyük bir artış gözlenmiştir. En fazla çelik ürünleri ithalatının yapıldığı ülke ise Çin Halk Cumhuriyeti olmuştur. Diğer taraftan, Türkiye'nin AB'ye yönelik yassı ürün ihracatında da ciddi bir artış yaşanmış ve Türkiye AB'ye en fazla sıcak haddelenmiş geniş şerit ihraç eden ülke olmuştur (Eurofer, 2011:54).

AB demir - çelik sektörü imalat yöntemleri açısından değerlendirildiğinde ise, 2010 yılında AB'de gerçekleşen ham çelik üretiminin % 57,7'si entegre tesislerde, % 41,9'unun ise elektrik ark ocaklı tesislerde üretildiği görülmektedir. Bu oranlar AB'nin, demir çelik sektöründe yüksek katma değere sahip yassı ürünler ve vasıflı çelik ürünlerinin üretiminde yoğunlaştığını göstermektedir.

Avrupa ülkelerindeki ham çelik üretim miktarları Tablo 1.8'de gösterilmiştir:

Tablo 1.8. Avrupa Ülkelerindeki Ham Çelik Üretimi (Milyon Ton)

	2008	2009	2010	2011 (8 aylık)
Avusturya	7	6	7	5
Belçika	11	6	8	6
Bulgaristan	1	1	1	1
Çek Cumhuriyeti	6	4	5	4
Finlandiya	4	3	4	3
Fransa	17	12	15	11
Almanya	44	32	42	31
Yunanistan	3	2	2	1
Macaristan	2	1	2	1
İtalya	29	19	25	19
Lüksemburg	3	2	3	2
Hollanda	7	5	7	5
Polonya	9	7	8	6
Romanya	5	3	4	3
Slovakya	5	4	5	3
Slovenya	1	1	1	1
İspanya	18	14	16	11
İsveç	5	3	4	3
Birleşik Krallık	13	10	10	7
A.B. (diğer) (27)	2	2	2	1
Avrupa Birliği (27)	191	135	167	121
Türkiye	27	25	29	22
Diğerleri	3	2	3	2
Diğer Avrupa	30	27	32	24
Avrupa (Toplam)	221	162	199	145

Kaynak: World Steel in Figures, World Steel Association, 2011, s.14.

1.4.4. Türkiye’de Demir - Çelik Sektörü

Sektörce kullanılan hammaddelerin yurt içi kaynaklardan sağlanabilecek olması demir çelik sanayisinin Türkiye’de kurulması düşüncesini doğurmuş ve Cumhuriyet’in ilk yıllarında uygulamaya konulan sanayi planında demir çelik tesislerinin kurulmasına karar verilmiştir (BYDK, Demir – Çelik Fabrikaları Raporu, 1995:1).

Ülkemizde ilk demir çelik sanayisi kuruluş çalışmalarına 1932 yılında Kırıkkale’de savunma sanayisinin çelik ihtiyacını karşılamak amacıyla, Askeri Fabrikalar Müdürlüğü’ne bağlı fabrikanın kurulmasıyla başlanmıştır. Bu fabrikada her türlü takım çelikleri, makine yapı çelikleri ve az miktarda da olsa inşaat demirinin üretilmesiyle birlikte demir - çelik sanayisinin temel alt yapısı da oluşturulmaya başlanmıştır.

Birinci Dünya Savaşı ve Kurtuluş Savaşı sonrasında ulusal bir demir - çelik sanayisine duyulan ihtiyacın sonucu olarak 1935 yılında Sümerbank'a bağlı Karabük Demir Çelik Fabrikaları kurulmuştur. Türkiye Cumhuriyeti'nde demir cevheri üretimi, Karabük Demir Çelik Fabrikası'nın kurulmasıyla başlamıştır. Karabük Demir Çelik Fabrikası, maden kömürü havzasına yakın oluşu, demiryolu güzergâhına yakın olması ve jeopolitik bakımdan elverişli durumda bulunması nedeniyle, 3 Nisan 1937'de Karabük'te kurulmuş ve işletme üniteleri, 1 Haziran 1939 yılından itibaren, 150.000 ton çelik üretim kapasitesi ile faaliyete geçmiştir. 1955 yılında kurulan Türkiye Demir Çelik İşletmeleri (TDÇİ) Genel Müdürlüğü'ne Divriği madenleri de bağlanarak Karabük Demir Çelik Fabrikası'nın gelişmesi sağlanmıştır (TOBB, 2011:11).

Karabük Demir Çelik Fabrikaları, Türkiye'nin uzun ürün talebini karşılamak üzere ilk entegre demir çelik tesisi olarak faaliyetlerine başlarken, yassı ürün talebini karşılamak için de ikinci entegre tesis olan Ereğli Demir Çelik Fabrikası 1965 yılında işletmeye açılmıştır. 1977 yılında, uzun ürün ve yarı mamul talebini karşılayabilmek amacıyla da, Türkiye'nin üçüncü entegre tesisi olan İskenderun Demir Çelik Fabrikaları üretime başlamıştır (BYDK, Demir – Çelik Fabrikaları Raporu, 1995:1).

1960'lı yıllardan itibaren özel sektöre ait elektrik ark ocaklı tesislerin de faaliyete geçmeye başlaması ve 1970'li yıllarda İSDEMİR' in ve özel sektöre ait 5 ark ocaklı kuruluşun işletmeye açılması ile 1980 yılında Türk demir çelik endüstrisi, yıllık 4.2 milyon ton ham çelik üretim kapasitesine ulaşmıştır. 1980'li yıllarda yaşanan ekonomideki liberalleşme hareketleri, sadece Türk ekonomisi açısından değil demir çelik endüstrisinin gelişimi açısından da bir dönüm noktası olmuştur. Türk demir çelik sektörü 1980'li yıllarda yeni elektrik ark ocaklı tesislerin kurulmasına ve ekonomik yapıdaki gelişmelere paralel olarak büyük bir gelişme göstermiştir.

Türkiye'nin ilk yassı çelik üreticisi olan Erdemir, 500 000 ton ham çelik ve 400 000 ton yassı mamul kapasitesiyle 1965 yılında üretim faaliyetlerine başlamıştır. Türkiye'nin en büyük entegre demir – çelik tesisine sahip olan kuruluş, 2002 senesinde İsdemir'i de bünyesine katarak faaliyetlerine devam etmektedir (Çelik Dergisi, Ağustos, 2012:4).

“Divriği demir yataklarının Maden Tetkik Arama Enstitüsü (MTA) tarafından 1937 yılında bulunması ve 1938 yılından itibaren üretime geçmesiyle birlikte demir cevheri üretimi demir ve çelik tesislerinin gereksinimine paralel olarak artmış,

günümüze kadar bu tesislerin hammadde gereksinimlerinin büyük bir bölümünü karşılamıştır. Türkiye’de bugüne kadar yaklaşık 900 adet demir oluşumu saptanmış, bunlardan ekonomik olabileceği düşünülen 500 kadarının etüdü yapılmıştır. Ülkemizde entegre demir çelik tesislerinde kullanılacak özellikteki demir cevheri rezervleri Sivas-Erzincan, Kayseri-Adana, Malatya, Kırşehir-Ankara ve Balıkesir bölgelerinde yer almaktadır” (TOBB Demir - Çelik Raporu, 2011:13).

Türkiye’nin bilinen madenlerindeki demir cevheri rezerv miktarı yaklaşık 113 milyon tondur (MTA, 2009:7). Sektörün ihtiyacını karşılayabilecek kaliteli cevher miktarı sınırlı olduğu için zenginleştirme tesislerine ihtiyaç duyulmaktadır. Mevcut cevherlerin üretimde ithal cevher ile karıştırılarak kısıtlı oranda kullanımı mümkün olmaktadır (Metal Sektörü Yeterlilikler Kılavuzu 2011:37).

Türk demir çelik sektöründe 3 çeşit üretici mevcuttur. Bunlardan birincisi demir cevheri ve kok kömürü kullanarak üretim yapan entegre tesisler, ikincisi hurdadan üretim yapan ark ocaklı tesisler, üçüncüsü ise diğer üreticilerden kütük almak ya da ithal etmek suretiyle üretim yapan bağımsız haddehanelerdir.

Türkiye’de ham çelikten nihai mamul üreten üreticiler Marmara, Ege, Akdeniz, Karadeniz ve İç Anadolu bölgesinde faaliyet göstermekte olup, üreticilerin çoğunluğu Marmara, Ege, Akdeniz sahil şeridinde yer almaktadır. Demir çelik sektöründe yaklaşık 150’ye yakın firma faaliyet göstermektedir. Bunların içerisinde kapasiteleri 50.000 ton ile 3.500.000 ton arasında değişen elektrik ark ocaklı tesisler ile toplam kapasiteleri 8.500.000 ton olan entegre tesisler bulunmaktadır. (Sanayi Genel Müdürlüğü, 2011: 6).

Türkiye’nin üretim yöntemi bazında ham çelik üretim miktarı Tablo 1.9’da gösterilmiştir:

Tablo 1.9. Türkiye'nin Üretim Yöntemi Bazında Ham Çelik Üretimi (Milyon Ton)

	2000	2005	2008	2009
Entegre Tesisler	5	6	7	8
Erdemir	2	3	3	4
İsdemir	2	2	3	3
Kardemir	1	1	1	1
Ark Ocaklı Tesisler	9	15	20	18
TOPLAM	14	20	27	26

Kaynak: Demir Çelik Üreticileri Derneği, 2010.

Çelik üretim tesislerinin bölgelere göre dağılımı Tablo 1.10'da gösterilmiştir:

Tablo 1.10. Çelik Üretim Tesislerinin Bölgelere Göre Dağılımı

Bölge	Adet
Akdeniz	9
Marmara	8
Ege	7
İç Anadolu	2
Karadeniz	3
Toplam	29

Kaynak: TOBB, 2011

“Türkiye 1939 yılında 150.000 ton/yıl ile başlayan demir çelik üretimini 1999 yılında 14 milyon ton/yıl üretime taşıyarak büyük bir kalkınma sağlamıştır. Bu süreç içinde birçok sanayi kolunun gelişmesine, inşaat sektörünün bugünkü noktasına gelmesine, 40.000'e yaklaşan insanımıza iş olanağı sağlamasıyla özellikle ağır sanayinin ihtiyaç duyduğu nitelikli eleman yetiştirilmesine, demir madenciliğinin gelişmesine öncülük etmiş ve bu bağlamda demir çelik sektörü Türkiye'de dev bir sektör haline gelmiştir” (TOBB Demir Çelik Raporu, 2011:27).

Demir- çelik sektörünün bu hızlı yükselişi ülkenin işgücünü de arttırarak, insanlara yeni iş kapıları açmış ve istihdamı arttırmıştır.

Türkiye’de demir çelik sektöründe istihdam değerleri Tablo 1.11’de gösterilmiştir:

Tablo 1.11. Türkiye’de Demir Çelik Sektöründe İstihdam Değerleri (Kişi Sayısı)

	1990	2000	2005	2006	2007	2008	2009	2010
EO	10 525	9 239	13 153	14 505	16 187	16 205	14 834	17 216
BOF	33 145	17 459	17 293	16 264	17 328	16 831	14 948	16 170
Toplam	43 670	26 698	30 446	30 769	33 515	33 036	29 782	33 386

Kaynak: TOBB, 2011

Türk demir çelik sektörü, 2011 yılında 34 milyon ton ham çelik üretimi ve 17 milyon dolar ihracatı ile endüstriyel sektörlerdeki toplam istihdamın % 1’ine, sektörlerin toplam ihracatının % 10’una sahip bulunmaktadır. Bu yönü ile sektör en çok ihracat yapan sektörler arasında yer almaktadır. Sektörün en önemli sorunlarından biri ağırlıklı olarak ithal girdiyle çalışması olup, Elektrik Ark Ocaklı (EAO) gibi kuruluşlarda hammadde olarak kullanılan hurdanın % 70 civarındaki bölümü ithal edilmektedir. 2011 yılında 9,8 milyar dolarlık hurda ithal edilmiş ve bu ithalatın büyük bir kısmı ABD, Rusya, Ukrayna ve AB ülkelerinden yapılmıştır. Entegre tesislerde ise, hammadde olarak 1,1 milyar dolarlık (4 milyon ton) taş kömürü ve 1,2 milyar dolarlık demir cevheri ithal edilmiştir (T.C. Bilim, Sanayi Ve Teknoloji Bakanlığı Demir Çelik Sektörü Raporu, 2012).

2010 yılı itibariyle Türkiye’nin ham çelik üretim kapasitesi 48,7 milyon ton civarındadır. 2011 yılında dünya çelik sektöründeki büyüme, kriz etkisinden çıkılan 2010 yılına kıyasla önemli ölçüde yavaşlayarak % 6,8 seviyesinde gerçekleşmiştir. Dünya çelik üretimindeki büyümenin 2010 yılına kıyasla % 15’ten % 6,8’e gerilediği 2011 yılında, Türk çelik sektörü, önceki yıla kıyasla iki puanlık bir artışla % 17 oranında büyümüştür. Bu büyüme oranı ile Türkiye, 2011 yılında çelik üretimini en fazla arttıran ülke olmuştur. Bu yönüyle Türkiye, Çin, Hindistan, Güney Kore’nin ardından, kriz öncesi üretim seviyesinin üstüne çıkan dördüncü ülke olmuştur. Ayrıca 2011 yılında Türkiye’nin kütük üretimi % 1,8 oranında artışla, 24,4 milyon ton seviyesine yükselmiştir (Demir Çelik Dergisi Mart Eki, 2012:3).

Türkiye’de ham çelik üretiminin % 73,6’sı elektrikli ark ocaklarında (EAO) gerçekleştirilirken; % 26,4’ü entegre tesislerde gerçekleştirilmektedir. 2010 yılında 29,1

milyon ton ile bu kapasitenin % 59,7'si kullanılmıştır. 2010 yılı itibariyle nihai mamul üretimi 26,3 milyon ton iken; nihai mamul tüketimi 23,6 milyon ton olarak gerçekleşmiştir. Miktar olarak demir-çelik ürünlerinin ihracatı, 17,6 milyon ton olarak gerçekleşmiş olup bu rakamın % 52'sini uzun ürünler oluşturmaktadır. Miktar olarak demir-çelik ürünlerinin ithalatı ise 11 milyon ton civarında gerçekleşmiş olup bu rakamın % 62'sini yassı ürünler oluşturmuştur (TOBB, 2011).

Yeni tesis yatırımlarının elektrikli ark ocaklı tesislere yoğunlaşması sebebiyle, 2011 yılında üretimde yaşanan toplam 4,96 milyon tonluk artışın yüzde 88'i elektrikli ark ocaklı tesislerde (EAO) gerçekleşmiştir (Demir Çelik Dergisi Mart Eki, 2012:3).

2013 yılından itibaren 57 milyon ton seviyelerine ulaşması beklenen ham çelik üretim kapasitesinin, 43,5 milyon tonluk bölümünün elektrikli ark ocaklarında, 13,5 milyon tonluk bölümünün ise entegre tesislerde gerçekleştirilmesi beklenmektedir. Üretim, entegre tesislerde demir cevheri, elektrik ark ocakları ise hurda tabanlıdır. Bununla birlikte, entegre tesislerde de düşük miktarda olmakla birlikte hurda tüketilmektedir. Türkiye dünyadaki en büyük hurda ithalatçısı olan ülkedir. Demir çelik üretimimiz, üç adet entegre tesis (Erdemir, İsdemir, Kardemir) ve 26 adet elektrik ark ocaklı tesis olmak üzere 29 tesis tarafından gerçekleştirilmektedir (Metal Sektörü Yeterlilikler Kılavuzu, 2011:15).

Çelik üretiminin başlıca girdi kalemleri olan hurda, demir cevheri ve kömür üretimlerinin ülkemizde yetersiz olması, bu ürünlerde yurtdışına bağımlılığın süreceğini göstermektedir. 2010 yılında elektrik ark ocaklı tesislerin üretimlerinin artmasıyla, hurda ithalatı % 22 oranında artışla, 19,2 milyon tona ulaşmıştır. Türkiye'nin, Dünya hurda ithalatındaki payı ise % 18 olarak tahmin edilmektedir (World Steel Association, 2011).

Türkiye, AB'ye, başlıca olarak, yarı bitmiş veya uzun ürünler ihraç etmektedir. Buna rağmen, Türkiye AB'den yüksek değere sahip yassı ve özel çelik ürünleri ithal etmektedir. Türkiye'nin en fazla çelik ihracatı yaptığı üçüncü yer Avrupa Birliği olup, Türkiye 2010 yılında Avrupa Birliği'ne 17.3 milyon ton çelik ihraç etmiştir (CPS Corporate & Public Strategy Advisory Group, 2011).

Türkiye, 40 yıl önce 1.170.000 ton ile dünya çelik üretiminde 38. sırada, yirmi yıl önce 20. sırada iken, 2005 yılında 11. sıraya, 2010 ve 2011 yıllarında onuncu sıraya yükselerek, yirmi yılda üretimini üç katın üzerinde artıran üçüncü ülke olmuştur.

Türkiye'nin coğrafik konumu itibarıyla küresel pazarlarda önemli bir rekabet avantajına sahip olması nedeniyle Türk demir çelik sektörünün dünya üretim sıralamasında daha ön sıralarda yer alma potansiyeli bulunmaktadır (Bakka, 2011)

Türk demir çelik sektörünün yıllara göre ham çelik üretimi Tablo 1.12'de gösterilmiştir:

Tablo 1.12. Türk Demir - Çelik Sektörünün Yıllara Göre Ham Çelik Üretimi (Milyon Ton)

Yıllar	Üretim Miktarı
1991	9
1995	13
2000	14
2005	21
2007	26
2008	27
2009	25
2010	29
2011*	16

Kaynak: World Steel Assocoiation, 2011. (*2011 verileri altı aylıktır)

2011 yılında üretimini % 17,2 oranında artıran sektör, bu dönemde de dünya çelik üretiminde ilk 10 üretici arasına girmiştir. Kütük üretimi miktar bazında % 11 oranında artışla, 22,1 milyon tona, slab üretimi ise, % 36 oranında artışla 8,9 milyon tona yükselmiştir. Nihai mamul üretiminde, 2011 yılında, Türkiye üretimini toplam % 21,5 oranında artışla, 2010 yılındaki 26.30 milyon tondan, 31,9 milyon tona yükseltmiştir. Yeni kapasitelerin de katkısıyla, en yüksek üretim artışı % 36,9 oranında artışla, 6.63 milyon tondan, 9.08 milyon tona ulaşan yassı ürünlerde gözlenmiştir. Aynı dönemde, uzun ürün üretimi ise, % 16,3 oranında artışla, 19.67 milyon tondan, 22.87 milyon tona ulaşmıştır. 2011 yılında, 31.94 milyon tonluk toplam nihai çelik ürünleri üretiminin % 71,6 oranındaki kısmı uzun ürünlerden, % 28,4 oranındaki kısmı yassı ürünlerden oluşmuştur. 2011 yılında elde edilen toplam 5.64 milyon tonluk üretim artışının, % 57 oranındaki kısmı uzun ürünlerde, % 43 oranındaki kısmı yassı ürünlerde

gözlenmektedir (DÇÜD, T.C. Bilim, Sanayi Ve Teknoloji Bakanlığı Demir Çelik Sektörü Raporu, 2012:22)

Türkiye'nin ürün bazında ham çelik üretimi Tablo 1.13'de gösterilmiştir:

Tablo 1.13. Türkiye'nin Ürün Bazında Ham Çelik Üretimi(Milyon Ton)

	2000	2005	2008	2009	Değişim (%)	Pay (%)
Uzun Ürün	12	18	23	21	-9,3	81,2
Yassı Ürün	2	3	4	5	14,6	18,8
TOPLAM	14	21	27	26	-5,6	100

Kaynak: Demir Çelik Üreticileri Derneği, 2010

Türkiye'nin yıllar itibariyle ham çelik kapasitesi ve kullanım oranları Tablo 1.14.'te gösterilmiştir:

Tablo 1.14. Türkiye'nin Ham Çelik Kapasitesi (Milyon Ton) ve Kapasite Kullanım Oranları (%)

	2000		2005		2008		2009	
	Kapasite	KKO	Kapasite	KKO	Kapasite	KKO	Kapasite	KKO
Entegre Tesisler	6	84,3	6	97,1	7	88,0	9	89,0
Erdemir	3	79,6	3	103,2	3	81,0	4	106,0
İsdemir	2	89,3	2	93,4	3	86,0	4	79,0
Kardemir	1	87,5	1	87,9	1	88,0	1	73,0
Ark Ocaklı Tesisler	14	66,7	19	78,0	26	76,0	30	59,0
TOPLAM	29	72,2	25	82,8	33	79,0	39	66,0

Kaynak: Demir Çelik Üreticileri Derneği, 2010

1980 yılında 4,2 milyon ton olan Türkiye'nin ham çelik üretim kapasitesi, 1990 yılında 11,3 milyon ton, 2000 yılında 19,8 milyon ton seviyelerine ulaşmıştır. Türkiye'nin toplam ergitme kapasitesi yıllardır artarak 2010 yılında 42,7 milyon tona ulaşmıştır. Bu artışın gelecek yıllarda devam etmesi ve 2015 yılında Türkiye'nin toplam

ergitme kapasitesinin 55 milyon tona ulaşması beklenmektedir. Benzer şekilde 2010 yılında 48,7 milyon ton olan toplam kütük ve slab üretim kapasitesinin yeni yatırımların devreye girmesiyle 2015 yılında 62 milyon tona ulaşması beklenmektedir (Türkiye Demir Çelik Üreticileri Derneği, 2011).

2011 yılının ilk sekiz ayında kaydedilen rakamlara göre hem üretimde hem de tüketimde 2010 yılına göre ciddi artışlar gerçekleşmiştir. Bu rakamlar Türkiye demir çelik sektörünün kriz öncesinde yakaladığı büyüme ivmesine 2011 yılında tekrar kavuştuğunun çok kuvvetli sinyalleri olarak algılanabilir. 2005 yılından bu yana uzun ürünlerde her sene üretim fazlası olmasına rağmen yassı ürünlerde sürekli bir üretim açığı bulunmaktadır. 2010 yılında devreye giren yeni yatırımlar ile üretim 6.6 milyon tona çıkmıştır (Eurofer, 2011:28).

Türkiye'nin toplam çelik üretim verileri Tablo 1.15'de gösterilmiştir:

Tablo 1.15. Türkiye'nin Toplam Çelik Üretimi (Milyon Ton)

	2005	2006	2007	2008	2009	2010	2011 (ilk 8 ay)
Uzun	15.8	19.2	21.8	22.1	20.7	19.7	14.7
Yassı	3.8	4.1	4.3	4.5	4.4	6.6	5.9
Toplam	19.6	23.4	26.1	26.7	25.1	26.3	20.6

Kaynak: DÇÜD, 2012.

Türkiye'nin toplam çelik tüketim verileri Tablo 1.16'da gösterilmiştir:

Tablo 1.16. Türkiye'nin Toplam Çelik Tüketimi (Milyon Ton).

	2005	2006	2007	2008	2009	2010	2011 (ilk 8 ay)
Uzun	9.1	10.5	12.1	10.3	9.7	11.6	8.9
Yassı	9.4	10.7	11.7	11.2	8.3	11.9	8.4
Toplam	18.5	21.2	23.8	21.5	18.0	23.6	17.3

Kaynak: DÇÜD, 2012.

Türkiye'nin çelikhane kapasitesi ve kapasite kullanım oranları Tablo 1.17'de gösterilmiştir:

Tablo 1.17. Türkiye'nin Çelikhane Kapasitesi (Ton) ve Kapasite Kullanım Oranları (%)

Firmalar	2000			2010		
	Kapasite	Üretim	KKO	Kapasite	Üretim	KKO
Asil Çelik	260.000	200.148	77	485.000	287.463	59
Çebitaş	700.000	417.160	60	750.000	168.800	23
Çemtaş	172.000	133.587	78	172.000	141.220	82
Çolakoğlu	1.522.000	1.570.053	103	3.000.000	2.303.745	77
Çukuroca	1.775.000	439.167	25	0	0	
Diler	906.000	262.794	29	1.500.000	1.301.266	87
Ede	0	0	0	780.000	85.084	11
Ege Çelik	0	0	0	2.000.000	613.686	31
Ege Metal	840.000	559.387	67	0	0	0
Ekinciler	1.000.000	404.227	40	1.000.000	684.431	68
Habaş	1.530.000	1.324.024	87	4.800.000	2.726.629	57
İçdaş	1.800.000	1.384.678	77	5.267.600	3.612.700	69
İlhan (Platinum)	0	0	0	200.000	56.696	28
İzmir D.Ç	850.000	742.548	87	1.500.000	1.095.584	73
Kaptan	-	-	-	1.350.000	1.073.799	80
Kroman	1.100.000	626.023	57	1.350.000	1.084.000	80
MKEK	60.000	5.763	10	60.000	3.531	6
Nursan	0	0	0	1.200.000	931.559	78
Özkan *	0	0	-	700.000	140.770	20
Sider	0	0	0	720.000	686.939	95
Sivas D.Ç.	0	0	-	550.000	432.431	79
Tosçelik	0	0	0	2.000.000	1.320.000	66
Yazıcı	817.000	824.271	101	1.000.000	984.712	98
Yeşil Yurt	300.000	202.367	67	1.000.000	426.242	43
Yolbulan-Baştuğ *	0	0	0	1.500.000	638.620	43
Diğer EAO**	0	0	0	500.000	105.000	21
EAO	13.632.000	9.096.197	67	33.384.600	20.904.907	66
Erdemir	3.000.000	2.388.009	80	3.850.000	3.538.898	92
İsdemir	2.200.000	1.965.100	89	4.000.000	3.564.495	89
Kardemir	1.000.000	875.429	88	1.500.000	1.134.566	76
BOF	6.200.000	5.228.538	84	9.350.000	8.237.959	88
Toplam	19.832.000	4.324.735	72	42.734.600	29.142.866	71

Kaynak: Türkiye Demir ve Demir Dışı Metaller Meclisi Sektör Raporu 2010, TOBB, 2011, s.6.

*Özkan, Yolbulan Baştuğ ve Diğer EO (Elektrikli Ocaklar) grubu, kapasite ve üretim toplamları içerisinde yer almasına rağmen, 2010 yılında deneme üretimi yaptıklarından, KKO (Kapasite Kullanım Oranı) hesaplamasına dâhil edilmemiştir.

** EO: EAO (Elektrikli Ark Ocakları) ve İO'ları (İndüksiyon Ocakları) kapsamaktadır.

Türkiye'nin nihai ürün üretiminin büyük bölümü (% 75) uzun ürün üretimidir. Dolayısıyla en büyük ihracat kalekimiz % 52'lik pay ile uzun ürünlerdir.

Demir - çelik ürünleri içinde en önemli ihracat kalemi olan uzun ürünlerde AB ülkeleri 2007 yılında % 33'lük ihracat oranı ile önemli bir müşteri iken, 2010 yılında uzun ürün ihracatımızın sadece % 6'sı AB ülkelerine yapılmıştır. Miktar olarak azalış 3 milyon tonun üzerindedir. Uzun ürün açısından Türkiye'nin en önemli müşterileri yıllardan beri ortalama % 50'lik payla Orta Doğu ve Körfez ülkeleridir (Çelik Dergisi 2010 yılı Raporu, 2010:11).

2. VERİ MADENCİLİĞİ

2.1. VERİ MADENCİLİĞİNİN TANIMI

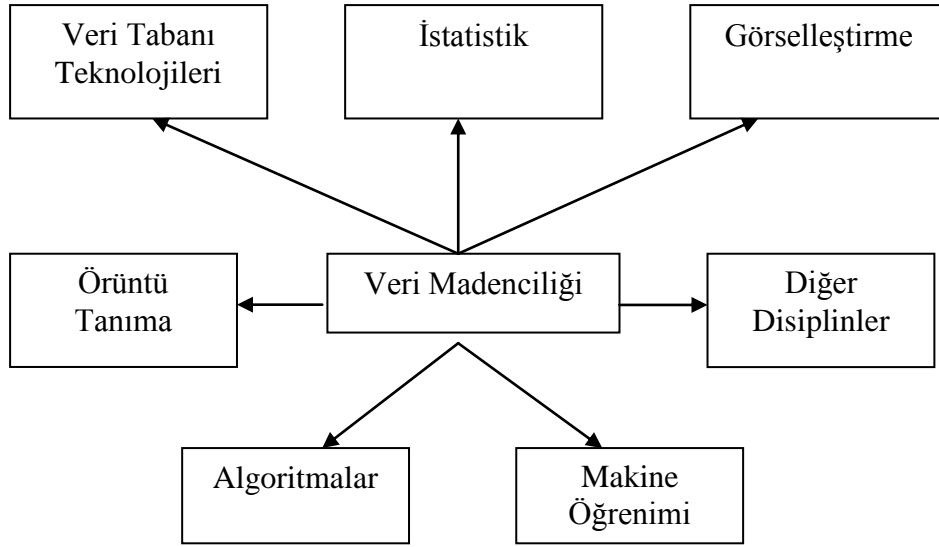
Veri madenciliği; istatistik, veri tabanı teknolojisi, makine öğrenimi, yapay zekâ ve görselleştirmenin kullanıldığı disiplinler arası bir çalışmadır.

Günümüzde etrafımızda oldukça fazla veri bulunmakta olup bu verilerin farklı tekniklerle işlenip anlamlı hale getirilmesi gerekir. İşte veri madenciliği tam olarak bu noktada devreye girer. Barkot sistemlerindeki gelişmeler ile veri tabanı sistemlerinin gelişmesi alışveriş verilerinin kolaylıkla elde edilmesine olanak tanımakta, hastanede veya bankada yapılan her işlemin kolayca saklanabilmesini sağlamaktadır. Yapılan telefon görüşmeleri kaydedilerek müşterilerin kimlerle ne kadar süre konuştukları tespit edilebilmekte, GSM şirketlerinin müşterilerine yönelik kampanyalar yapabilmesine olanak sağlamaktadır (Gürsoy, 2009:25).

Veri madenciliği büyük veri tabanlarından önceden bilinmeyen ancak potansiyel olarak kullanılacak bilgilerin çıkarılmasıdır. Veri madenciliği, doğrulama amaçlı veri madenciliği ve keşif amaçlı veri madenciliği olmak üzere ikiye ayrılır. Doğrulama amaçlı veri madenciliği, öne sürülen bir hipotezin geçerlilik kazanabilmesi için bilgi çıkarma sürecidir. İstatistik ve çok boyutlu analiz gibi teknikleri içerir. Keşif amaçlı veri madenciliği ise kuralların otomatik olarak çıkarılması için kümeleme, ortaklık keşfi ve öğreticili tümevarım gibi araçları kullanır (Sumathi ve Sivanandam, 2006:189).

Gelişmiş veri analiz sistemleri altında yer alan veri madenciliği; kümeleme, sınıflama, birliktelik analizleri ve OLAP (Online Analytical Processing – Çevrimiçi Analitik İşleme) vb. gibi teknikleri kullanır. Veri madenciliğinin ileri seviye uygulamaları ise görsel veri madenciliği, web madenciliği, metin madenciliği ve bio madenciliktir (Cihan, 2009:24).

Veri madenciliği ile ilişkili alanlar Şekil 2.1’de gösterilmiştir:



Şekil 2.1. Veri Madenciliği ile İlişkili Alanlar

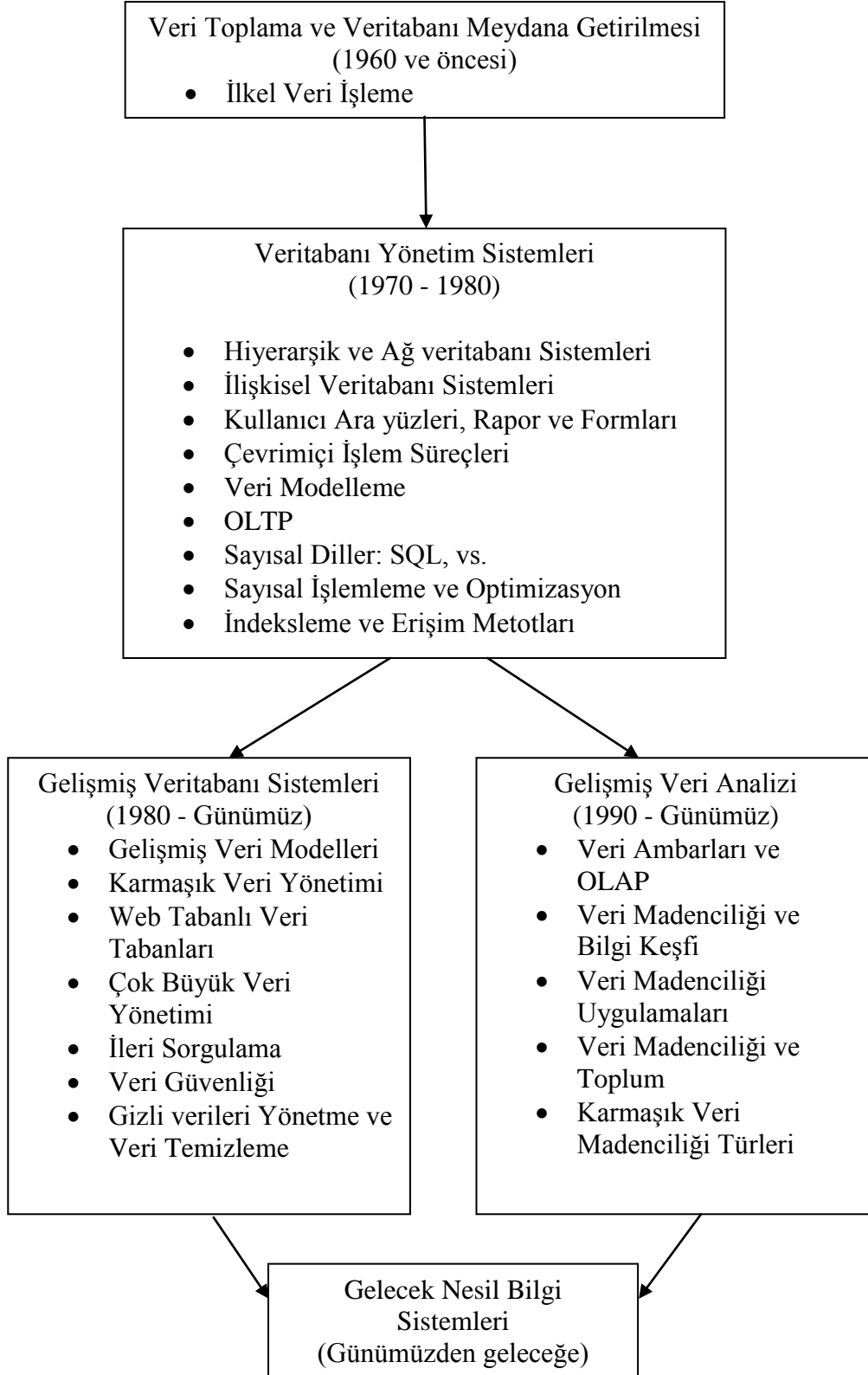
2.2. VERİ MADENCİLİĞİ GELİŞİM SÜRECİ

Veri madenciliği esas olarak üç temel alan üzerinde gelişmiştir. Bu alanlardan en önemlisi ve en eskiye dayananı klasik istatistik bilimidir. Klasik istatistiksel yöntemler bugün kullanılmakta olan veri madenciliği araç ve metotlarının esasını oluşturur (Çetin, 2009:6).

Veri madenciliğinin üzerinde olgunlaştığı bir diğer alan ise yapay zekâdır. Yapay zekâ, sezgisel yaklaşımı temel alarak, insan gibi düşünebilme prensibiyle, istatistikten farklı metotlarla, problemlerin çözümlerinde kullanılır (Coşkun, 2010.9).

Üçüncü alan ise yine köklerini istatistik ve yapay zekâdan alan makine öğrenmesidir. Aslında makine öğrenme, yapay zekânın sezgisel yöntemleri ileri düzey istatistiksel yöntemlerle harmanlayıp evrimleşerek geliştiği ileri düzey halidir. Makine öğrenme, uygulandığı bilgisayar sistemlerinde, istatistiksel ve yapay zekâ algoritmaları kullanarak eldeki verinin değerlendirilmesine, bu verilerden sonuçlar çıkarılmasına ve bu sonuçlara bakılarak kararlar alınmasına olanak sağlar (Coşkun, 2010.9).

Veritabanı sistem teknolojisi gelişim süreci Şekil 2.2’de gösterilmiştir:



Şekil 2.2. Veritabanı Sistem Teknolojisi Gelişim Süreci

Uyarlandığı Kaynak: HAN, J., KAMBER, M., PEI, J., “Data Mining Concepts and Techniques”, **Morgan Kaufmann**, USA, 2012, s.3.

2.3. VERİ MADENCİLİĞİNİN KULLANIM ALANLARI

Veri madenciliğinin en önemli özelliği çok büyük miktarlardaki verilerin istenildiği gibi işlenebilmesine imkân sağlamasıdır. İşletmeler yoğun rekabet içerisinde bir adım öne geçebilmek için bilgileri etkin ve verimli bir şekilde kullanmak ister. Bu noktada veri madenciliği çok büyük önem kazanmaktadır. Veri madenciliği bankacılık, sigortacılık, elektronik ticaret, sağlık, iletişim, ulaştırma, savunma, dolandırıcılık tespiti ve eğitim gibi pek çok alanda kullanılmaktadır (Özkan, 2010:216, Meyer ve Cannon, 1998:186).

Verilerin toplanması ve saklanmasıdaki kolaylık, veritabanı yönetim sistemlerindeki teknolojik gelişmeler, bilgi işlem maliyetlerindeki meydana gelen düşüşler ile birlikte kullanılacak analitik araçların artması da veri madenciliği uygulamalarına olan ilgiyi arttırmaktadır (Park, Selwyn ve Shaw, 2001:205-222).

Veri madenciliği, çeşitli sektörlerde çok geniş bir kullanım alanına sahiptir. Veri madenciliği yöntemlerinin kullanıldığı sektörler aşağıda gösterilmiştir (Chen, Sakaguchi ve Frolick, 2006:1-6):

- Kredi kartı şirketleri: Kredi kartı şirketleri, kredi kartı başvurularını onaylamak, kart sahiplerinin satın alma davranışlarını çözümlmek ve dolandırıcılık durumunu algılamak için veri madenciliğini kullanır.
- Perakendeciler: Müşterilerin satın alma alışkanlıkları ve tercihlerini anlama perakendecilerin esas stratejisidir. Veri madenciliği, perakendecilere bu bilgiyi sağlayabilir. Perakende ortamında etkili veri madenciliği uygulamaları olarak, Pazar Sepet Analizi (MBA) ya da Alışveriş Sepeti Analizi (SBA) yaygın olarak kullanılır. Elektronik Satış Noktasına (EPOS) gelen müşterilerin alışveriş sepeti özelliklerinin analiz edilmesiyle birlikte etkili promosyon ve reklam için bulgular elde edilebilir. Bazı mağazalarda EPOS verileri kullanılarak kasiyer davranışını algılamaya yönelik veri madenciliği teknikleri de uygulanmaktadır.
- Finansal Servisler: Güvenlik analistleri yatırım stratejilerini geliştirmek, ticaret ve risk modelleri inşa etmek, devasa finansal verileri analiz etmek için veri madenciliğini yoğun bir şekilde kullanırlar. Finans sektöründeki şirketlerin bazı ürünleri veri madenciliği ile test edilmiş ve olumlu sonuçlar alınmıştır. Bu

ürünler döviz ticareti, otomatik poliçe, hisse senedi seçimi, kredi puanlama, tanımlama dolandırıcılık desenleri ve ipotek tarama ve diğerleridir.

- Bankalar: Veri madenciliği bankacılık alanında büyük potansiyel göstermiş olmasına rağmen, henüz çok erken bir aşamadır. Bankaların sadece yüzde 11'i veri madenciliği faaliyetlerini kolaylaştırmak için işlevsel bir veri ambarına sahiptir.
- Havayolları: Havayolları, yeni uçuş hatları belirlemek, hizmet kalitesi ve standartlarını yükseltmek gibi stratejik kararlar alabilmek adına veri madenciliği tekniklerini kullanırlar.
- Üreticiler: Veri madenciliği teknikleri teknik üretim süreçlerini kontrol etmek ve planlamak için imalat sanayinde kullanılmaktadır. Üçüncü en büyük demir çelik şirketi olan LTV Steel Corp, potansiyel kalite sorunlarını tespit için veri madenciliği tekniklerini kullanarak ürün hatalarını yüzde 99 oranında azaltmayı başarmıştır.
- Sigorta Şirketleri: Sigorta sektörünün veri yoğunluğu veri madenciliğini zorunlu kılar. Bu bilgiler sigortacıların müşterileri daha iyi tanımasına ve sigorta dolandırıcılığını daha etkili bir şekilde tespit etmesine olanak tanır.

Veri Madenciliği uygulamaları bankacılık, sigorta, tıp, gibi çeşitli alanlarda çok geniş bir alanda uygulama fırsatı bulmasına rağmen üretim alanında aynı ilgiyi görememiştir. Bunun sebepleri aşağıdaki gibi açıklanabilir:

- Üretim alanında çalışan araştırmacıların çoğunun veri madenciliği algoritmaları ve araçlarına yabancı olmaları,
- Veri Madenciliği araştırmacılarının çoğunun karmaşık üretim alanlarına yabancı olmaları,
- Veri Madenciliği ve üretim alanında çalışma yapan az miktardaki araştırmacının da üretim verilerine erişme imkânlarının olmayışı,
- Veri Madenciliği tekniklerini üretim sektörüne uygularken geçerlilik ve fayda değerlendirilmesinin zorluğudur (Wang, 2007:489).

Veri madenciliği uygulamaları gelişmiş üretim sistemlerinin aşağıdaki kısımlarında uygulanır:

- Üretim sistemlerinin modellemesi,
- Üretim süreç kontrolü,
- Kalite kontrol,
- İzleme ve teşhis,
- Güvenlik değerlendirme,
- Süreç planlama,
- Üretim kâr optimizasyonu,
- Montaj seçimi,
- Malzeme ihtiyaç planlanması,
- Önleyici makine bakımı (Bilekdemir, 2010:38).

2.4. VERİ TABANLARINDA BİLGİ KEŞFİ SÜRECİ

1960'larda veri toplama sistemleri popülerken, 1970'lerde ilişkisel veri tabanları, 1980'lerde ise ilişkisel veritabanı modelleri kullanılmıştır. 1990'larda adından söz ettirmeye başlayan veri madenciliğinin önemi 2000'li yıllarda oldukça artmıştır (Akın, 2008:43).

Geleneksel sorgu veya raporlama araçlarının büyük veri yığınları karşısında yetersiz kalması; hem Veri Madenciliği hem de Veri Tabanlarında Bilgi Keşfi (VTBK) uygulamalarını doğurmuştur. Veri madenciliği ile VTBK arasında amaçları bakımından çok büyük benzerlik vardır. Veri madenciliğinin en önemli farkı, ikincil belleklerde saklı büyük veri yığınlarının kullanılmasıdır (Yarımağan, 2000:56).

Veri Madenciliği kendi başına oldukça büyük bir alana sahip olmasına rağmen aslında veri tabanlarında bilgi keşfi (VTBK) adı verilen sürecin bir parçasıdır. Çeşitli şekillerde depolanan veriler VTBK sürecinde VM ile anlamlandırılmaya çalışılır.

Birçok kişi veri madenciliği ile diğer popüler terim olan veritabanlarında bilgi keşfini eşanlamlı kabul ederler. Alternatif olarak diğerleri; veri madenciliğini, veritabanında bilgi keşfi sürecinde adımlardan biri olarak kabul ederler (Çetinyokuş, 2008:36).

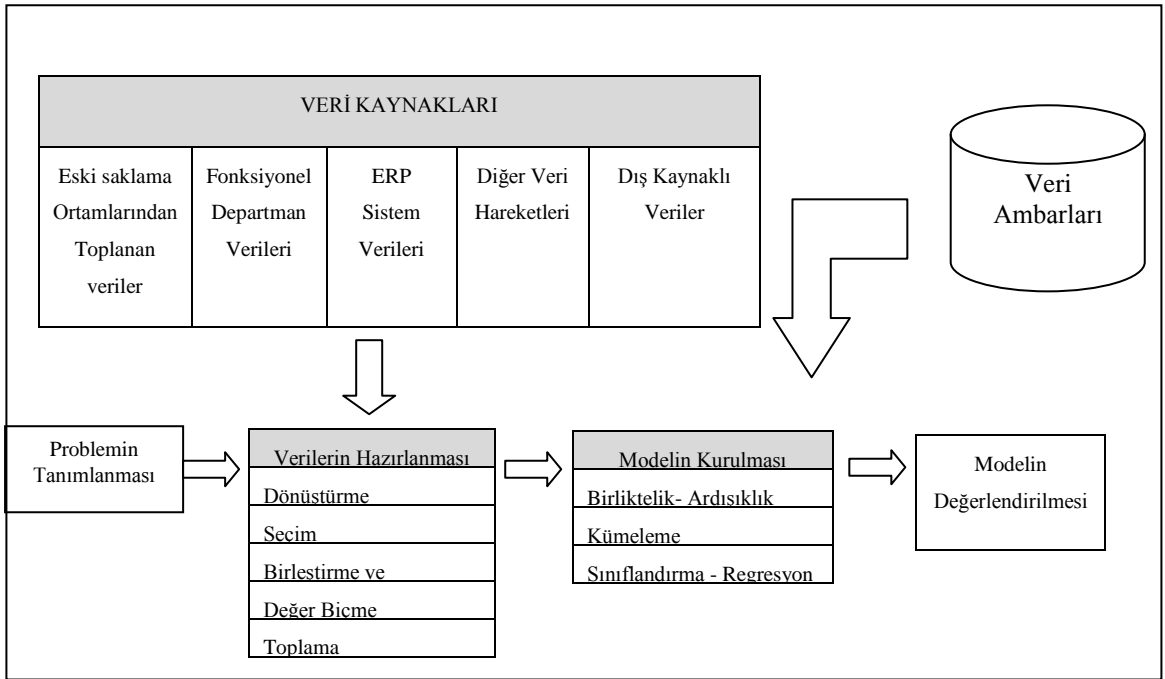
2.5. VERİ MADENCİLİĞİ SÜRECİ

Kurumlar, veri madenciliği sürecini oluştururken kendi içyapılarına, verilerine, amaç ve kaynaklarına göre hareket ederler. Veri madenciliği uygulamalarında başarılı

olmanın en önemli şartı uygulamanın, kurumun hangi amaçları için yapılacağını net bir şekilde belirlenmiş olmasına bağlıdır.

Çeşitli sektörlerde toplanan veriler veritabanlarında birikir. Bu veriler içerisindeki gizli verilerin ortaya çıkarılması kolay değildir. Alt işlemlerden meydana gelen bu süreç 1996 yılında The Cross- Industry Standart Process for Data Mining (CRISP – DM) tarafından oluşturulmuştur (<http://www.crisp-dm.org>).

VM, Şekil 2.3'te gösterilen süreçlerden meydana gelir:



Şekil 2.3. Veri Madenciliği Süreci.

Uyarlandıđı Kaynak: Akpınar, H., "Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliđi", İ.Ü. İşletme Fakóltesi Dergisi, Sayı:1, 07.09.2012, s.1-22. Çevrimiçi <http://www.isletme.istanbul.edu.tr/surekliyayinlar/dergiler/nisan2000/1.htm>.

Amaç ve niteliđi ne olursa olsun VM ile yapılan çalışmalar Şekil 2.3'te belirtilen aşamaların tamamından geçmek zorundadır.

2.5.1. Problemin Tanımlanması

Problemin tanımlanması aşaması, araştırmanın ve veri madenciliđinin amacını, mevcut durumun değerlendirilmesiyle planlama sürecinin belirlenmesini kapsar. Bu aşamada ihtiyaçlar net bir şekilde tanımlanmalıdır. Bu basamakta amaçlar gerçekleştirilirken dikkat edilecek olan performans ölçütlerinin neler olduđuna ve son

olarak bu süreç sonunda ortaya çıkan sonucun hangi durumlar için kullanılacağına karar verilir (Sumathi ve Sivanandam, 2006:195-216).

Veri madenciliği uygulamalarındaki başarısızlıklar genelde problemler için kısa ve uzun vadeli hedeflerin net olarak tanımlanmamasından kaynaklanır. VM uzmanı, hedefleri net bir dilde olduğu kadar kurumun hedefini ve veri madenciliğinin nasıl bir fonksiyonunun olacağını belirterek açıklamalıdır (Fernandez, 2003:48).

2.5.2. Verilerin Hazırlanması

Kurumun bilgi sistemleri üzerinde ürettiği sayısal bilgiyi iyi analiz edebilmesi ve veriler ile mevcut iş problemi arasındaki ilişkiyi net bir şekilde ortaya koyabilmesi gerekir. Verilerin hazırlanması süreci kendi içerisinde verilerin toplanması, temizlenmesi, bütünleştirilmesi, dönüştürülmesi ve indirgenmesi aşamalarından meydana gelir (Han vd., 2012:87).

2.5.2.1. Verilerin Toplanması

Modelin kurulabilmesi için gereken bilgilerin hazırlandığı bu aşamada güvenilir kaynaklardan gelen veriler işletme amaçlarını gözeterek şekilde alınır. Dikkat edileceği üzere veriler henüz işlenmemiş değerlerken, yapılan işlemler neticesinde, süreç sonunda anlamlı değerlere yani bilgiye dönüşür.

2.5.2.2. Verilerin Temizlenmesi

Yanlış girilmiş gürültülü verilerin, uç değerlerin, eksik veya yanlış verilerin temizlenmesi olarak tanımlanan bu süreçte eksik değerleri doldurma, kayda almama veya eksik değere değişkenin ortalama değerini verme, regresyon veya karar ağacı teknikleriyle eksik verileri tahmin gibi tekniklerle hatalar düzeltilebilir (Han vd., 2012:88)

2.5.2.3. Verilerin Bütünleştirilmesi

Veri bütünleştirme işlemi, farklı veritabanlarından alınan verilerin tek bir veritabanına toplanması işlemidir. Toplanan bu verilerin aynı standarda sahip olmaması veya aynı veritabanından çekilen verilerin zaman içerisinde revizyonlara uğramış olması bazı sıkıntılar doğuracaktır.

Bu gibi durumlardan kaçınmak için “meta veri” kullanılır. Meta veriler mevcut veriler hakkındaki tüm detayları bünyesinde bulundurur.

2.5.2.4. Verilerin Dönüştürülmesi

Verilerin ilgilenilen özellikleri korunarak modele uygun hale getirilmesi adıdır. Verilerin veri madenciliğine uygun bir forma dönüştürülme işlemi olan bu süreçte düzeltme, birleştirme genelleştirme ve normalleştirme basamakları yer alır (Han vd., 2012:94).

2.5.2.5. Verilerin İndirgenmesi

Modele girmesi muhtemel gereksiz verilerin azaltılmaya çalışıldığı bu süreçte değişken sayısının çok fazla olduğu ve değişkenlerin önem derecesine göre sıralanması gereken durumlarda faktör analizi, temel bileşenler analizi gibi yöntemlere başvurulur boyut indirgenerek model daha yalın ve net bir hale getirilir. Bu da modelin daha sağlıklı sonuçlar vermesine sebep olur.

Veri küpü, veri sıkıştırma, örnekleme, nitelik birleştirme, kesikli hale getirme, boyut indirgeme gibi teknikler veri indirgeme tekniklerinden bazılarıdır.

2.5.3. Modelin Kurulması

Modelleme aşaması, veri madenciliği yazılımı yardımıyla uygun teknikler kullanılarak farklı durumlar için sonuçlar üretilmesi aşamasıdır (Olson ve Shi, 2007:24). Bu aşamada analize uygun hale getirilen verilerle birçok model kurularak en iyi model seçilmeye çalışılır. En iyi performansı verecek olan model ve algoritmanın tespit edilmeye çalışıldığı bu süreçte muhtemel tüm modeller kurulup karşılaştırılmalıdır (Chapman, Clinton, Kerber, Khabaza ve Reinartz, 2009:29).

Genellikle ilk olarak kümeleme analizi ve verinin görselleştirilmesi teknikleri kullanılmaktadır. Verinin tipine göre daha sonra çeşitli modeller uygulanabilir. Eğer amaç verinin gruplandırılması ise ve gruplar belirli ise, diskriminant analizi uygun olabilir. Eğer amaç tahmin ise ve veri sürekli ise regresyon analizi, eğer veri sürekli değilse lojistik regresyon uygun olabilir. Her iki amaç için de yapay sinir ağları kullanılabilir. Verinin sınıflandırılması için karar ağaçları tekniği de kullanılabilir (Olson ve Shi, 2007:24).

Model kurulurken kullanılan değişkenlerin birbirleri ile ilişki düzeyleri gerektiğinden fazla ise, daha anlamlı olan değişkeni modele almak faydalı olacaktır. Eğer bu süreç sonucunda en iyi modelin kurulduğu düşüncesi oluşmuşsa modelin değerlendirilmesi aşamasına geçilebilir.

2.5.4. Modelin Değerlendirilmesi

Kurulan modellerin karşılaştırıldığı ve en iyi modelin seçildiği basamaktır. Kurulan modeller içinden en doğru sonucu vereninin yani doğruluk derecesi yüksek olanının bulunabilmesi için bazı teknik ve yöntemler vardır. Özellikle sınıflama problemleri için kurulan modellerin doğruluk derecelerinin değerlendirilmesinde risk matrisleri kullanılmaktadır. Kaldıraç (Lift) oranı ve grafiği ise, bir modelin sağladığı faydanın değerlendirilmesinde kullanılan önemli bir yardımcıdır (Çetin, 2009:19).

Kurulan modelin doğruluk derecesi ne denli yüksek olursa olsun, gerçek dünyayı tam anlamı ile modellediğini garanti edebilmek mümkün değildir. Yapılan testler sonucunda geçerli bir modelin doğru olmamasındaki başlıca nedenler, model kuruluşunda kabul edilen varsayımlar ve modelde kullanılan verilerin doğru olmamasıdır. (Çetin, 2009:19).

2.5.5. Modelin Uygulanması

Seçilen ve işletme ihtiyaçlarına uygunluğu değerlendirilen modelin ne ölçüde etkin, etkili ve verimli kullanılabildiğinin belirlendiği bu basamak sürecin son basamağıdır. Seçilen model tek başına çalışabilecek bir sistem olabileceği gibi bir sistemin alt basamağı da olabilir. Kurulan modeller, risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında karar vericiye destek amacıyla doğrudan kullanılabileceği gibi, malzeme ihtiyaç planlaması, kurumsal kaynak planlaması gibi süreçlerin alt sistemleri biçiminde de kullanılabilir.

2.6. VERİ AMBARLARI

Veri madenciliği çalışmalarında veri ve veritabanı kavramları ön plana çıkar. Fakat işletmelerde kullanılan işlemsel veritabanları (transactional database) veri madenciliği uygulamalarında doğrudan kullanılamaz. Bu verilerin veri madenciliği amacıyla kullanılabilmesi için düzenlenip kullanıma uygun hale getirilmesi gerekir. Belli bir döneme ait, çalışmaya uygun olarak düzenlenmiş, işletmelere ait bu veritabanlarına veri ambarları denir (Silahtaroglu, 2008:15).

Veri ambarı, analitik işlemlerin yapılması için ihtiyaç duyulan bilgi temelinin sağlar (Inmon, 2007:68). Yönetimin karar verme sürecini desteklemede verinin zaman değişkenli, bütünleşik, konu yönelimli ve içerisinde değişikliğe izin vermeyen, bilgilerin sadece okunabildiği bir toplamıdır (Özkan, 2011:22).

Veri ambarları veritabanlarında olduğu gibi verilerin depolandığı alanlardır. Ama veriler, veritabanlarından farklı olarak daha da özet bir hale getirilmiştir. Olap gibi analitik işlemler için kullanılır. Özne tabanlı, bütünleşmiş ve zaman dilimli, yöneticinin karar işlemlerinde yardımcı olacak şekilde toplanmış, sabit veriler topluluğudur (Alpat, 2006:89).

Farklı kaynaklardan toplanan veriler VM süreçlerinden geçirilerek veri ambarlarına aktarılır. Veri ambarları üzerinde artık hiçbir değiştirme işlemi yapılamaz. Yani veri ambarlarındaki veriler silinemez veya yeni veri eklenemez. Veri ambarlarının belirli aralıklarla güncellenmesi oluşturulan modelin geçerliliği açısından büyük önem taşır.

Veri ambarları konu odaklı olup, aynı olayı veya varlığı ilgilendiren veriler birbirlerine bağlanmıştır. Örneğin bir veri ambarı müşteri, ürün gibi varlıklar ya da satış, reklam, pazarlama, teslimat gibi olaylara dönük düzenlenmiş olabilir. Birden fazla veri tabanı bir araya getirilerek farklı bilgiler bu veritabanlarına aktarılmış ve bütünleştirilmiştir. Veriler belli bir döneme ait olup verilerde zaman içindeki değişiklikler raporlamaya uygun bir haldedir (Silahtaroglu, 2008:15-16).

2.6.1. Veri Ambarı Mimarisi

İşletmelerin ihtiyaçlarının iyi anlaşılması ve analiz edilmiş olması iyi bir veri ambarı oluşturmanın en temel koşuludur. İyi planlanmış bir veri ambarında veriler ve sistemler arasındaki iletişim ve işlemler net bir şekilde gösterilmiş olmalıdır.

Veri ambarları kullanılarak raporlamalar, sorgulamalar, görselleştirmeler ve alt veri depoları yani data martlar oluşturulabilir. Data martlar küçük boyutlu bölümsel ambarlardır. Organizasyonun (işletmenin) belirli kullanıcıları için ayrılmış ve onlara ait verileri içerir. Data martlar pazarlama alanına yönelik özgün odaklı tematik veritabanlarıdır. Bu anlamda yeni ve / veya potansiyel müşterilere bağlı tüm bilgileri içeren bir iş arşivi olarak kabul edilebilir (Giudici, 2003:22).

2.6.2. Veri Modelleme

Veri ambarının kuruluşunda veriyi anlamak, veriyi anlamak için ise önce bir veri modeli kurmak gerekir. Veri modeli, verinin tanımını, yapısını ve verideki örüntüleri ifade etmeye yarar. Veri modellemenin amacı, verinin taşıdığı anlamı, veriler arasındaki ilişkileri, verinin niteliklerini açıkça belirtmek olarak genellenebilir. İyi bir veri modeli

varlıkları, varlıkların niteliklerini, veriler arasındaki ilişkileri, birincil, ikincil veya alternatif anahtarları, kurumuna faaliyet konularını ve konular arasındaki ilişkileri iyi bir şekilde yansıtabilmelidir (Gürsoy, 2009:10).

• Çevrimiçi Analitik İşleme (OLAP - On Line Analytical Processing)

Veri ambarları üzerinde çeşitli konularla ilgili kararlar almaya yardımcı olan veri analiz ve sorgulama işlemlerinin bütününe OLAP denir. OLAP sistemlerinin terminolojisi ve bazı yetenekleri milyonlarca insan tarafından kullanılan tablolarlama programları içinde kullanılmaya başlanmıştır. OLAP sistemlerinin interaktif veri analizi noktasında güçlü bir etkisi vardır ve genellikle veri görselleştirme ve özet istatistikler üretmek için geniş olanaklar sağlar (Tan vd, 2002:131).

• Çok Boyutlu Çevrimiçi Analitik İşleme (MOLAP)

MOLAP veri tabanı, dizi tabanlı boyutlu depolama motorları aracılığıyla çok boyutlu verileri desteklemektedir. Doğrudan veri küpüne dizi yapıları çok boyutlu matrisleri eşleştirmekte kullanılırlar. Bir veri küpü kullanmanın avantajı, özetlenen verilerin hızlı indekslenmesini sağlamasıdır. Veri seti seyrek ise çok boyutlu veri depoları ile depolama kullanımını düşük olabilir. Bu gibi durumlarda, seyrek matris sıkıştırma tekniklerinin kullanılması gerekir (Han, Kamber, Pei, 2012:164).

MOLAP, çok boyutlu bir veri ambarlama sistemine sahip olup verilerin saklanması için iki yöntem kullanılır. Bunlardan biri veri küplerinin diğeri ise standart ilişkisel veri tabanı yönetim sistemlerinin kullanılmasıdır. Bu yöntemler veriyi “Yıldız Şema” kullanarak saklarlar (Gürsoy, 2009:15).

• İlişkisel Çevrimiçi Analitik İşleme (ROLAP)

MOLAP’a alternatif bir teknoloji olan ROLAP dinamik çok boyutlu verilerin analizinde kullanılır. ROLAP teknolojisinde, verilerin düzenlenmesi için gerçek tabloları ve boyut tabloları içeren şemalar kullanılır. Bunlar; Kar tanesi şeması, Yıldız şeması ve Gerçek takımyıldızları şemasıdır (Han, Kamber, Pei, 2012:164).

○ Yıldız Şeması

Sıklıkla kullanılan modellerden biri olan Yıldız Şeması, merkez tablo ve birden fazla boyut tablosunu içermektedir. Şemada bulunan merkez tabloda sayısal değerler

bulunurken, boyut tablolarında ise merkez tablosundaki başlıklar ayrıntılı olarak alt başlıklar şeklinde yer alır (Esen, 2009:71).

○ **Kar Tanesi Şeması**

Kar Tanesi Şeması, bir merkez tablo ile boyut tablolarından oluşmaktadır. Bu şema boyut tablolarına bir hiyerarşi eklemeye olanak tanır (Esen, 2009:71).

○ **Gerçek Takım Yıldızları Şeması**

Bazı karmaşık uygulamalarda boyut tablolarını paylaşan birden fazla merkezi tabloya ihtiyaç duyulabilir. Takımyıldızı şemada bazı referans tablolar normalize, bazıları da denormalize edilmiştir. Bu tip şemalar, birden fazla yıldız şemanın bir araya gelmiş haline benzer (Esen, 2009:71).

2.7. VERİ MADENCİLİĞİNDE KULLANILAN YÖNTEMLER

Veri madenciliği yöntemleri işlevlerine göre aşağıdaki gibi üç temel grupta tanımlanmaktadır.

- Sınıflama (Classification),
- Kümeleme (Clustering),
- Birliktelik kuralları ve sıralı örüntülerdir (Association rules and sequential patterns).

Gerek tanımlayıcı gerekse tahmin edici modellerde yoğun olarak kullanılan belli başlı istatistiksel yöntemler; sınıflama (classification) ve regresyon (regression), kümeleme (clustering), birliktelik kuralları (association rules) ve ardışık zamanlı örüntüler (sequential patterns), bellek tabanlı yöntemler, yapay sinir ağları ve karar ağaçları olarak gruplandırılabilir. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir (Dumouchel, 1999).

Tahminleyici yaklaşımlar içinde; en önemli iki teknik sınıflandırma ve karar ağaçlarıdır. Ayrıca yapay sinir ağları, genetik algoritmalar ve Bayes yöntemleri de bu amaç için üretilmiş algoritmalar. Verilerin ortak olan özelliklerine göre ayrılması işlemi olan sınıflandırma, ürünün özellikleri ile müşteri özelliklerinin eşleştirilmesi için kullanılır. Böylece her iki taraf için de ideal olanın belirlenmesi hedeflenir (Gürsoy, 2009:42).

Dolandırıcılık tespiti, pazarlama, kalite kontrol çalışmaları, hastalık tespitleri gibi konuların oldukça fazla çalışıldığı alanlardan biri olan sınıflandırma, en çok bilinen ve kullanılan VM yöntemlerinden birisidir (Dunham, 2003:8).

Veri madenciliğinde verilerin önceden verilen sınıflara göre ayrılması, aslında ileride elde edilecek sonuçların yani sınıfların tahminidir. İstatistiksel tahmin modelleri kullanılarak bu sınıflandırma işlemleri yapılabilir (Silahtaroglu, 2008:60).

İstatistiksel yöntemler, birçok VM uygulamasında yoğun bir şekilde kullanılır. Regresyon analizi, lojistik regresyon analizi, zaman serileri analizi ve Bayesyen yaklaşım gibi teknikler istatistiğe dayalı algoritmalarıdır.

2.7.1. Regresyon Analizi

Regresyon analizi, herhangi bir bağımlı değişkenin bir veya birden fazla bağımsız değişkenle arasındaki ilişkinin matematiksel bir fonksiyon olarak yazılmasıdır. Bu fonksiyona regresyon denklemi adı verilir (Orhunbilge, 2010:61).

Bir ya da daha çok değişkenin başka değişkenler cinsinden tahmin edilmesini sağlayacak ilişkileri bulma ve tanımlama işlemine regresyon analizi denir. Regresyon analizinde, gözlenen bir olay değerlendirilirken bu olayın hangi olaylardan etkilendiğinin belirlenmesi önemlidir. Bu olaylar bir ya da birden çok olabileceği gibi etki düzeyleri farklı seviyelerde de olabilirler. Regresyon analizinde verilerin matematiksel bir fonksiyon şeklinde tanımlanması gerekmektedir. Oluşturulan matematiksel fonksiyon bir bağımlı değişken ve bir veya birden çok bağımsız değişkenden oluşabilir. Değişkenler sayılabilir veya ölçülebilir nitelikte olabilir. Tek değişkenli modeller basit doğrusal regresyon, birden fazla bağımsız değişkenli modeller ise çoklu regresyon modelleri olarak adlandırılır (Argüden ve Erşahin, 2008:48).

Regresyon analizi, standart dağılım, standart sapma, diskriminant analizi, güven aralıkları gibi verileri ve veriler arasındaki ilişkiyi inceleyen yöntemler klasik istatistik çalışmalarıdır. Bu yöntemler ileri düzey istatistiksel analizin temelini oluşturmuştur.

“Eldeki verilerden hareket ederek geleceğin tahmin edilmesinde faydalanılan ve veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olan sınıflama ve regresyon modelleri arasındaki temel fark, tahmin edilen bağımlı değişkenin kategorik veya süreklilik gösteren bir değere sahip olmasıdır. Ancak çok terimli Lojistik Regresyon (Multinomial Logistic Regression) gibi kategorik değerlerin de tahmin

edilmesine olanak sağlayan tekniklerle, her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün hale gelmiştir” (Çetin, 2009:66).

2.7.1.1. Basit Doğrusal Regresyon Analizi

Basit doğrusal regresyonda iki değişken arasındaki ilişkinin tanımlanması amaçlanır. Eldeki verilerin kullanılması ile bir doğru oluşturulur. Doğru tüm veri noktalarından tahmin edilen eğriye olan uzaklığın kareleri minimize edilmek suretiyle optimize edilir. Doğru elde edildikten sonra iki değişken arasındaki ilişkinin gücü r korelasyon katsayısı değeri ile ölçülür.

Örneğin, tek bir bağımsız değişkenli (X) ve iki parametrelili (B_0 ve B_1) verilmişse ana kütle için basit doğrusal regresyon denklemi aşağıdaki gibi yazılır:

$$Y = B_0 + B_1 X + \mathcal{E}$$

Denklemden B_0 doğrusal fonksiyonun sabiti, B_1 ise doğrusal fonksiyonun eğimidir. Fonksiyonun eğimi regresyon katsayısı olarak da adlandırılır ve X ' teki bir birimlik değişimin, Y üzerinde yine Y cinsinden meydana getireceği değişikliği gösteren katsayıdır. Denklemden Y bağımlı değişken olup, sınıfları temsil eder. X ise bağımsız değişkeni, yani nitelikleri temsil eder (Orhunbilge, 2010:64).

- **Regresyon Denklemindeki Bilinmeyenlerin Bulunması**

Ana kütlede rastgele bir örneklem elde edilirse, ana kütle parametreleri için örneklem tahminleri bulunur ve aşağıdaki örneklem doğrusal regresyon denklemi elde edilir:

$$y = b_0 + b_1 x + e$$

Burada e terimi örneklemde elde edilen hata ve y' tahmini bağımlı değişken değeri olmak üzere $e = y - y'$ ile gösterilir.

- **En Küçük Kareler Yöntemi**

Parametre kestirimleri, hataların kareleri toplamının minimum değeri bulunarak elde edilir. Bu yöntem “En Küçük Kareler Yöntemi” olarak adlandırılır ve hatalar minimum kareler toplamı aşağıdaki şekilde ifade edilir:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y'_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

Bu fonksiyonun minimum değerini bulmak için her parametrenin birinci türevi alınarak sıfıra eşitlenir. Böylece her kısmısal türeve göre bir denklem elde edilir. Örneğin iki parametrelili doğrusal regresyon için iki değişkenli iki denklem elde edilir.

$$\frac{de}{db_0} = 2\sum(-1)(y - b_0 - b_1x) = -\sum y + nb_0 + b\sum x = 0$$

$$\frac{de}{db_1} = 2\sum(-x)(y - b_0 - b_1x) = -\sum xy + b_0\sum x + b\sum x^2 = 0$$

Negatif işaretli terimler eşitliğin sağ tarafına geçirilerek normal denklemler elde edilir.

$$\sum y = nb_0 + b_1\sum x$$

$$\sum xy = b_0\sum x + b_1\sum x^2$$

Bu iki bilinmeyenli denklemlerde cebirsel eleme veya matris yöntemiyle çözüm yapılarak b_0 ve b_1 katsayılarının değerleri elde edilir.

Denklemlerde x ve y gözlem değerleri yerine bu değerlerin aritmetik ortalamadan farkları konulabilir. Aritmetik ortalamadan farkların cebirsel toplamı sıfır olduğundan değişkenlerin toplamını ifade eden terimler sıfır olur (Orhunbilge, 2010:67).

$$\sum(y - \bar{y}) = 0 \text{ ve } \sum(x - \bar{x}) = 0$$

Böylece birinci denklemde $b_0 = 0$ ve ikinci denklemde ise

$$b_1 = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2} \text{ elde edilir.}$$

Burada \bar{y} , y gözlem değerlerinin ortalaması ve \bar{x} ise x gözlem değerlerinin ortalamasıdır.

- **Tahminlerin Standart Hatası**

Genellikle S ile gösterilen standart hata, regresyon doğrusuna uyumun bir göstergesidir. Çünkü $Y_i - Y_i'$ farklarının küçük çıkması Y_i' kestirimlerinin Y_i gözlem değerlerine yaklaştığının yani regresyon doğrusunun verilere uygunluğunun bir göstergesidir (Alpar, 2010:167).

Ana kütleler için basit doğrusal regresyon denkleminin standart hatası:

$$S_{yx} = \sqrt{\frac{\sum(y-\bar{y})^2}{N}} \text{ olur.}$$

2.7.1.2. Çoklu Doğrusal Regresyon Analizi

Bireylerin, diğer canlıların, nesne, kurum ve kuruluşların değişik özelliklerinin bir arada analiz edilmesine olanak veren bu yöntemler topluluğuna “Çok değişkenli istatistik yöntemler (Multivariate Analysis)” denir (Orhunbilge, 2010:3).

Basit regresyon analizindeki tüm varsayımlar çoklu regresyon analizi için de geçerlidir. Çoklu doğrusal regresyonda iki veya daha çok sayıda bağımsız değişken veya bağımsız değişken fonksiyonu bulunur. Örneğin, önce verilmiş olan regresyon modeli yeni bir terim eklenerek değiştirilirse; aşağıdaki çoklu doğrusal regresyon modeli ortaya çıkar:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_i x_i + e_i \quad i=1, n$$

Çoklu regresyon modeli $x_{1i}, x_{2i}, \dots, x_{pi}$ girdi değerlerine bağlı olarak y_i değerinin tahmin edilmesini sağlar. Modelde yer alan regresyon katsayıları, en küçük kareler yöntemi kullanılarak hesaplanır.

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - y'_i)^2$$

Çoklu doğrusal regresyon modelinin varsayımları aşağıda gösterilmiştir:

- Normal dağılıma sahip olması,
- Doğrusallık olması,
- Hata terimlerinin ortalamasının 0 olması,
- Sabit varyansın olması,
- Otokorelasyon olmaması,
- Çoklu bağlantı olmamasıdır.

Bağımsız değişken olarak seçilmesi planlanan değişkenlerin modele girebilmesi için bağımlı değişken ile çoklu korelasyon katsayılarının incelenmesi neticesinde karar verilebilir. Bağımsız değişkenler ile bağımlı değişken arasındaki çoklu korelasyon katsayısının 1'e yakın olması bağımsız değişkenlerin bağımlı değişkeni etkilediği ve birlikte modele girebilecekleri anlamına gelir. Ancak bu kararın alınabilmesi için bağımsız değişkenlerin de birbirlerinden etkilenmemesi gerekir.

- **Çoklu Açıklayıcılık Katsayısı (Belirtme Katsayısı) (R^2)**

Açıklanabilen değişimin toplam değişim içindeki payı olan açıklayıcılık katsayısı, bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişkenler tarafından açıklanabildiğini gösterir. R^2 , 0 ile 1 arasında değerler alır. Değişkenler arasında doğrusal bir ilişki varken R^2 değerinin 1'e yaklaşması bağımlı değişkendeki değişimin büyük bir kısmının bağımsız değişkenler tarafından açıklandığını ve varsayımlar sağlandığında modelin uygunluğunu gösterir (Alpar, 1997:173).

$$R^2 = \frac{\text{Açıklanabilen Değişim}}{\text{Toplam Değişim}} \text{ denkleminde hesaplanır.}$$

Eğer R^2 değeri sıfıra yakınsa, toplanan verilere uygulanan modelin uygun olmadığı sonucu çıkarılır ve modelin değiştirilmesi gerekir. Eğer R^2 değeri 1'e yakınsa, modelin uygun olduğu sonucu çıkarılır.

Regresyon varsayımlarının yerine getirilmediği durumlarda, uygun olmayan modeller için de R^2 'nin yüksek değerler alması mümkündür. Dolayısıyla R^2 model uygunluğu için güvenilir bir ölçü değildir. Düzeltilmiş R^2 model uygunluğu ölçümünde göreceli olarak daha iyi bir performansa sahiptir ve birçok istatistik yazılımı bu ölçütü kullanmaktadır (Özçınar, 2006:28).

- **Düzeltilmiş R^2**

R^2 belirtme katsayısı çoklu modellerde genellikle yeterli değildir. Çünkü çoklu regresyon modelleri için denkleme yeni değişken ilave edilmesi durumunda R^2 değeri genellikle artar. Bu yüzden anlamlı bir test yapabilmek için çoklu modellerde düzeltilmiş R^2 , model seçim kriteri olarak dikkate alınır.

n: gözlem sayısı, k: modeldeki değişken sayısı (bağımsız değişken + bağımlı değişken) olmak üzere;

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} \text{ denkleminde belirlenir.}$$

Tekli regresyon modellerinde olduğu gibi belirlilik katsayısı 1'e ne kadar yakın ise mevcut model o kadar anlamlıdır.

2.7.1.3. Regresyon Varsayımları

Çok değişkenli istatistik yöntemlerinde, otokorelasyon (hataların bağımsız olması), çoklu doğrusal bağlantı, normallik, eşit varyanslılık ve doğrusallık olmak üzere beş varsayım vardır (Orhunbilge, 2010:29).

1. Otokorelasyon

Otokorelasyon, çoklu regresyon analizinde hata teriminin birbirini izleyen değerleri arasında ilişki bulunması halidir. Bu durum, genel doğrusal regresyon modelinin önemli bir varsayımından sapmadır. Genel doğrusal regresyon modeli varsayım gereği olarak, hata terimleri arasında bir ilişki yoktur.

Otokorelasyon, uygulamada daha çok zaman serilerinde ortaya çıkmakla birlikte, yatay kesit verilerinde de otokorelasyona rastlanabilir (Serper,1986:303). Zaman serilerinde otokorelasyon, zaman periyodunun büyüklüğü veya küçüklüğüne göre değişebilir. Periyot bir aylık veriye dayanıyorsa, otokorelasyon büyük, iki aylıksa biraz daha küçük ve yıllıksa daha da küçüktür.

Otokorelasyonun nedenleri aşağıda gösterilmiştir:

- Bazı açıklayıcı değişkenlerin modele alınmaması,
- Modelin matematiksel biçiminin yanlış seçilmesi,
- Açıklanan değişkende ölçme hatası olması,
- Verilerin işlenmesi,
- Hata teriminin yanlış belirlenmesi,

Otokorelasyon sonucunda ise;

- Parametre tahminleri sapmasız olmakla birlikte etkin değildir.
- Hata teriminin varyansı, olduğundan küçük tahmin edilmektedir.
- E.K.K. tahminlerine göre yapılan öngörüler etkin değildir.

• Durbin-Watson Testi

Durbin-Watson istatistiği pozitif ve negatif otokorelasyonların belirlenmesinde kullanılan testlerden biridir (Orhunbilge, 2010:48). Dört aşamalı bir test olan Durbin-Watson testi sadece birinci tip otokorelasyonun varlığını sınamaktadır.

1. Hipotezlerin kurulması aşaması

- $H_0: \rho = 0$ (otokorelasyon yoktur)
- $H_1: \rho \neq 0$ (otokorelasyon vardır)

2. Tablo değerlerinin bulunması aşaması

Bu aşamada, seçilen bir anlamlılık düzeyi ile gözlem sayısı ve açıklayıcı değişken sayısına göre, Durbin-Watson tablosundan, d istatistiğinin alt (d_{alt}) ve üst ($d_{üst}$) sınırları bulunur.

3. d istatistiğinin hesaplanması aşaması

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (\text{Orhunbilge, 2010:48})$$

4. Karşılaştırma ve karar aşaması

İkinci aşamada bulunan tablo değerleri ile üçüncü aşamada hesaplanan d istatistiğinin karşılaştırılarak otokorelasyon hakkında bir sonuca ulaşıldığı aşamadır. d_L alt sınır, d_U üst sınır olmak üzere; karar vermede kullanılan eşitsizlikler aşağıda gösterilmiştir:

- $0 < d < d_L$ ise pozitif otokorelasyon vardır.
- $d \leq d_{sub} > d_L \leq d_U$ ise karar verilemez.
- $d_U \leq d < 4 - d_U$ ise otokorelasyon yoktur.
- $4 - d_U \leq d \leq 4 - d_L$ ise karar verilemez.
- $4 - d_L < d < 4$ ise negatif otokorelasyon sonuçları vardır (Serper, 1986:319).

2. Çoklu Doğrusal Bağlantı Olmaması

Bağımsız değişkenlerin modele girebilmeleri için aralarında ilişki olmaması, yani aralarında çoklu bağlantı olmaması gerekir. Örneğin X_1 ve X_2 bağımsız değişkenleri arasındaki çoklu bağlantı analizi aşağıda gösterilmiştir:

r_{12} , X_1 ve X_2 değişkenleri arasındaki basit doğrusal korelasyon katsayısı olmak üzere;

$$r_{12} = \frac{\sum d_{12}}{\sqrt{\sum d_1^2 \sum d_2^2}}$$

r_{12} 'ün 0 veya 0'a çok yakın olması gerekir.

Kullanılan veri derleme yöntemi, modelde veya örneklemedeki sınırlamalar, değişkenler arasındaki iktisadi ilişki, çalışılan veri azlığı gibi sebepler çoklu doğrusal bağlantıya sebep olurlar.

Çoklu doğrusallık sorununun belirlenmesi için çeşitli sezgisel ve sayısal yöntemler kullanılabilir. Bu yöntemlerden ilki bağımsız değişkenler arasındaki korelasyon değerlerinin incelenmesidir. Korelasyon değerlerinin 0,90' dan büyük olması durumunda çoklu doğrusallık durumundan şüphe edilebilir. Diğer bir yöntem ise düşük t istatistiğinin, yüksek standart hata ve geniş güven aralıklarının olmasıdır. Bu gibi testlerin yanı sıra çoklu doğrusallık sorununun tespitinde kullanılacak diğer yöntemler ise VIF değeri, Tolerans değeri ve CI koşul endeksinin belirlenmesidir. Çoklu bağlantı varsayımlarının testleri aşağıda gösterilmiştir:

- **VIF Değeri (Varyans Artış Faktörleri):**

Bağımsız değişkenler arasında bir bağıntı olup olmadığını ve her bir bağımsız n tane değişkenin diğer bağımsız değişkenler ile ayrı ayrı regresyon işlemi gerçekleştirilmesi sonucu hesaplanan R^2 değerlerinin kullanılmasıyla elde edilen VIF değeri hangi değişken veya değişkenlerin çoklu doğrusallık durumuna yol açtığı belirlenmesinde kullanılan bir istatistiktir. Herhangi bir bağımsız değişkenin diğerleri ile ilişkisi arttıkça VIF değeri de artacaktır. VIF değeri 1 ile ∞ arasında değerler alır. Genellikle VIF değerlerinin 5'in üzerinde olmaması istenir (Alpar, 2010:303).

Her bağımsız değişken (n) için:

$$VIF_n = \frac{1}{(1-r_n^2)}$$

değeri ile hesaplanır.

- **Tolerans Değeri:**

Tolerans değeri, diğer tahminleyiciler dikkate alındığında elimizdeki tahminleyiciye ait varyans yüzdesidir. VIF değerinin tersi olan bu değer her bir bağımsız değişken için $1-r_n^2$ denklemiyle elde edilir. Tolerans değerinin $> 0,2$ olduğu durumlarda çoklu doğrusal bağlantı önemsiz kabul edilir (Orhunbilge, 2010:54).

- **CI Koşul Endeksi:**

Çoklu doğrusallık sorununun tespitinde kullanılan bir diğer test olan CI koşul endeksi, maksimum özdeğerin (λ_{\max}) minimum özdeğere (λ_{\min}) bölümünün karekökü alınarak hesaplanır. $CI \geq 30$ iken yüksek dereceden çoklu bağlantı sorunu vardır.

$$CI = \sqrt{\frac{\text{maksimum özdeğer}}{\text{minimum özdeğer}}} = \sqrt{\frac{(\lambda_{\max})}{(\lambda_{\min})}}$$

- **Özdeğer Vektörleri (Eigenvalues):**

Yukarıdaki yöntemlerin yanı sıra varyans kovaryans Σ matrisinin özdeğerlerine bakılarak çoklu bağlantı sorunları tespit edilebilir. Σ matrisinin birden fazla özdeğeri 0'a yakın değerler alıyorsa çoklu bağlantı sorunundan bahsedilebilir.

- **Çoklu Bağlantının F ve t Testi Karşılaştırması ile Belirlenmesi**

Katsayılara ilişkin t istatistiklerinin tümü anlamsız iken F istatistiğinin anlamlı çıkması çoklu bağlantı sorununa işaret edebilir. Ancak anlamlı derecede bağlantı sorunu olan birçok veri seti bu şekilde bir sonuç sergilemediği için bu yaklaşım kuşku olabilmektedir (Montgomery, 1991:45).

3. Hataların normal dağılımı

Normallik varsayımı tüm istatistik yöntemler için geçerli olan bir varsayımdır. Normallik varsayımından sapmanın söz konusu olduğu durumlarda ana kütle için anlamlılık testleri yapılamaz ve tahminlerde güven aralıkları oluşturulamaz. Çok değişkenli regresyon analizinde hataların normallikte sapma göstermesinin sebepleri;

- Önemli bir değişkenin modele alınmaması,
- Örnek birim sayısının yetersiz olması,
- Uygulanan yöntemin diğer varsayımlarından sapma söz konusu olması olarak gösterilebilir (Orhunbilge, 201:38).

Normallikten sapma durumunda önemli değişkenin modele dahil edilmesi veya örneklem sayısının arttırılmasının çözüm getirmemesi durumunda yöntemin diğer varsayımlarından sapma olup olmadığı incelenir. Sapma saptanırsa gereken düzeltmeler yapıldığında hatalar da normal dağılım gösterebilir (Orhunbilge, 201:38).

4. Doğrusallık

İlişkinin doğrusallığı, bağımlı değişkendeki değişmelerin bağımsız değişkene bağımlılığının derecesini gösterir. İlişkinin doğrusal olduğu durumda regresyon katsayıları sabit kalır. Standardize edilmiş hatalarla tahmini bağımlı değişken grafiğindeki doğrusal ilişkinin, seçilen modelde söz konusu olmadığı aksine doğrusal olmayan bir ilişkinin varlığının gözleneceği söylenebilir. Çoklu regresyon analizinde, doğrusal olmayan eğilimin hangi bağımsız değişkenden kaynaklandığını belirleyebilmek için bağımsız değişkenlere ait regresyon denkleminin standardize hataları için bu inceleme yapılmalıdır (Orhunbilge, 2010:57).

2.7.1.4. Korelasyon Analizi

Korelasyon bağımlı değişkenle bağımsız değişken veya değişkenler arasındaki ilişkinin gücünü gösteren bir katsayıdır (Orhunbilge, 2010:62).

Korelasyon katsayısı, iki değişken arasındaki doğrusal ilişkinin ölçüsü olup, incelenen değişkenlerin birimlerinden bağımsızdır ve örneklem için $-1 \leq r \leq 1$ dir (Alpar, 2010:57).

Korelasyon katsayısının 0'a yaklaşması değişkenler arasında zayıf ilişkinin varlığını gösterir. Değişkenler birlikte artıyor veya azalıyorsa pozitif yönde, değişkenlerden biri artarken diğeri azalıyorsa ise negatif yönde bir ilişki vardır.

Korelasyon katsayılarının türlerini belirlemek için anakütle analizlerinde ρ 'ya, örnek analizlerinde ise r 'ye endisler ilave edilir. Örneğin r_{yx} basit korelasyon katsayısını, r_{yyx}^2 ise ikinci dereceden korelasyon katsayısını göstermektedir.

r_{12} ; x_1 ve x_2 bağımlı değişkenleri arasındaki basit korelasyon katsayısı, r örneklem korelasyon katsayısı, ρ ise anakütle korelasyon katsayısı olmak üzere;

$$r_{12} = \frac{\sum x_1 x_2 - \frac{\sum x_1 \sum x_2}{n}}{\sqrt{\left(\sum x_1^2 - \frac{(\sum x_1)^2}{n}\right) \left(\sum x_2^2 - \frac{(\sum x_2)^2}{n}\right)}} = \frac{\sum d_{12}}{\sqrt{\sum d_1^2 \sum d_2^2}} \text{ denkleminde bulunur.}$$

Çoklu korelasyon katsayısı ise aşağıda verilen denklem ile ifade edilir:

$$\rho = \sqrt{1 - \frac{\sum (Y - Y')^2}{\sum (Y - \bar{Y})^2}}$$

Korelasyon katsayısı önemli ve anlamlı bulunduğu kullanılır.

Bir bağımsız ve bir bağımlı değişken arasında doğrusal regresyon kurulduğunda korelasyon katsayısının karesi belirtme katsayısına eşittir (Alpar, 1997:173).

$$r^2 = R^2$$

- **Çoklu Korelasyon Katsayısının Testi**

Çoklu korelasyon katsayısının testinde aşağıdaki aşamalar takip edilir:

1. Hipotezin yazılması aşaması

$H_0; \rho_{y,x_i} = 0$ Anakütlede bağımsız değişkenler birlikte bağımlı değişkeni etkilememektedir.

$H_0; \rho_{y,x_i} \neq 0$ Anakütlede bağımsız değişkenler birlikte bağımlı değişkeni etkilemektedir.

2. α anlamlılık düzeyinin belirlenmesi aşaması

3. F (varyans analizi) istatistiğinin hesaplanması aşaması

f_1 ve f_2 serbestlik dereceleri olmak üzere;

$$F = \frac{\text{Açıklanan değişkenlik} / f_1}{\text{Açıklanamayan değişkenlik} / f_2} \quad F = \frac{\frac{r^2 y, x_i}{f_1}}{\frac{(1-r^2 y, x_i)}{f_2}}$$

4. $F_{\alpha; f_1 - f_2}$ tablo değeri $\geq F$ ise H_0 kabul (ilişki ana kütlede anlamsızdır)

$F_{\alpha; f_1 - f_2}$ tablo değeri $< F$ ise H_0 red (ilişki ana kütlede anlamlıdır)
(Orhunbilge, 2010:143).

Bağımlı ve bağımsız değişkenler arasındaki ilişki düzeyi Tablo 2.1.'de gösterilmiştir:

Tablo 2.1. Korelasyon Aralığına Göre İlişki Düzeyi

Korelasyon Aralığı	İlişki Düzeyi
(-0,25) - 0,00 ve 0,00 - 0,25	İlişki Yok
(-0,49) - (-0,26) ve 0,26 - 0,49	Zayıf
(-0,69) - (-0,50) ve 0,50 - 0,69	Orta
(-0,89) - (-0,70) ve 0,70 - 0,89	Yüksek
(-1,00) - (-0,90) ve 0,90 - 1,00	Çok Yüksek

2.7.2. Karar Ağaçları

Sınıflandırma ve regresyon modellerinde kullanılan başlıca tekniklerden biri olan karar ağaçları, ağaca benzeyen yapısı ile kökünden yapraklarına doğru kurallar dizini içeren bir akış diyagramıdır (Chien, Wang ve Cheng, 2006:57).

Karar ağaçlarının anlaşılabilir kurallar üretebilmesi, aşırı hesaplamaya gerek kalmadan sınıflandırma yapabilmesi, hem sürekli ve hem de kesikli değişkenler için uygun olması, sınıflandırma ve kestirim için hangi alanların en önemli olduğunu açık bir şekilde gösterebilmesi karar ağaçlarının güçlü yönleri olarak bilinir (Oğuzlar, 2001:80).

Kurallar genellikle bilginin ifade edilmesinde kullanılır. Kural öğrenme, eğitim örnekleri kümesinden kural türetme sürecidir. Kural öğrenmede birçok algoritma, bir örneği daha önceden belirlenmiş sınıflardan birisine atamak için kural araştırması yapar. Bu araştırmayı yaparken Chaid, Quest, Cart, ID3 ve C5.0 gibi algoritmaları kullanırlar (Wong ve Leung, 1999:66).

Karar ağacının en fazla kullanıldığı uygulamalar şunlardır:

- Demografik grupların hangilerinin mektup aracılığıyla yapılan pazarlama uygulamalarında yüksek cevaplama oranına sahip olduğunun belirlenmesi.
- Kredi geçmişlerinin kullanılmasıyla bireylere ilişkin kredi kararlarının verilmesi,
- İşletmeye en faydalı olan bireylerin özelliklerinin kullanılmasıyla işe alma süreçlerinin belirlenmesi,
- Tıp ile ilgili gözlem verilerinden hareketle en etkin kararların verilmesi,
- Satışları hangi değişkenlerin etkilediğinin belirlenmesi,

- Ürün hatalarına yol açan değişkenlerin belirlenmesi (Oğuzlar, 2001:80).

Karar ağaçları sınıflandırma ve kestirim için güçlü ve popüler araçlardır. Karar ağaçlarının çekici olan yönü, bir takım kuralları temsil etmesidir. Kurallar, karar ağacından kolaylıkla okunabilir. Karar ağaçları, kuruluş fiyatlarının uygunluğu, kolay yorumlanabilmeleri, veritabanı sistemlerine kolayca entegre edilebilmeleri ve güvenilirlikleri nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahip tekniktir (Bilekdemir, 2010:24).

Veri miktarının çok büyük olduğu durumlarda eğer veri çok sayıda değişkene sahip, karmaşık ve doğrusal olmayan yapıda ise, değişkenler arasında birliktelik ve ilişkileri bulmak, çıktıları tahmin etmek gerekiyorsa problemlerin çözümünde veri madenciliği tercih edilmektedir (Han, Kamber ve Pei, 2012:89)

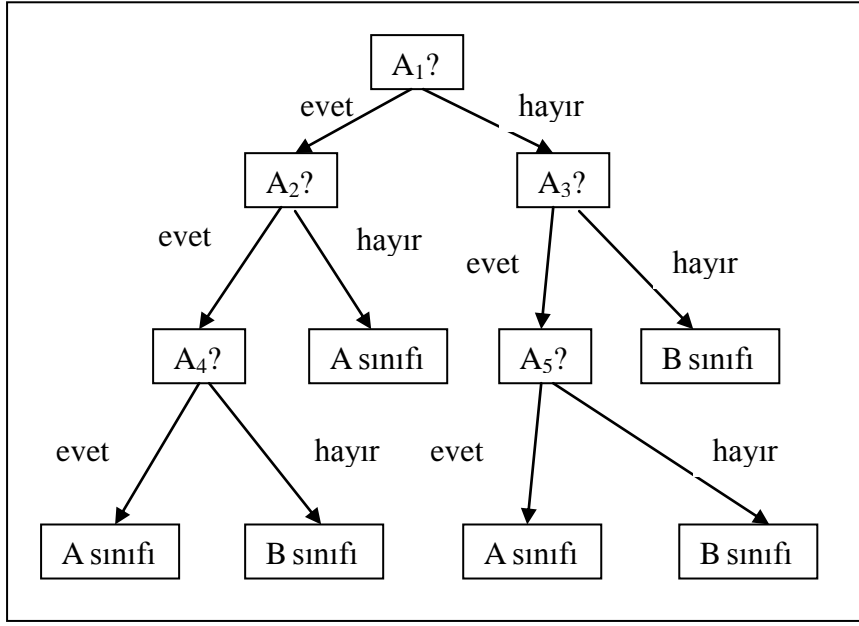
2.7.2.1. Karar Ağaçlarında Dallanma ve Budama Kriterleri

Karar ağacı; karar düğümleri, dallar ve yapraklardan oluşur. Karar düğümü, gerçekleştirilecek testi belirtir. Bu testin sonucu ağacın veri kaybetmeden dallara ayrılmasına neden olur. Her düğümde test ve dallara ayrılma işlemleri ardışık olarak gerçekleşir ve bu ayrılma işlemi üst seviyedeki ayrımlara bağlıdır. Ağacın her bir dalı sınıflama işlemi tamamlamaya adaydır. Eğer bir dalın ucunda sınıflama işlemi gerçekleşmiyorsa, bir karar düğümü oluşur. Ancak belirli bir sınıf oluşuyorsa, o dalın sonunda yaprak vardır. Bu yaprak, veri üzerinde belirlenmek istenen sınıflardan biridir. Karar ağacı işlemi kök düğümünden başlar ve yukarıdan aşağıya doğru yaprağa ulaşana dek ardışık düğümleri takip ederek gerçekleşir (Bilekdemir, 2010:24).

Karar ağaçlarındaki en büyük sorunlardan birisi herhangi bir kökten itibaren dallanmanın hangi kriterlere göre yapılacağıdır. Aslında her farklı kriter için bir karar ağacı algoritması karşılık gelir (Özkan, 2010:49).

Budama ağaçta oluşmuş fakat sınıflamaya hiçbir etkisi olmayan dalların ağaçtan çıkarılması işlemidir. Budama işlemi ağaç kurulurken yapılabileceği gibi kurulduktan sonra da yapılabilir. Budama işleminin hangi ölçüte göre yapılacağı önemlidir. Algoritmaların çoğunda varsayılan değerden daha düşük anlamlılık gösteren değerler budanır (Silahtaroglu 2008:51).

Şekil 2.4'te budanmamış karar ağacı örneği gösterilmiştir:



Şekil 2.4. Budanmamış Karar Ağacı Örneği

Uyarlandığı Kaynak: Han, Kamber ve Pei, 2012, s.345.

2.7.2.2. Bazı Karar Ağacı Algoritmaları

Karar ağaçlarına yönelik geliştirilmiş birçok algoritma vardır. Bu algoritmalar birbirlerinden kök, düğüm ve dallanma kriteri seçimlerinde izledikleri yol bakımından ayrılırlar (Silahtaroglu 2008:51).

• Chaid Algoritması

1980 yılında Kaas tarafından geliştirilmiştir. Tahmin değişkenleri hedef değişkene uyan bir çiftin içinde istatistikî olarak anlamlı bir fark kalmayınca kadar birleştirilmesiyle en iyi sınıfın hesaplanması tekniğine dayanır. En iyi sınıfın seçilmesinde Ki-kare (Chi-square) testi kullanılır (Albayrak ve Yılmaz, 2009:31-52).

Bu analiz sonucunda bağımlı değişken üzerine etkisi istatistiksel olarak önemli bulunan bağımsız değişkenler içinde en yüksek F değerine sahip olan değişken, Chaid algoritması diyagramında ilk sırayı almaktadır (Koyuncugil, 2007:1-17).

Chaid, Ki-kare metriği vasıtasıyla ilişki düzeyine göre farklılık rastlanan grupları ayrı ayrı sınıflamakta ve ağacın yaprakları, ikili değil, verideki farklı yapı sayısı kadar dallanmaktadır.

Chaid algoritması, sürekli ve kategorik tüm değişken tipleri ile çalışabilmesi nedeni ile yaygın olarak kullanılmaktadır. Bu yöntemde, sürekli tahmin edici

değişkenler bu uygulama tarafından otomatik olarak analizin amacına uygun olarak kategorize edilmektedir.

Chaid ile diğer yöntemler arasındaki en önemli farklılıklarından birisi, ağaç türetimidir. ID3, C4.5 ve CART ikili ağaçlar türetirken, Chaid ikili olmayan çoklu ağaçlar türetir

Chaid algoritmasının kullanılmasının avantajlarından biri kategorik (nominal) ve sürekli olan değişken tipleriyle çalışılabilir olması ve kategorik değişkenler için gözlenen sıklık değerlerinin X^2 analizini yaparak bu değişkenlerin modele ne kadar iyi cevap verebileceğine karar verebilmesidir (Mattison, 1997:254).

Chaid algoritması çalışmalarında kategorik bağımsız değişkenler tercih edildiğinden, modele giren bağımsız değişkenlerin sürekli olmaları durumunda bu değişkenleri bölerek kategorik hale getirir. Eğer bağımsız değişken çok fazla kategoriye sahip ise, bu durumda da kategori sayısı indirgenerek ağacın basitleştirilmesi yoluna gidilir (Çelik, 2009:64).

Chaid algoritması, tahmin edici değişkenin tüm değerlerini dikkate alarak işlem yapan bir analiz tekniğidir. Hedef değişkene benzeyen değişkenleri birleştirir ve farklı olan değişkenler ile işleme devam eder. Bir sonraki aşamada karar ağacının ilk dalını oluşturmak için en iyi tahmin değişkenini seçer. Ardışık tüm düğümler tamamlandığında karar ağacı da tamamlanmış olur (Gürsoy, 2009:76).

Chaid algoritmasının işleyişi aşağıda aşamalı olarak ifade edilmiştir.

1.Adım: Her bir X bağımsız değişkeni için, X' in, Y karar değişkenini dikkate alan en az öneme sahip (en büyük p değerine sahip) kategori çifti bulunur. Yöntem, Y'nin ölçüm düzeyine bağlı olarak p değerlerini hesaplar.

a. Bağımlı değişken Y sürekli ise F testi kullanılır.

b. Eğer Y isimsel ise, X' in kategorileri satırlarda ve Y' nin kategorileri sütunlarda olacak biçimde iki yönlü çapraz tablo düzenlenir. Pearson ki-kare testini veya olabilirlik oranı testi kullanılır.

c. Eğer Y sıralı ise, bir Y birliktelik modeli oluşturmak için olabilirlik oran testi kullanılır.

2.Adım: En büyük p değerine sahip X' in kategori çifti için, p değerini önceden belirlenmiş alfa düzeyi $\alpha_{\text{birleş}}$ ile kıyaslanır.

a. Eğer p değeri $\alpha_{\text{birleş}}$ ' den büyük ise X' in kategori çifti bir tek kategori altında birleştirilir. X'in yeni kategori kümesi için süreç Adım 1'den yeniden başlatılır.

b. Eğer p değeri $\alpha_{\text{birleş}}$ ' den küçük ise Adım 3'e gidilir.

3.Adım: X' in ve Y' nin kategori kümesi için uygun Bonferroni çarpanı kullanılarak, düzeltilmiş p değeri hesaplanır.

4.Adım: En küçük düzeltilmiş p değerine sahip en önemli X tahmin edici değişkeni seçilir. X' in p değeri önceden tanımlanmış alfa düzeyi $\alpha_{\text{böl}}$ ile kıyaslanır.

a. Eğer p değeri, $\alpha_{\text{böl}}$ değerinden küçük veya eşit ise düğüm X' in kategori kümesi temel alınarak bölünür.

b. Eğer p değeri, $\alpha_{\text{böl}}$ değerinden büyük ise düğüm bölünmez. Bu düğüm uç düğümdür. Ağaç büyütme süreci durma kuralları görülene kadar sürdürülür (Demirel, 2010:59).

• CART (Clustering and Regression Tree) Algoritması

CART, karar ağacı algoritmalarının içinde en fazla kullanılanlardan birisi olup 1984 yılında Breiman ve arkadaşları tarafından önerilen bir algoritmadır. Sınıflandırma ve regresyon analizi için kullanılan bir algoritmadır. CART bağımlı değişken kategorik iken sınıflandırma modeli, sürekli iken ise tahminleme modeli kuran bir algoritmadır (Larose, 2005:109).

CART, ID3 algoritmasında olduğu gibi en iyi dallara ayırma ölçütlerini seçmek için entropi tekniğini kullanır (Dunham, 2003:8).

Tüm algoritmalarda olduğu gibi CART algoritması için de amaç en doğru modeli kurabilmektir. Modelin doğruluğu, bağımlı değişkenin kategorik olması durumunda doğru tahmin edilen kayıtların oranı, sürekli olması durumunda ise ortalama hata kareler yöntemi ile belirlenir.

CART veriyi iki alt kümeye ayırarak oluşturulan karar ağacının daha homojen olmasını sağlar. CART algoritması karmaşık bir algoritma olup verilerin büyüklüğüne göre süreç de uzayabilir (Gürsoy, 2009:75).

• ID3 Algoritması

ID3 Algoritması Quinlan tarafından geliştirilmiştir. Temeli makine öğrenmesi ve bilişim teorisine dayanır.

ID3 Algoritması en ayırıcı özelliğe sahip değişkeni ararken eldeki bilgileri sayısallaştırma yolunu seçer yani entropi kavramından yararlanır. Entropi kavramının amacı bir veri kümesi içerisindeki belirsizliği ve rastgeleliliği ölçmektir. Entropi, bir sistemdeki belirsizliğin ölçüsüdür (Dunham, 2003:77).

ID3 algoritması, bilgi kazancı en fazla olan karakteristiği seçerek bu karakteristiği ilk aşamada ağacın kökü, diğer aşamalarda ise bir düğüm (alt küme için kök) olarak alır ve buna göre karar ağacını dallandırır (Quinlan, 1993).

• C4.5 Algoritması

ID3 karar ağacı algoritmasının bir uzantısı olan C4.5, bir diğer popüler karar ağacı modelleme sistemidir. C4.5 algoritması, sayısal değerli niteliklerle de karar ağacı oluşturulabilmesine imkan sağlamaktadır. Bununla birlikte bilinmeyen nitelikteki örnek kümeleri ile oluşturulacak karar ağaçları için de yol gösterir (Özkan, 2010:76).

C4.5 Algoritması, CART Algoritmasından farklı olarak, iki dallı karar ağacı ile sınırlı değildir. CART ikili bir karar ağacı yapısı kurarken, C4.5 Algoritması, çok daha fazla değişken bir yapıya sahip karar ağaçları kurabilir (Larose, 2005:116).

ID3 algoritması karar ağacının her düğümü için bilgi kazancı değerlerini hesaplar, C4.5 ise bilgi kazancı ile beraber alt sette yer alan karakteristiklerin bilgi kazanç oranlarını da hesaplayarak bilgi kazanç oranı en yüksek olan karakteristiği düğüm noktası olarak seçer.

• QUEST Algoritması

İkili büyüyen bir karar ağacı algoritması olan QUEST algoritması ayrı ayrı değişken ve ayırım noktası seçimi ile ilgilenir. Birim değişken ayırıcı tarafsız değişken seçimini tahmini olarak yapar. İkili dal yapısına sahip olduğu için karar ağacı çok büyük olabilir (Gürsoy, 2009:76).

• SLIQ Algoritması

SLIQ (Supervised Learning In Quest – Araştırmada Denetimli Öğrenme) Algoritması hem sayısal hem de kategorik verilerin sınıflandırılmasında kullanılabilen

önce genişlik ilkesine göre hareket ederek aynı anda birçok karar ağacı yaprağı oluşturabilir. Bu algoritma hızlı olmasının yanı sıra çok iyi sonuçlar veren karar ağaçları da üretebilir. Ayrıca bellekte tutulması güç olan çok büyük verileri, belleğe almadan bir kerede tek bir ağaç olarak sınıflandırabilir. ID3 ve C5 gibi algoritmalar önce derinlik ilkesine göre çalışırken, Sliq algoritması önce genişlik ilkesiyle hareket ederek aynı anda birçok yaprak oluşturabilir (Silahtaroglu, 2008:58).

2.7.3. Yapay Sinir Ağları

Yapay Sinir Ağları (YSA); sınıflandırma, tahminleme ve kümeleme için kullanılabilen genel amaçlı ve güçlü araçlardır. Oldukça geniş bir çerçevede uygulama alanına sahip olan YSA, süreçlerin girdileri ve çıktıları arasındaki karmaşık ilişkileri modelleyebilme özelliğinden dolayı çeşitli endüstriyel süreçlerin modellerini kurmak ve tahminlerde bulunabilmek için uygulanır. Yapay Sinir Ağlarının güçlü bir teknik olması, geniş bir alandaki problemlere uygulanabilmesiyle kategorik ve sürekli değişkenleri ele alabilmesinden kaynaklanır (Emel, Taşkın ve Kılıçarslan, 2004:205).

Yapay sinir ağları kavramı beynin çalışma ilkelerinin bilgisayarlar üzerinde kurgulanması mantığı çerçevesinde oluşmuş ve ilk çalışmalar beyindeki nöronların matematiksel olarak modellenmesi üzerine yoğunlaşmıştır (Efe, 2011:89)

Yüksek oranda parametrelerle ifade edilen istatistiksel sınıflandırma modellerinden biri olan YSA bu özelliklerinden ötürü oldukça esnekler ve bu yüzden fonksiyonlardaki nispeten küçük aykırılıkları doğru bir şekilde modelleyebilirler. Diğer yandan böylesi bir esneklik ciddi bir aşırı-uyum (Over Fitting) tehlikesini de beraberinde getirir (Han, Kamber ve Pei, 2012: 399-400).

Doğrusal olmayan problemlerde tahmin yapma açısından istatistik uygulamalara göre daha kolay ve doğru sonuçlar veren YSA, biyolojik sistemin üstünlüklerini de modelleyebilmesi sebebiyle karmaşık problemlerin çözümünde de sıkça kullanılmaktadır. Klasik problem çözme yöntemlerinin aksine YSA'da çok sayıda problem çok hızlı bir şekilde çözülebilmektedir. Aynı anda çalışabilme özelliğinin yanı sıra birbirinden bağımsız olarak da çalışabilen yapay sinir hücrelerinden herhangi birindeki bir sorun ağın geri kalanını etkilememektedir (Esen, 2009:68).

Bir YSA modeli; birbirleriyle bağlantılı olan sinirlerin bulunduğu genel olarak giriş katmanı, gizli katman ve çıkış katmanı olarak adlandırılan üç bölümden oluşur.

Giriş değişkenleri giriş katmanından YSA modeline girer. Giriş katmanından geçerek çıktı katmanına iletmek ile görevli olan gizli katmana geçen bu değişkenler buradan da son katman olan çıktı katmanına iletilirler.

2.7.4. Bayes Sınıflandırma Algoritması

Naive Bayes algoritması tüm ölçütlerin sonuç üzerindeki etkilerinin olasılıksal olarak hesaplanması temeline dayanan bir veri madenciliği modelidir. Naive Bayes algoritmasında bir sonucun çıkma olasılığı o sonucu etkileyen tüm faktörlerin o sonucu sağlama olasılıklarının çarpımıdır (Bilekdemir, 2010:34).

Bayes sınıflandırıcılar istatistiksel sınıflandırıcılardır ve belirli bir değişkenler grubunun belirli bir sınıfa ait sınıf üyeliği olasılıklarını tahmin ederler. Bu sınıflandırma Thomas Bayes'in teoremine dayanmaktadır (Han, Kamber ve Pei, 2012:350).

Bir sınıflandırma probleminde amaç belirli kestirimci değişkenler kümesi verildiği durumda, her sınıfa ilişkin üyelik olasılıklarını tahmin etmektir. Bu tür bir olasılık koşullu olasılık olarak adlandırılır. Y olayının verilen X olayına göre koşullu olasılığı $P(Y|X)$ sadece X olayının gerçekleştiği durumlarda Y olayının da gerçekleşmesi olasılığını temsil eder. X ve Y' nin bağımsız rastsal değişkenler çifti olduğu düşünülürse, bunların bileşik olasılıkları $P(X = x, Y = y)$ ' dir. Bunların koşullu olasılığı ise bir rastsal değişkenin diğer rastsal değişkenin değerinin bilindiği durumda vereceği sonuç olarak tanımlanabilir. Örneğin $P(Y = y | X = x)$ koşullu olasılığı Y değişkeninin, X 'in x değerini aldığı gözlemlendiği bir durumda, y değerini alacağı durumdur. X ve Y 'nin bileşik ve koşullu olasılıkları aşağıdaki biçimde ilişkilendirilebilir (Tan vd, 2006:228, Irmak 2009:32).

$$P(X, Y) = P(Y | X) \times P(X) = P(X | Y) \times P(Y)$$

Yukarıdaki formülde son iki ifadenin yeniden düzenlenmesi ile Bayes teoremi olarak bilinen aşağıdaki formül elde edilir:

$$P(Y | X) = \frac{P(X | Y) \times P(Y)}{P(X)}$$

2.7.5. Genetik Algoritmalar

Genetik algoritmalar çok değişkenli fonksiyonların optimizasyonu amacıyla veya sınıflandırma için kullanılan araçlardır. Geleneksel optimizasyon yöntemlerine

göre farklılıkları olan genetik algoritmalar, parametre kümesini değil kodlanmış biçimlerini kullanırlar (Yavuz, 2009:78).

GA, biyoloji biliminden esinlenilerek geliştirilmiş makine öğrenimi yöntemlerinden birisidir. Değişik tiplerdeki verileri işleme özelliğine sahip olan genetik algoritmalar optimizasyon amacı ile de kullanılabilirler. Olumlu yönlerine rağmen genetik algoritmaların kullanımlarında bazı sıkıntılar da yaşanabilmektedir. Bunlardan en belirgin olanı karmaşık sorunların genetik kodlamasının zorluğu ve en iyi sonucun üretildiğine dair bir garantinin de verilememesidir (Dolgun, 2006:62).

Genetik algoritmalar, veri yığınlarındaki örüntüleri bulmak için tek başlarına kullanılmazlar. Örneğin sınır ağları gibi öğrenme temelli veri madenciliği algoritmalarına rehberlik etme amaçlı kullanılabilirler (Çetin, 2009:80).

2.8. VERİ MADENCİLİĞİ YAZILIMLARINDAN BAZILARI

Çoğu yazılım sadece özelleştirilmiş sistemlere cevap verebilmek amacıyla oluşturulmuştur. Veri madenciliğinin amacı daha önceden bilinmeyen ilişkileri aramak ve analiz için uygun yöntemleri karşılaştırmaktır. Dolayısıyla, özelleştirilmiş sistemler, çözüm için uygun değildir. Geçerli veri madenciliği yazılımı, farklı teknikleri kullanıp karşılaştırabilen, karmaşık veri tabanı yönetim yazılımları ile bütünleşebilen bir veri madenciliği modeli geliştirmelidir (Giudici, 2003:11). Aşağıda veri madenciliği yazılımlarından bazıları verilmiştir:

2.8.1. IBM SPSS Statistics 19.0

IBM SPSS Statistics, temel iş ve araştırma sorunlarına yanıt vermek açısından iş yöneticileri ve analistler için gerekli olan temel istatistik yordamlarını sunar. Bu yazılım, kullanıcının verileri hızlı bir şekilde görüntülemesini, ek testler için hipotezleri formüle etmesini ve değişkenler arasındaki ilişkileri netleştirmek, kümeler oluşturmak, eğilimleri belirlemek ve tahminler yürütmek üzere yordamları yürütmesini sağlar. IBM SPSS Statistics 19.0 aşağıdaki temel yetenekleri içerir:

- **Doğrusal modeller** karmaşık ilişkileri tanımlayan verilerin doğal özelliklerine uygun olacak şekilde tasarlanan bir dizi regresyon ve gelişmiş istatistiksel yordamlar sunar.
- **Doğrusal olmayan modeller** verilerinize daha gelişmiş modeller uygulama yeteneği sağlar.

- **Özelleştirilmiş tablolar** kullanıcıların verileri kolayca anlamasını ve sonuçlarını farklı hedef kitleler için farklı biçimlerde hızla özetlemesini sağlar (<http://www.spss.com>).

2.8.2. IBM SPSS Modeller (Clementine 11.1)

Çok kullanılan bir veri madenciliği aracı olan IBM SPSS Modeller, CRISP-DM modeli çerçevesinde tahminleme modellerini geliştirebilen bir veri madenciliği yazılımıdır. Clementine, karar ağaçlarının yanı sıra, faktör, regresyon analizi gibi analizleri yapabilen, yapay sinir ağlarını kullanabilen ve veri tabanına rahatlıkla bağlanabilen bir araçtır (Özyirmidokuz, 2009:70).

Clementine, müşteri ilişkileri yönetimi, kimya sektöründe maddelerin aşındırıcılık tahmininde ve bankacılık alanında kredi kartı dolandırıcılıkları gibi konularda kendine uygulama alanı bulmuştur.

IBM SPSS Modeller (Clementine) programı görselliğe önem verilerek tasarlanmıştır. Çalışma ekranında sürükle bırak ile nesne yerleştirme ve nesnelere birbirine bağlama işlemleri kolaylıkla yapılabilmektedir. Clementine, veri madenciliği aşamalarını gelişmiş bir teknoloji ile gerçekleştirme imkânı sunmaktadır.

Clementine, SPSS in farklı veri madenciliği projelerine dair ciddi bir iş deneyimini kullanıcılarına aktarmayı amaçlayarak hazırlanmış bir programdır (Şekil 2.5). Verinin anlaşılması aşamasında veri kaynaklarına bağlanma, veriyi tanıma, verinin kalitesini anlama ve verinin grafiksel olarak incelenmesi, hipotezleri oluşturma amaçlı veri gruplarını değerlendirme çalışmalarında Clementine grafikler ve tablolar belli bölgelerin seçimini yapma seçeneği sunmaktadır. Clementine içerisinde yer alan histogram, çizgi grafik (line plot), nokta grafiği (point plot), istatistik (statistics), dağılım grafikleri (distribution graphs), veri çıktısı (data audit) işlemcileri verinin ön incelemesinde sıkça kullanılan işlemcilerden bazılarıdır (Şekil 2.6) (Şekeroğlu, 2010:88).

Clementine içerisinde tahmin edici modellerden;

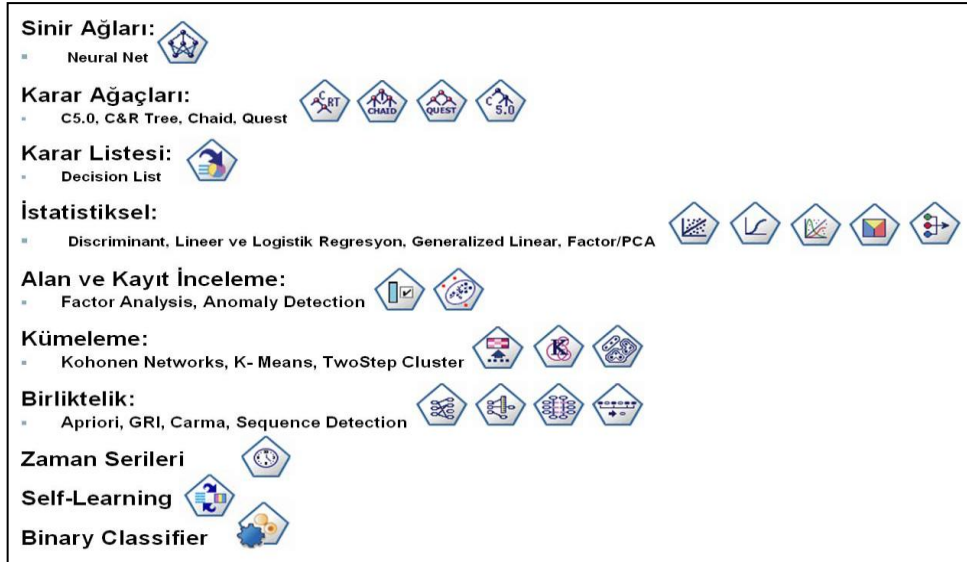
- Sinir Ağları,
- Karar Ağaçları: C5.0, CART, QUEST, CHAID,
- Regresyon,
- Lojistik Regresyon,

- Ardışıklık Tespiti Yöntemleri,
- Kümeleyici modellerden;
- Kahonen Ağları,
- K-Ortalama (K-Means),
- İki Adımlı Kümeleme Yöntemleri,

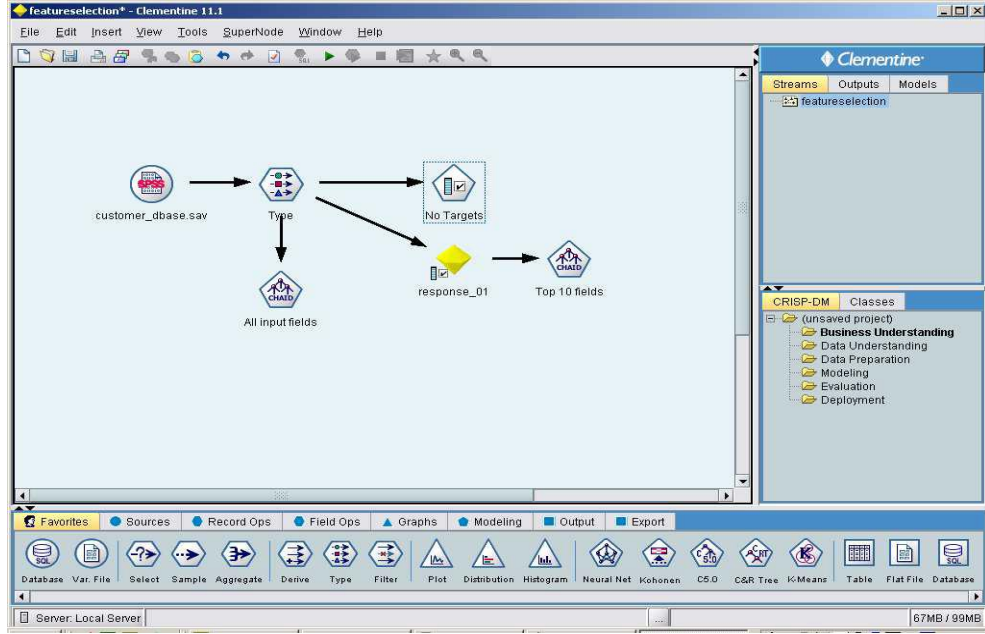
Birliktelik modellerinden ise

- Apriori,
- Gri yöntemleri için gerekli araçlar bulunmaktadır (Ataseven, 2009:95).

Şekil 2.5 ve 2.6'da Clementine tuvalinde gözükten veri madenciliği modelleri ve kullanıcı ara yüzü verilmiştir.



Şekil 2.5. Clementine'de kullanılan Veri Madenciliği Modellerinin Simgeleri



Şekil 2.6. Clementine 11.1 Kullanıcı Ara yüzü

Diğer veri madenciliği yazılımlarından bazıları aşağıda gösterilmiştir:

• Weka

WEKA, Yeni Zelanda'daki University of Waikato tarafından Java programlama dili kullanılarak oluşturulmuş, içerisinde birçok makine öğrenmesi algoritması barındıran bir veri madenciliği programıdır. WEKA'nın asıl kullanım alanı birçok makine öğrenmesi algoritmasının uygulandığı sınıflandırma problemi olmasına rağmen bünyesinde çeşitli kümeleme, birliktelik kuralları ve regresyon analizi algoritmaları da barındırmaktadır (Özçınar, 2006:83).

• DB Miner

Çevrimiçi analitik işleme (OLAP) yeteneğini VM algoritmalarıyla birleştirebilme özelliği ile ön plana çıkan DB Miner, OLAP ve VM yöntemlerini dinamik bir şekilde seçebilme imkânına sahiptir. Ayrıca kullanıcıların kolay kullanabileceği bir ara yüze sahiptir. DB Miner'in diğer genel amaçlı VM araçlarına göre bir avantajı, geliştirilen DMQL (Data Mining Query Language) dilini kullanabiliyor olmasıdır. DMQL, SQL benzeri bir VM sorgulama dilidir (Bozkır, 2009:101).

• Data Logic / R

Kümeleme ve sınıflama analizleri için kullanılan ticari bir veri madenciliği aracı olan DataLogic/R artık nitelik ve verilerin temizlenmesi işlemlerini yapabilmektedir. Sistemin en güçlü tarafı üretilen kuralların öğrenme-test geçerliliği ve güvenlik gibi kriterlerde değerler üretebilmesidir. Bu değerler üretilen kuralların kalitesini belirlemek için kullanılmaktadır (Çil, 2010:65).

2.9. LİTERATÜR ÇALIŞMASI

1995 yılında, King vd. çalışmalarında, çeşitli veri kümeleri arasında hangi algoritmaların daha iyi sonuçlar ürettiğini incelemek amacıyla, karar ağaçları algoritmaları (CART, C4.5, NEWID, ITrule, Cal5), istatistiksel algoritmalar (NaiveBayes, K-en yakın komşuluk, doğrusal diskriminant, kuadratik diskriminant, lojistik regresyon, Bayes Ağları) ve yapay sinir ağları (geri yayılım, merkez tabanlı uzaklık fonksiyonu) metotları arasında karşılaştırma yapmışlardır. Daha iyi sonuçlar üreten algoritmanın, üzerinde araştırma yapılan veri kümesine bağlı olduğu sonucuna ulaşıldığı çalışmada örnek olarak ikili değerli niteliklerin % 38 oranının üzerinde olduğu veri kümelerinde sembolik öğrenme algoritmalarının daha verimli sonuçlar ürettiğini belirtmişlerdir (King, Feng ve Sutherland, 1995:289-333).

1999 yılında, Özden vd. çalışmalarında, plastik sanayisinde reçine üretiminde kullanılan enjeksiyonlu döküm işleminde ürün kalitesini arttırmak amacıyla veri madenciliği yöntemlerini kullanmışlardır. Ürün ve üretim tasarımı, işlem sorunları, ürün kalitesi ve sağlamlığı için sinir ağları modelleri oluşturmuşlardır. Model oluşturmada kullanılan değişkenler; sıcaklık, tutunma basıncı, enjeksiyon hızı, soğutma zamanı, döküm sıcaklığı ve sıcaklık değişimidir. Değişkenlerin azaltılması amacıyla C4.5 ve CART karar ağaçlarının kullanıldığı çalışmanın sonucunda oluşturulan sinir ağları modellerinin ürün kalitesini arttırıcı bir yöntem olduğu belirtilmiştir (Özden ve Chen, 1999:17).

1999 yılında, Mieno vd. çalışmalarında, yarı iletken üretiminde ortaya çıkan hata nedenlerini belirleyebilmek için veri madenciliği yöntemlerinden olan regresyon analizini kullanmışlardır. Değişkenler olarak fırın numarası, transistör, eşik gerilimi ve geçit voltaj değerlerinin kullanıldığı çalışmada üretimde meydana gelen hataların elektriksel olduğu düşünülerek değişkenler incelenmiş ve elektriksel hatalar analiz

edilmiştir. Bu yöntem, normalde işlemin takibi esnasında görülemeyen hata nedenlerinin altı saat içinde otomatik olarak belirlenebilmesini sağlamıştır (Mieno, Sato, Shibuya, Odagiri, Tsuda ve Take, 1999:391-394).

2002 yılında, Skinner vd. çalışmalarında, levha üretim işleminin levhaların kalitesi ve üretim verimindeki etkisini incelemiş, düşük verimin nedenini belirlemede iki çeşit geleneksel çok değişkenli istatistiksel yöntem CART ile karşılaştırılmıştır. Çalışmada kullanılan değişkenler; lot ve levha numarası, elektriksel parametreler, belli sıcaklıktaki kapasitans ve direnç, ve ortalama levha verimidir. Çalışma sonucunda CART'ın düşük verimi engelleyici bir yöntem olduğu rapor edilmiştir (Skinner, Montgomery, Runger, Fowler, McCarville ve Rhoads, 2002:523-530).

2002 yılında, Deng vd. çalışmalarında, çelik tabaka üretiminde mekanik özelliklere etki eden faktörlerin ayarlanarak ürün kalitesini arttırmak ve maliyeti azaltmak için veri madenciliği yöntemlerinden SAS/EM yöntemini kullanmışlardır. Hedef değişkenler olarak germe ve gerilim kuvvetini almışlardır. Diğer değişkenler karbon, mangan miktarı ve soğutma sıcaklığıdır. Analiz sonucunda mangan ve karbon miktarının mekanik özellikler üzerine aynı etkiye sahip olduğu görülmüştür. Karbon fiyatının daha düşük olması nedeni ile üretimde karbon miktarı arttırılmış ve mangan miktarı azaltılmıştır. Böylece üretim maliyeti düşürülmüştür (Deng ve Liu, 2002:25-42).

2003 yılında, Li vd. çalışmalarında, kaplanmış cam üretiminin geliştirilmesi amaçlı, mekaniksel, kimyasal, elektriksel ve manyetik işlemlerdeki değişkenleri CART ve yapay sinir ağları metotları ile incelemişlerdir. Değişkenler olarak voltaj, güç, gaz basıncı ve dağılımı gibi makine ayarları alınmıştır. CART ve sinir ağları yaklaşımları, üretim hattındaki makine ve kalite ölçümleri arasındaki ilişkiyi modellemede ve cam kalitesini arttırmada umut vaat eden sonuçlar vermiştir (Li, Feng ve Sethi, 2003:731-734).

2004 yılında, Emel vd. çalışmalarında, çelik üretim sürecinin kalitesini önceden tahmin edebilmek amacıyla veri madenciliği tekniklerinden yapay sinir ağı modelini kullanmışlardır. Yapay sinir ağlarının kullanılması ile sürecin sonucunu önceden tahmin edebilme ve sonuca etkisi olan değişkenleri ve sonuç üzerindeki etkisini görebilme imkânı oluşmuştur. Çalışmada kullanılan değişkenlerden bazıları; örneklem numarası, döküm no, pota numarası, miktar, fırın kükürt miktarı ve hedef kükürt miktarıdır. Analizi yapılan sürecin ilgili girdi değişkenleri yapay sinir ağları ile eğitilerek üretim

çıktılarının kalitesini tahmin edecek bir model oluşturulmuştur. Yapay sinir ağı, test kümesinin % 80'ini doğru olarak sınıflamıştır (Emel, Taşkın ve Kılıçarslan, 2004:205-225).

2006 yılında, Çoban çalışmasında, üretim sektöründe faaliyet gösteren bir şirketin verilerini kullanarak işletmenin tedarikçi seçimine dair yaptığı modelleme çalışmasında Yapay Sinir Ağları, CART, Chaid ve Quest, kümeleme algoritmalarından ise K-Means ve Kohonen yöntemlerini kullanmıştır. Veri setini oluşturan değişkenlerden bazıları firma adı ve kodu, firmanın bulunduğu il, kazandığı ihale sayısı, kuruluş şekli, firma tipi, sektörü, kalite belgesidir. Analizlerde elde edilen etkinlik grafikleri incelendiğinde YSA algoritmasının başlangıçta, karar ağacı algoritmalarının ise süreç içerisinde daha etkin olduğu gözlenmiştir (Çoban, 2006).

2006 yılında, Özçınar çalışmasında, KPSS sınavına girecek öğrencilerin sınav sonuçlarını, veri madenciliği tekniklerinden yapay sinir ağları ve çoklu regresyon tekniklerini kullanarak önceden tahmin etmeye yönelik bir model oluşturmuştur. Modellerin ürettikleri hata değerleri farkının çok yüksek olamasa da yapay sinir ağı modelinin çalışmada kullanılan bütün veri kümelerinde regresyon analizi modelinden daha başarılı olması, Yapay Sinir Ağı tekniğinin öngörüye dayalı eğitim araştırmaları için klasik istatistik yöntemlere bir alternatif oluşturabileceğini göstermiştir (Özçınar, 2006).

2007 yılında, Rajavarman vd. çalışmalarında, geliştirdikleri genetik algoritma tabanlı sınıflandırma metodunu yaygın olan ID3, ID3 boosting ve Yapay Sinir Ağları algoritmaları ile karşılaştırarak sonuçlarını yayınlamışlardır. Üç farklı veri kümesi ile yaptıkları değerlendirme sonuçlarına göre kendi genetik algoritma sonuçlarının diğer algoritmalar karşısında daha başarılı olduğunu ifade etmişlerdir (Rajavarman ve Rajagapalan, 2007:555-561).

2007 yılında, Kurşun çalışmasında, işgücü yoğun olan özel bir hazır giyim işletmesinin gömlek dikim hattını ele alıp darboğaz noktalarını belirleyerek bu noktalara tezgâh, işçi ekleme çıkarma kararlarıyla hattın dengelenmesi ve alınacak yatırım kararlarına destek olacak öneriler sunulmasını amaçlamıştır. Bunun için öncelikle işletmede iş-zaman etüdü yapılmış, gerek kayıtlardan gerekse iş-zaman etüdünden elde edilen veriler bağımsızlık testi, teorik dağılımlara uygunluk testi ve Kolmogorov-Smirnov testi gibi istatistiksel testlere tabii tutularak elde edilen verilerle kurulan

modelin geçerliliği sınanmıştır. İkinci olarak üretim hattındaki darboğaz noktaları belirlenmiştir. Darboğazları yok etmek için olası senaryolar, çeşitli “what-if” analizleriyle yatırım kararlarını desteklemek için kullanılmıştır. Sonuç olarak sisteme entegre edilecek çeşitli sayılarda makine ve işçilerle bitirilen iş sayısının % 39 oranında, kuyrukta bekleyen iş sayısının % 45 oranında ve kuyrukta bekleme süresinin de % 64 oranında azaldığı belirtilmiştir (Kurşun, 2007).

2007 yılında, Onurlubaş çalışmasında, bitkisel yağ sanayisindeki gelişmeler, uygulamalar ve bunların sonuçlarına yer vermiştir. Üretim için oluşturulan modeller değişik matematiksel formlarla (üssel, kuadratik, lineer) denenmiş, bağımlı değişkendeki değişimin, bağımsız değişkenler tarafından açıklanan oranını ifade eden belirtme katsayısının (R^2) en yüksek olanı, hata terimlerinin büyüklüğü, parametrelerin büyüklük ve işaretleri, birlikte dikkate alınarak, doğrusal model kullanılmıştır. Ekonometrik analizde bağımlı değişken olarak üretim; bağımsız değişkenler olarak ise ihracat, ithalat, sabit sermayeye ilaveler, iç talep, işyeri sayısı, kukla değişkeni kullanılmıştır. Ekonometrik analizin sonucunda üretimi etkileyen en önemli değişkenlerin ithalat, iş yeri sayısı ve kukla değişkenleri olduğu görülmüştür. Çoklu regresyon analizi sonuçlarına göre, bitkisel yağ sanayinin üretimi, iç talebi, katma değeri, sabit sermayeye ilaveler, girdi, çıktı değerlerinin 2014 yılında artış göstereceği tahmin edilmiş ayrıca, sektörün gelişiminde ve kapasite kullanım oranlarının artırılmasında özellikle küçük ve orta ölçekteki işletmelerin desteklenmesi sektörün geleceği açısından önemli görülmüştür (Onurlubaş, 2007).

2008 yılında, Asilkan çalışmasında, ikinci el otomobillerin pazarda gelecekteki fiyatlarını tahmin etmek amacıyla, 2005–2007 yılları arasındaki verilerle regresyon analizi ve yapay sinir ağları tekniklerini kullanarak modeller oluşturmuştur. Her iki yöntemin tahmin doğrulukları karşılaştırıldığında, yapay sinir ağlarının regresyon tekniğine göre daha iyi sonuçlar verdiği sonucuna varılmıştır (Asilkan, 2008).

2009 yılında, Söylemez çalışmasında, SVM (Scorecard Veri Madenciliği), lojistik regresyon ve karar ağaçları tekniklerini kullanarak bir bankanın kredi değerlendirme sonuçlarını tahminleyici bir scorecard modeli kurmuştur. Bireysel müşterilerin scorecard analizlerinin gerçekleştirildiği bu çalışmada en iyi model SVM olarak bulunmuştur. Kredi almak isteyen müşterilere ait bilgilerin bu modele girilmesiyle kredilerini batırıp batırmayacakları büyük olasılıkla tahmin edilebilmiştir.

% 95,25 doğruluk oranına sahip SVM modelinden sonra % 82,83 oranıyla karar ağaçları ve % 76,57 oranıyla lojistik regresyon modeli gelmiştir (Söylemez, 2009).

2009 yılında, Çetin çalışmasında, bir üretim işletmesinde üretilen ürünlerin yanlış ayrılmasının nedenlerini belirlemek ve yapılan hataları azaltıcı stratejiler geliştirmek amacıyla karar ağaçları ve yapay sinir ağları tekniklerini kullanarak modeller geliştirmiştir. Vardiya düzeni, üretim periyodu, çalışanlar, üretim sıklığı, miktar, fabrika, makine arızası, ürün grubu, müşteri, çalışan kadro ve profil ve üretim türü karar değişkenleri olarak seçilmiştir. Tahminleme sonuçlarına göre karar ağacı teknikleri yapay sinir ağlarına oranla daha başarılı sonuçlar vermiştir. Hata türü tespitinde yapay sinir ağı % 43 gibi düşük bir oran ile doğru sonucu tahmin edebilmiştir (Çetin, 2009).

2009 yılında, Özyirmidokuz çalışmasında, halı imalat sürecini ve ürün kalitesini iyileştirmek amacıyla veri madenciliği tekniklerini uygulamıştır. Kullanılan değişkenlerden bazıları elektrik arızası, makine arızası, temizlik, bıçak hatası, eksik yün, apre, dokuma, eksik yün olup modellemede karar ağacı yöntemi olarak C4.5 algoritması ve yapay sinir ağları kullanılmıştır. Hata verisinden elde edilen karar ağacının doğruluk oranı % 74,44 olarak tespit edilmiştir. Makine duruş verileri ile yapılan ikinci uygulama çalışmasında % 99,24 doğruluk oranı ile bir yapay sinir ağı modeli geliştirilmiştir. Yapay sinir ağı modeline ait hata değerlerinin çok düşük olması, tekniğin bu gibi çalışmalar için klasik istatistiksel yöntemlere iyi bir alternatif olabileceğini göstermiştir (Özyirmidokuz, 2009).

2010 yılında, Coşkun çalışmasında, göğüs kanseri vaka kayıtlarını içeren SEER veri kaynağını üzerinden J48, Naive Bayes, Lojistik Regresyon ve KStar algoritmalarının karşılaştırmasını yapmıştır. Yapılan karşılaştırma sonucuna göre eldeki veriler üzerinden çalıştırılan J48 algoritmasının ürettiği modelin diğer algoritmalarından daha iyi karşılaştırma ölçütlerine sahip olduğu sonucuna ulaşılmış, ancak modeller arasında belirgin bir farklılık oluşmadığı da ayrıca vurgulanmıştır (Coşkun, 2010).

2010 yılında, Bilekdemir çalışmasında, su sayaçlarının üretim süresinin tahmininde karar ağaçları algoritmalarından C4.5 algoritması ile sınıflandırma tekniklerini kullanmıştır. Üretim süresini etkileyen faktörlerin hangi nedenlerden oluştuğu analiz edilmiş, gerekli önlemleri alarak teslim süreleri hakkında daha doğru kararların alınabileceği vurgulanmıştır. Üretim süresinin tahmininde performansın

arttırılabilmesi için makine verilerinin yanında işçiler ile ilgili verilerin de toplanması ve hesaba katılmasının gerekliliğine vurgu yapılan çalışmada tahmin için yapay sinir ağları, çizelgeleme problemlerinin çözümü için de genetik algoritmalar gibi başka veri madenciliği teknikleri kullanılmıştır. Bu çalışmada veri madenciliği tekniklerini kullanarak üretim süresinin tahmin edilebileceği gösterilmiştir (Bilekdemir, 2010).

2011 yılında Ülengin vd. çalışmalarında, Türkiye'nin rekabet gücünü belirleyen parametreler ile demir çelik sektörü arasındaki ilişkiyi analiz etmek amacıyla Bayes Nedensel Ağı tekniğini kullanmışlardır. Oluşturulan bu model ile demir çelik sektörü konusunda, uzmanların, değişkenlerin çeşitli durumları arasındaki nedensel ilişkileri görsel olarak irdeleyebilecekleri ve herhangi bir karar değişkeninin durumunda bir değişikliğe gidildiğinde bunun diğer değişkenlerde ne tür bir değişime sebep olacağını anında saptayabilecekleri bir model ortaya konmuştur. Çalışmada kullanılan bazı değişkenler yerel tedarikçi kalitesi, yerel tedarikçi yoğunluğu, teknolojinin firma düzeyinde benimsenmesi, genel altyapının niteliği, son teknolojilerin yaygınlığı, üretim süreci gelişmişliği ve liman altyapısının kalitesi olarak belirlenmiştir. Model sonucu, mevcut değişkenler üzerinde meydana gelecek olan değişiklikler ile sektörün uluslararası piyasadaki rekabet gücünü arttırma oranları göz önüne alınarak belirlenmiştir (Ülengin vd, 2011:25).

2011 yılında, Gu vd. çalışmalarında, AISI 1045 karbon çeliğinin performansını geliştirmek için uygulanan plazma yüzey sertleştirme işleminde, sertleştirme niteliklerini belirlemek amacıyla yapılan deney sonuçlarını kullanarak genetik algoritma-BP sinir ağlarından oluşan tahmin ve optimize edici bir model oluşturmuşlardır. Bu modeli oluşturmada sertleştirme hızı, plazma gaz akış hızı ve sertleştirme aralığı ana parametreler olarak seçilmiştir. Çalışma sonucunda, sertleştirme katmanları ve işlem parametreleri arasında doğrusal olmayan bir ilişki bulmuşlardır. Deneylerden elde edilen veriler çerçevesinde, tahmin edilen verilerin birbirine uyması nedeni ile önerilen Genetik Algoritma modelinin plazma yüzey sertleştirme işlemindeki sertleştirme niteliklerini tahmin ve optimize etmede güvenilir olduğunu belirtmişlerdir (Gu vd 2011:57-64).

2012 yılında, Çiftlikçi ve Özyirmidokuz çalışmalarında halı üretim işletmesinde üretilen halıların kalitesini arttırmak için hata nedeni, tarih, işçi, kalite, tasarım, renk, genişlik, yükseklik, hata cinsi gibi nitelikler üzerinde elde edilen değişkenlerde,

sınıflama ağaç modelleri için C4.5, hata nedenlerini tahmin etmek için ise C5.0 algoritmalarını kullanmışlardır. Karar ağaçları kullanılarak, işlemdeki hatalar bulunmuş ve hatalı ürünlerin üretimi engellenmiştir (Çiftlikçi ve Özyirmidokuz, 2012:1181-1200).

2012 yılında, Escribano vd. çalışmalarında, çeliğin haddeleme işlemindeki değişkenlerin ve parametrelerin hızlı ve kesin tahminlerinde kullanılabilecek modelleri elde etmek için sonlu eleman modeli (Finite Element Model) ve veri madenciliği yöntemlerini birleştirerek yeni bir model önermişlerdir. Tahmin modellerini oluşturmada kalınlık, silindir ve levha arasındaki sürtünme katsayıları, silindir çapı, açılmal hız ve dikey hareket gibi değişkenlerden faydalanmışlardır. Önerilen yöntemin, çeliğin bu işlemde kuvvet, tork, çıkış kalınlığı, en yüksek temas basıncı ve temas alanı gibi parametrelerinin tahmininde kullanılabileceğini belirtmişlerdir (Escribano, 2012:43-49).

2012 yılında, W. Hong-Bing vd. çalışmalarında, temel oksijen fırınında eritilmiş çelik içerisindeki fosfor miktarını tahmin etmek amacıyla, tahmin edici ve kümeleyici aşamaları içeren melez bir yöntem önermişlerdir. Önerilen bu yöntemi, GMDH (Group Method of Data Handling) polinomial sinir ağları ve BP sinir ağları (Back Propagation Neural Network) ile karşılaştırmışlardır. Kümeleri oluşturmak için ise homojen veriler ile ağırlıklı K-ortalamlarını kullanmışlardır. Her bir küme için GMDH uygulanmış ve melez bir yöntem oluşturulmuştur. Önerilen bu yöntemin fosfor miktarını belirlemede her iki yöntemle göre daha etkili ve üstün olduğunu belirtmişlerdir (Hong-Bing, 2012:11-16).

2012 yılında, Abhang vd. çalışmalarında, EN-31 çelik alaşımının metal kesme ve torna işlemi sonrası yüzey pürüzlüğünü tahmin etmek için çoklu regresyon ve yapay sinir ağları yöntemlerini kullanmışlardır. Mil hızı, besleme hızı, kesik derinliği ve takım uç yarıçapı değişkenlerinin kullanıldığı çalışmada sonuç olarak yapay sinir ağlarının deneysel verilerle daha uyumlu ve daha iyi sonuçlar verdiği bulunmuştur. Ayrıca bu yöntemin daha hızlı ve daha basit hesaplamalı olduğu da belirtilmiştir (Abhang ve Hameedullah, 2012:317-322).

3. UYGULAMA

3.1. ARAŞTIRMANIN YÖNTEMİ

Bu çalışmada, dünyada ve Türkiye’de demir çelik sektörüne konu olan ürünlerin gelişimine ve mevcut yapısına yer verilmiş olup, ülkemizin önde gelen entegre demir çelik tesislerinden birinde kuruluş tarihinden özelleştirildiği tarihe kadar üretimi yapılan ve sektörde uzun ürün olarak adlandırılan ara mamul ve nihai ürünlerin kaydını içeren veri seti üzerinde, veri madenciliği ön işleme ve sınıflandırma algoritmaları ile analizler yapılmıştır.

Uzun ürünlerin üretimini etkileyen faktörler veri madenciliği yöntemleriyle değerlendirilmiştir. Üretim miktarı, bağımlı değişken olarak; satış miktarı, yılbaşı ve yılsonu stok miktarları, ihracat ve ithalat miktarları, çalışan sayısı, lojistik miktarı, üretim kapasitesi, kapasiteden yararlanma oranı ve çelik üretim miktarı ise bağımsız değişkenler olarak belirlenmiştir.

Bağımsız değişkenlerin bağımlı değişken olan üretim miktarı üzerindeki etkisinin incelendiği çalışmada üretim temelli bir modelleme oluşturulmuş ve demir çelik tesislerine üretim noktasında bir yansıtım oluşturulmaya çalışılmıştır.

Modelleme çalışmasında ön işleme için veri temizleme, birleştirme, dönüştürme işlemleri Excel veri seti üzerinde, verilerin durumu uzmanlara danışılarak gerçekleştirilmiştir. İndirgenen veriye Clementine veri madenciliği yazılımındaki karar ağaçlarından Chaid algoritması ve IBM SPSS Statistics yazılımı ile de çoklu regresyon uygulanarak analizler gerçekleştirilmiştir. Bu analizler sonucunda; bağımlı değişken olan üretim miktarı, bağımsız değişkenlerin durumuna göre sınıflandırılarak, en uygun üretim miktarı tahmini için kurallar bulunması amaçlanmıştır. Elde edilen kurallar, demir çelik sektöründe uzun ürünler olarak nitelendirilen ürünlerin üretimlerine ilişkin önemli bilgiler vermekte ve yeni üretim faaliyetlerin planlanması noktasında karar desteği sağlar.

3.2. UYGULAMA ADIMLARI

3.2.1. Problemin Tanımlanması

Problemin tanımlanması aşamasında, ilk olarak analizde kullanılacak olan girdi ve çıktılar belirlenmiştir. Problem, uzun ürünlerin üretim miktarları üzerinde nelerin ne kadar etkiye sahip olduğunun belirlenmesidir. Bu amaçla üretim miktarı, yılbaşı stok

miktarı, yılsonu stok miktarı, satış miktarı, çalışan sayısı, ihracat miktarı, ithalat miktarı, yıllık üretim kapasitesi, üretilen çelik miktarı, lojistik miktarı ve kapasiteden yararlanma oranı arasındaki ilişkiler ortaya konulmaya çalışılmıştır.

3.2.2. Verilerin Hazırlanması

Verilerin hazırlanması sürecinde aşağıda belirtilen aşamalardan geçilmiştir:

- **Verilerin Toplanması**

Tanımlanan problem için gerekli olan veriler BYDK arşivinden 1937 - 1995 yıllarını içeren ciltler detaylı incelenerek elde edilmiştir. Veriler; üretim miktarı, yılbaşı stok miktarı, yılsonu stok miktarı, satış miktarı, çalışan sayısı, ihracat miktarı, ithalat miktarı, yıllık üretim kapasitesi, üretilen çelik miktarı, lojistik miktarı ve kapasiteden yararlanma oranı olmak üzere 11 değişken olarak tanımlanmış ve toplanan tüm veriler bir Excel dosyası içerisinde tablolanmıştır. Oluşturulan veri setindeki bu veriler dikkate alınarak model kurulmuştur.

Veri hazırlığı sonucu oluşturulan değişkenlerin bir kısmı diğer değişkenler ile ilişkileri sebebi ile çıkarılmış, bir kısmı da anlamsız olduklarından sonradan modellere dâhil edilmemişlerdir.

Üretim Miktarı: Üretim miktarı değişkeni bağımlı değişkendir. Belirtilen yıl içerisinde, işyerinde üretilen ile hammadde ve malzemesini vermek üzere başka bir işyerine yaptırılan uzun ürün miktarıdır. Modeller, daha önce alınan bir kararı temel alarak her yeni bir yıl için bir tahmin kararı oluştururlar. Sonrasında seçilen modelin, bağımlı değişkeni, bağımsız değişkenlerin değerleri ile doğrudan olarak tahminleyebilmesi amaçlanmaktadır.

Satış Miktarı: Belirtilen yıl içerisinde, işyerinde üretilen ile hammadde ve malzemesini vermek üzere başka bir işyerine yaptırılan üretimden ve stoklardan yapılan satışlarla işyerinin bağlı bulunduğu kuruluşa ait, diğer işyerine devredilen ürünlerin miktarıdır.

Stok Miktarı: Belirtilen ay sonu itibarı ile işyerinin ürettiği tüm ürünlerin stoklardaki hesap kalanıdır. Stok miktarı aşağıda verilen formül ile hesaplanır:

Stok Miktarı = Bir önceki ayın stok miktarı + içinde bulunan ayın üretim miktarı - satış miktarı - işletmenin kendi içinde kullandığı miktar.

İthalat - İhracat Miktarı: Bir ülkede yerleşik kişi ve kurumların diğer ülkelere mal satmasına ihracat (dışsatım), diğer ülkelerden mal satın almasına ise ithalat (dışalım) denir.

Üretim Kapasitesi: Bir işletmenin belli bir zaman diliminde üretebileceği mal veya hizmet miktarına üretim kapasitesi denir.

Çalışan Sayısı: İşletmede işçi, memur, sözleşmeli statüsünde çalışan kişi sayısıdır.

Lojistik Miktarı: Her türlü hammadde, yardımcı madde ve ürünlerin müessese içinde ve dışında taşınma miktarıdır.

Çelik Üretim Miktarı: İşletmede üretimi yapılan toplam çelik miktarıdır.

- **Verilerin Temizlenmesi**

Veri seti programa tanıtıldıktan sonra, verilerde bir sapma, anormal bir değer olup olmadığının tespiti için öncelikle veri kalitesi (data audit) incelenmiştir. Verilerin tablolara aktarılması sonrasında elde edilen kayıtlarda belli sayıda hatalar ve/veya eksikler saptanmıştır. Bu sorunlardan bazıları ve çözüm için yapılan işlemler aşağıda kısaca özetlenmiştir:

Tekrar Eden Kayıt: Takip eden yıllara ait raporlarda bağımlı ve bağımsız değişkenlerin birden fazla kopyası görülmüştür. Raporlardan toplanan verilerin birleştirilmesi esnasında aynı veriden birden fazla sayıda tespit edildiğinde verilerden bir tanesi veritabanında tutulmuş, diğerleri silinmiştir.

Eksik Veri: Verilerin bir kısmında, çeşitli sebeplerden kaynaklanan eksik verilere rastlanmıştır. Bu tip veriler de analizleri olumsuz etkilememesi için ortalama alma yöntemiyle tahmin edilerek veritabanına eklenmiştir. Eksik verileri tamamlamak için SPSS programı üzerinden Transform/Replace Missing Values/Series Mean ile eksik veriler tamamlanmıştır.

- **Verilerin Bütünleştirilmesi**

Çalışmada çeşitli haddehane ve ocaklarda üretimi ayrı ayrı yapılan uzun ürün gruplarının üretim miktarları, yılbaşı ve yılsonu stok miktarları, satış miktarları, ithalat ve ihracat miktarları bütünleştirilerek veri setine aktarılmıştır.

- **Verilerin Dönüştürülmesi**

Başbakanlık Yüksek Denetleme Kurulu (BYDK)'ndan temin edilen raporlarda mevcut olan ithalat ve ihracat miktarlarından bir kısmı dönemin döviz kurları ve/veya ürün fiyatları göz önüne alınarak hesaplanmıştır.

- **Verilerin İndirgenmesi**

Çalışmada, ilk olarak BYDK arşivinde bulunan veri tabloları ve bu tablolarda bulunan sağlıklı analiz edilebilir verilerin neler olduğu tespit edilmiştir. Tablolarda bulunan her bir bilgi alanı tek tek gözden geçirilerek veri madenciliği çalışmalarında kullanılıp kullanılmayacağı incelenmiştir. Mevcut verilerden yola çıkarak elde edilebilecek yeni veriler üzerinde de yine bu aşamada durulmuştur. Veri ve ilişki anlama aşamasında, veri hazırlama için yapılan işlemler bir “Süper Düğüm” içerisinde toplanmış ve veri tabanına bağlanmıştır. “Type” Düğümü ile değişkenler yeniden tanımlanıp, girdi ve çıktı değişkenleri belirlenmiştir. Karar değişkeni olarak üretim miktarının seçildiği uygulamada her bir değişken ile karar değişkeni arasındaki ilişki, istatistiksel olarak ortaya konulmuştur.

3.2.3. Modelin Kurulması

Tüm değişkenlerin tanımlayıcı istatistikleri incelenmiş ve bilgi vermeyeceği düşünülen değişkenler elenmiştir. Uç değerler, o değişken için tüm verinin dağılımına bakılarak belirli bir noktaya çekilmiştir. Ayrıca boş değerler, o değişken içindeki diğer tüm kategorilerden daha yüksek miktarda ise tablodan çıkarılmıştır.

Tüm bu çalışmaların sonucunda, modellemede kullanılabilecek girdi değişkenleri ile “üretim miktarı” olarak ifade edilen çıktı değişkeni hazırlanmıştır. Analiz için toplanan verilerin düzenlenmesinin ardından, üretim miktarı ile karar değişkenleri arasındaki tüm ilişkileri kapsayan model bu adımda oluşturulmuştur. Oluşturulan modelde çoklu regresyon ve karar ağaçları ile analizler gerçekleştirilmiştir.

IBM SPSS Statistics 19.0 paket programı ile verilere çoklu regresyon yöntemi uygulanmış, model uygulanmadan önce gerekli korelasyon analizi yapılarak birbiri ile ilişkisi yüksek değişkenler elenmiştir. Çoklu regresyon modelde çıkan anlamsız değişkenler de elenerek modelin son haline ulaşılmış, diğer modeller de bu değişkenler üzerinden çalıştırılmıştır. Modelde anlamsız çıkan ve otokorelasyon oluşturan bağımsız

değişkenler; yılbaşı stok miktarı, üretim kapasitesi, lojistik miktarı, kapasiteden yararlanma oranı ve üretilen çelik miktarı olarak belirlenmiş ve modelden çıkarılmıştır.

Bu aşamada verilere, Çoklu Regresyon ile karar ağaçlarından Chaid algoritması uygulanmıştır. Hedef değişkeni olarak üretim miktarı alınmış, 1943-1994 yılları arasındaki uzun ürün verilerini içeren toplam veri kümesi içerisinde karar değişkeninin kategorileri eşit ağırlıklandırılmıştır.

Verilerin modellenmesi aşamasında Chaid ve Çoklu Regresyon sınıflandırma algoritmaları denenerek model geliştirilmiştir.

Chaid algoritması ve Çoklu Regresyon, çalışmanın amacına uygun olarak, bağımsız değişkenlerin bağımlı değişken ile arasındaki ilişkilerin en ince detayına kadar araştırılmasına olanak sağlamasından dolayı, çalışmanın amacına uygun algoritmalar olarak belirlenmiştir.

Uygulamada SPSS Inc. Şirketi tarafından veri madenciliği uygulamaları için tasarlanmış IBM SPSS Statistics 19.0 ve IBM SPSS Modeler (Clementine 11.1) paket programları kullanılmıştır.

Chaid algoritması SPSS Clementine 11.1 paket programı ile analiz edilerek yorumlanırken, çoklu regresyon modeli ise Clementine 11.1 paket programının çoklu bağlantı ve otokorelasyon çalıştırmaması yüzünden IBM SPSS Statistics 19.0 kullanılarak analiz edilmiş ve yorumlanmıştır.

3.2.3.1. Çoklu Regresyon Analizi

Çalışmada demir çelik sektöründe A işletmesinde üretimi yapılan uzun ürünlerin üretim miktarlarının tahminine ilişkin, sınıflandırma temelli çoklu regresyon analizi yapılmıştır. Regresyon modelinin oluşturulması için öncelikle modele anlamlı katkısı olan değişkenlerin belirlenmesine çalışılmıştır. Bunun için gerçekleştirilen çoklu regresyon analizi sonucunda oluşan modelin özeti Tablo 3.1'de gösterilmiştir. Bu tablo, üretim miktarının % 88'e yakın bölümünün bağımsız değişkenler tarafından açıklanabildiğini göstermiştir.

Tablo 3.1. Regresyon Model Özeti

Model	R	R ²	Düzeltilmiş R ²	Tahminin Standart Hatası
1	,937	,878	,865	67294,934

Bu regresyon modelinin anlamlı olup olmadığını anlamak için incelenen Anova testi sonuçları Tablo 3.2’de yer almaktadır.

Bağımsız değişkenlerin, bağımlı değişkeni tahminlediği düşünülen modelde şu hipotezler kurulmuştur:

H_0 =Model anlamsızdır.

H_1 =Model anlamlıdır.

Bu testin sonucunda ortaya çıkan F değerine karşılık gelen p anlamlılık seviyesi 0.05’ten küçük olduğu için söz konusu regresyon modelinin açıklayıcılığı istatistiksel açıdan önemlidir ve H_0 hipotezi reddedilir. Yani kurulan model anlamlıdır ($p=0,00<0,05$).

Tablo 3.2. Regresyon Modeli Anova Sonuçları

Model	Kareler Toplamı	df	Karelerin Ortalaması	F	Sig.
1 Regresyon	1,528E12	5	3,056E11	67,491	,000
Artıklar	2,128E11	47	4,529E9		
Toplam	1,741E12	52			

Ardından, mevcut modelin değişkenlerinin modele olan katkılarını belirlemek amacıyla regresyon analizinden elde edilen ve Tablo 3.3’te görülmekte olan katsayılar tablosu incelenmiştir.

Tablo 3.3. Çoklu Regresyon Modeli/Bağımsız Değişkenler ve Katsayıları

Model		Katsayılar		Standart	t	Sig.
		B	Std. Hata	Katsayılar		
				Beta		
1	Sabit	21770,844	23872,853		,912	,000
	İthalat miktarı	,051	,095	,044	,538	,049
	İhracat miktarı	,227	,073	,201	3,110	,003
	Satış miktarı	,632	,084	,711	7,483	,000
	Yılsonu stok miktarı	,227	,182	,079	1,246	,029
	Kişi sayısı	5,915	3,845	,137	1,538	,031

Katsayılar tablosunda önem değerleri incelendiğinde tüm değişkenlerin modele anlamlı katkısının olduğu görülmüştür ($p_{\text{sabit}}=0,00<0,05$, $p_{\text{ithalat}}=0,049<0,05$, $p_{\text{ihracat}}=0,03<0,05$, $p_{\text{satış miktarı}}=0,00<0,05$, $p_{\text{yılsonu stok miktarı}}=0,029<0,05$, $p_{\text{çalışan kişi sayısı}}=0,031<0,05$).

Modele dahil edilen değişkenlerin tanımlayıcı istatistikleri Tablo 3.4'te gösterilmiştir:

Tablo 3.4. Değişkenlerin Tanımlayıcı İstatistikleri

Değişkenler	Ortalama (Ton)	Standart Sapması	N
Üretim miktarı	366861,396	1,829	53
İthalat miktarı	221752,971	1,566	53
İhracat miktarı	196493,349	1,621	53
Satış miktarı	342249,358	2,058	53
Yılsonu stok miktarı	76944,188	63705,112	53
Çalışan kişi sayısı	9386,245	4238,446	53

Yukarıdaki tabloda bağımlı değişken olan üretim miktarının ortalaması ve standart sapması $366861 \pm 1,83$ ton olarak gerçekleşmiştir.

- İthalat miktarının ortalaması ve standart sapması $221753 \pm 1,566$ ton olarak gerçekleşmiştir.
- İhracat miktarının ortalaması ve standart sapması $196493 \pm 1,621$ ton olarak gerçekleşmiştir.

- Yılsonu stok miktarının ortalaması ve standart sapması 76944 ± 63705 ton olarak gerçekleşmiştir.
- Çalışan kişi sayısının ortalaması ve standart sapması 9386 ± 4238 kişi olarak gerçekleşmiştir.
- Satış Miktarının ortalaması ve standart sapması $342249 \pm 2,058$ ton olarak gerçekleşmiştir.

Değişkenler arasındaki korelasyon değerleri Tablo 3.5’de gösterilmiştir:

Tablo 3.5. Korelasyon Tablosu

		Üretim miktarı	İthalat miktarı	İhracat miktarı	Satış miktarı	Yılsonu stok miktarı	Çalışan kişi sayısı
Korelasyon	Üretim miktarı	1,000	,661	,429	,893	,374	,719
	İthalat miktarı	,661	1,000	,396	,683	,126	,305
	İhracat miktarı	,429	,396	1,000	,216	,046	,392
	Satış miktarı	,893	,683	,216	1,000	,289	,628
	Yılsonu stok miktarı	,374	,126	,046	,289	1,000	,548
	Çalışan kişi sayısı	,719	,305	,392	,628	,548	1,000
Sig. (1-tailed)	Üretim miktarı	.	,000	,001	,000	,003	,000
	İthalat miktarı	,000	.	,002	,000	,043	,013
	İhracat miktarı	,001	,002	.	,040	,042	,002
	Satış miktarı	,000	,000	,040	.	,018	,000
	Yılsonu stok miktarı	,003	,043	,042	,018	.	,000
	Çalışan kişi sayısı	,000	,013	,002	,000	,000	.

Pearson korelasyon tablosuna bakıldığında bağımlı değişken olan üretim miktarı ile yılsonu stok miktarı (0,374) ve ihracat miktarı (0,429) arasında zayıf, ithalat miktarı ile arasında (0,661) orta düzeyde, çalışan kişi sayısı (0,719) ve satış miktarı (0,893) ile arasında ise yüksek bir ilişki bulunmaktadır. Bağımlı ve bağımsız değişkenler arasındaki korelasyon katsayısının önemlilik düzeylerine bakacak olursak; üretim miktarı ile ithalat miktarı arasında korelasyon önemli ($p=0,00<0,05$) ve ilişkinin derecesi 0,661 bulunmuştur. Üretim miktarı ile ihracat miktarı arasında korelasyon önemli ($p=0,01<0,05$) ve ilişkinin derecesi 0,429 bulunmuştur. Üretim miktarı ile satış miktarı arasında korelasyon önemli ($p=0,00<0,05$) ve ilişkinin derecesi 0,893 bulunmuştur. Üretim miktarı ile yılsonu stok miktarı arasında korelasyon önemli

($p=0,03<0,05$) ve ilişkinin derecesi 0,374 bulunmuştur. Son olarak üretim miktarı ile çalışan kişi sayısı arasında korelasyon önemli ($p=0,00<0,05$) ve ilişkinin derecesi 0,719 bulunmuştur. İlişki düzeylerinin tümünün pozitif olduğu göz önüne alındığında bağımsız değişkenlerin tümündeki artış, bağımlı değişken değerini de arttırıcı bir unsur olarak değerlendirilebilir.

Regresyon Varsayımlarının İncelenmesi

- **Otokorelasyon**

Modelde bağımsız değişkenler arasındaki otokorelasyon göstergesi Durbin Watson değeri literatürde genellikle 1,5-2,5 arasında olmasını gerektirir. Tablo 3.6'da Durbin Watson değeri gösterilmiştir:

Tablo 3.6. Model Özeti Tablosundan Otokorelasyonun Tespiti

Model	R	R ²	Düzeltilmiş R ²	Tahminin Standart Hatası	Değişim İstatistikleri					Durbin-Watson
					R ² Değişimi	F Değişimi	df1	df2	F Önem Düzeyi	
1	,937	,878	,865	67294,934	,878	67,491	5	47	,000	1,533

Modelde Durbin Watson değeri 1,533 bulunduğundan hata teriminin birbirini izleyen değerleri arasında ilişki bulunmamakta yani otokorelasyon oluşturmamaktadır.

- **Çoklu Doğrusal Bağlantı Sorununun İncelenmesi**

Modelin koşul endeksi ve öz değer vektör değerleri Tablo 3.7'de gösterilmiştir:

Tablo 3.7. Modelin CI koşul İndeksi ve Özdeğer Vektör Değerleri

Model	Boyut	Özdeğer Vektörleri	Koşul İndeksi (CI)
	1	5,012	1,000
	2	1,412	3,489
	3	2,277	4,255
	4	,964	5,536
	5	1,305	6,918
	6	,331	9,638

CI koşul endeksinin >30 olduğu durumlarda ve özdeğer vektörlerinin değerlerinden birden fazlası sıfıra çok yakın değerler alıyorsa çoklu bağlantı

probleminden bahsedilebilir. Koşul indekslerinin tümü 30'dan küçük olup, özdeğer vektörlerinin değerlerinden sadece bir tanesi sıfıra yakın değer almaktadır.

Bu gibi testlerin yanı sıra çoklu doğrusallık sorununun tespitinde kullanılacak diğer yöntemler ise VIF değeri ile Tolerans değerinin belirlenmesidir.

Tablo 3.8'de Tolerans ve VIF değerleri gösterilmiştir:

Tablo 3.8. Tolerans ve VIF Değerleri

Model	Standart olmayan katsayılar		Standart katsayılar	t	Önem Düzeyi	95,0% B güvenlilik aralığı		Korelasyonlar			Eş doğrusallık İstatistikleri		
	B	Std. Hata	Beta			Alt sınır	Üst sınır	Sıfırcı Derece	Kısmi	Parça	Tolerans	VIF	
													Beta
1 Sabit	21770,84	23872,853		,912	,000	-26255	69796,830						
İthalat	,051	,095	,044	,538	,049	-,140	,242	,661	,078	,027	,394	2,541	
İhracat	,227	,073	,201	3,110	,003	,080	,374	,429	,413	,159	,622	1,608	
Satış	,632	,084	,711	7,483	,000	,462	,802	,893	,737	,382	,288	3,467	
Yılsonu- stok	,227	,182	,079	1,246	,029	-,139	,592	,374	,179	,064	,649	1,540	
Kişi- sayısı	5,915	3,845	,137	1,538	,031	-1,820	13,651	,719	,219	,078	,328	3,050	

Bu noktada çoklu bağlantı problemi oluşturabileceği için değişkenler arasında güçlü ilişki olması istenmez (Alpar, 2010:302). VIF değerleri 5'in altında ve tolerans değerleri 0,2'nin üzerinde olan verilerde çoklu doğrusal bağlantı yoktur (Orhunbilge, 2010:54). VIF faktörleri tüm bağımsız değişkenlerde 5'in altındadır. Tolerans istatistikleri ise tüm değişkenlerde 0,2'nin üzerindedir.

Bağımsız değişkenler arasındaki korelasyon değerlerinden tümünün 0,90'dan küçük olduğu da göz önüne alındığında modelde çoklu bağlantı sorunu olmadığı söylenebilir (Tablo 3.5).

Parametre değerlerine ait t istatistiklerinden modele dâhil edilen her bir değişken ayrı ayrı (%5 anlamlılık düzeyinde) anlamlıdır. Tablo 3.8'den de görüldüğü üzere sabit terim 21771 ± 23872 bulunmuştur.

Satış miktarına ait parametre değeri 0,632 bulunmuştur. Bunun anlamı yılbaşı stok miktarındaki bir birimlik artışın toplam üretim miktarını 0,632 birim arttıracaktır. Değişkenlerin b katsayılarının pozitif olması bu değişkenlerin bağımlı değişkene pozitif yönde etkisi olduğu sonucunu vermektedir.

Beta katsayılarına bakıldığında bağımlı değişken olan üretim miktarını etkileyen en önemli bağımsız değişkenin satış miktarı olduğu görülmektedir. Diğer değişkenler ise önem sırasıyla ihracat miktarı, çalışan kişi sayısı, yılsonu stok miktarı ve ithalat miktarı olarak sıralanmaktadır.

Sıfır dereceli korelasyonlar, basit korelasyon katsayılarıdır. Bu katsayılar bağımlı değişkenler ile bağımsız değişkenler arasındaki ilişkiyi gösterir. Örneğin üretim miktarı ile ithalat miktarı arasında 0,661'lik orta derecede bir korelasyon vardır ve pozitif yöndedir. En yüksek sıfır dereceli korelasyon katsayısı, üretim miktarı ile satış miktarı arasında olup pozitif yönde 0,883 derecesinde yüksek bir korelasyon meydana gelmiştir.

Artıkların istatistik değerleri Tablo 3.9'da gösterilmiştir:

Tablo 3.9. Artıkların İstatistiği

	Minimum	Maksimum	Ortalama	Std. sapma
Artık	-1,697	1,324	,000	63977,848

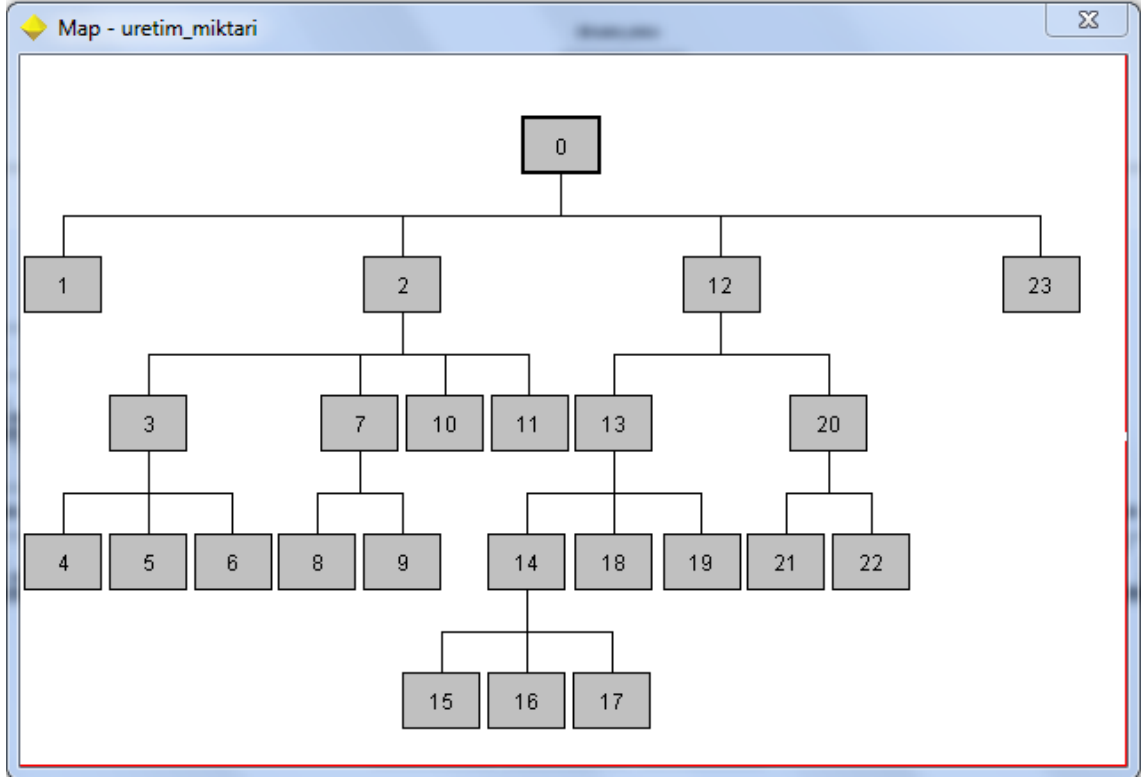
Artıklar bir regresyon modelinin uygulanmasından sonra açıklanmayan değişkenliği gösterir. Bağımlı değişken Y'nin gözlenen değeri ile tahmin edilen değeri (Y-Y') arasındaki sapma puanları olan, tahminde yapılan hatadır. Modelde artıkların minimum değeri -1,697, maksimum değeri ise 1,324 bulunmuştur. Artıkların ortalamasının 0 bulunması

Modeldeki çoklu regresyon denklemi aşağıda gösterilmiştir:

$$\text{Üretim Miktarı} = 21771 + (0,632 \times \text{Satış miktarı}) + (5,915 \times \text{Çalışan kişi sayısı}) + (0,227 \times \text{İhracat miktarı}) + (0,227 \times \text{Yılsonu stok miktarı}) + (0,051 \times \text{İthalat miktarı})$$

3.2.3.2. Chaid Algoritması Analizi

Türkiye'nin entegre demir-çelik tesislerinden biri olan A işletmesinde üretimi yapılan uzun ürün grubunun üretim faaliyetleriyle ilgili değişkenlerine karar ağacı yöntemlerinden Chaid algoritması ile analizler yapılmıştır. Belirtilen değişkenler ile Chaid algoritması sonucunda elde edilen karar ağacının üretim miktarına ilişkin bulgular Şekil 3.1'de gösterilmiştir:



Şekil 3.1. Üretim Miktarı Tahmini İçin Oluşturulan Karar Ağacı

Chaid Algoritması ile yapılan analiz sonucu bağımlı değişken olan üretim miktarını etkileyen en önemli bağımsız değişken satış miktarı olmuştur. Diğer değişkenler ise önem sırasına göre; çalışan kişi sayısı, ithalat miktarı, ihracat miktarı ve yılsonu stok miktarı olarak sıralanmaktadır.

Kurulan modelde, bağımlı değişken olan üretim miktarının tahmini sonuçları aşağıda gösterilmiştir:

İthalat miktarı (47 680 ton - 60 405 ton] aralığında, çalışan kişi sayısı $\leq 4\ 487$ ve satış miktarı (80 573 ton – 333 384 ton] aralığında iken tahmini üretim miktarı 226 623 tondur.

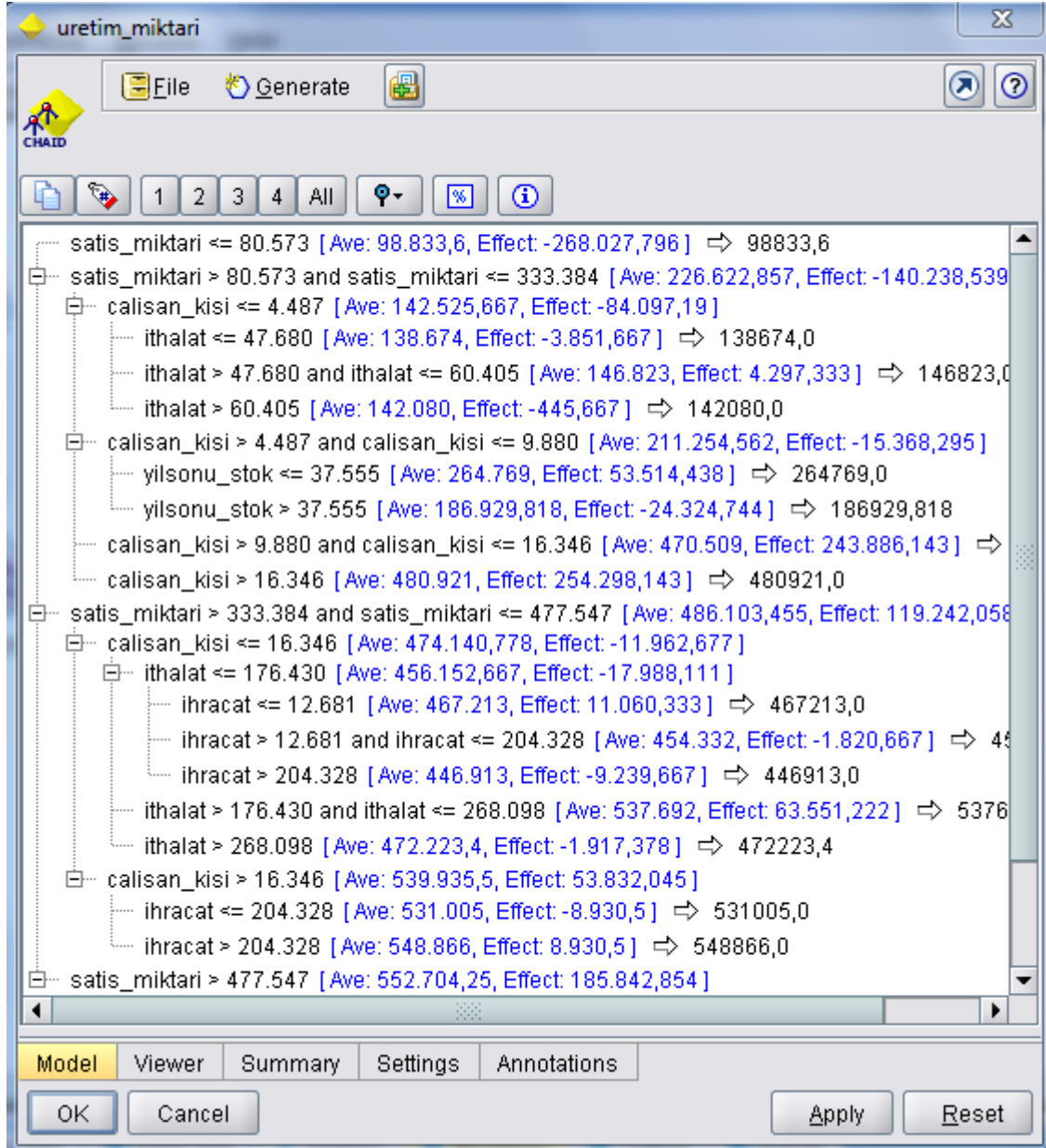
Yılsonu stok miktarı $\leq 37\ 555$ ton, çalışan kişi sayısı (4 487 – 9 880] ve satış miktarı (80 573 ton – 333 384 ton] iken tahmini üretim miktarı 226 623 tondur.

İhracat miktarı $\leq 12\ 681$ ton, ithalat miktarı $\leq 176\ 430$ ton, çalışan kişi sayısı $\leq 16\ 346$ ve satış miktarı (333 384 ton – 477 547 ton] arasında iken tahmini üretim miktarı 486 103 tondur.

İhracat miktarı $> 204\ 328$ ton, çalışan kişi sayısı $> 16\ 346$ ve satış miktarı (333 384 ton – 477 547 ton] iken tahmini üretim miktarı 486 103 tondur.

Yukarıdaki sıraya göre satış miktarı, üretim miktarını etkileyen en önemli bağımsız değişken olarak görülmüştür.

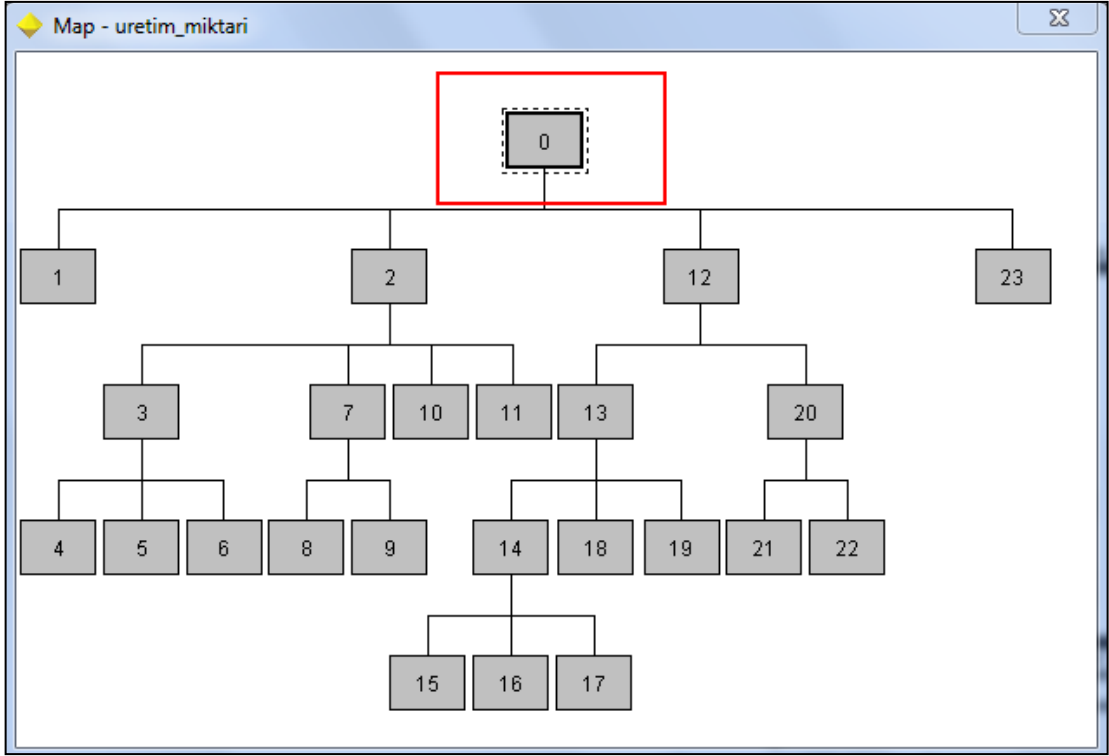
Chaid analizine ilişkin kurallar bütünü Şekil 3.2’de gösterilmiştir:



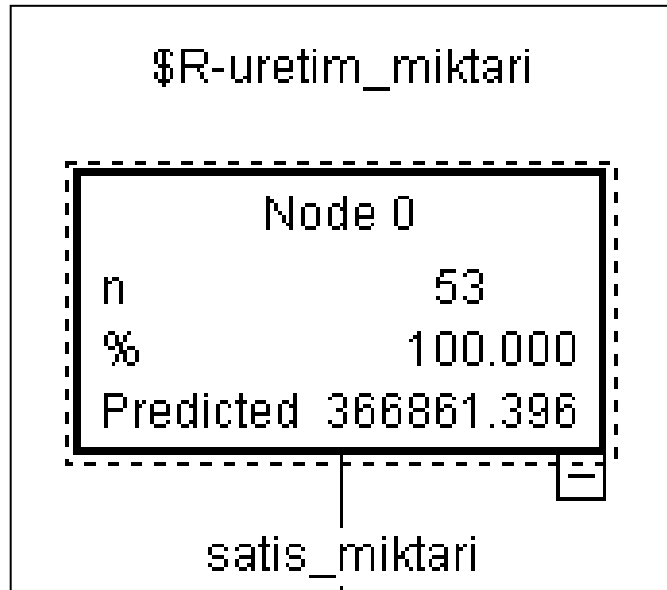
Şekil 3.2. Chaid Analizine İlişkin Kurallar Bütünü

Modelin doğruluğu, değişkenlerin, yaptıkları dallanmalarla istenilen kitleyi ne oranda yakalamış olduğuna göre ölçülebildiği gibi, ağaçtaki kuralların gerçek hayata ne derece uygun olduğu ile de ölçülebilir.

CHAID algoritması sonucunda elde edilen karar ağacının başlangıç düğümü (Node) Şekil 3.3’te gösterilmektedir.

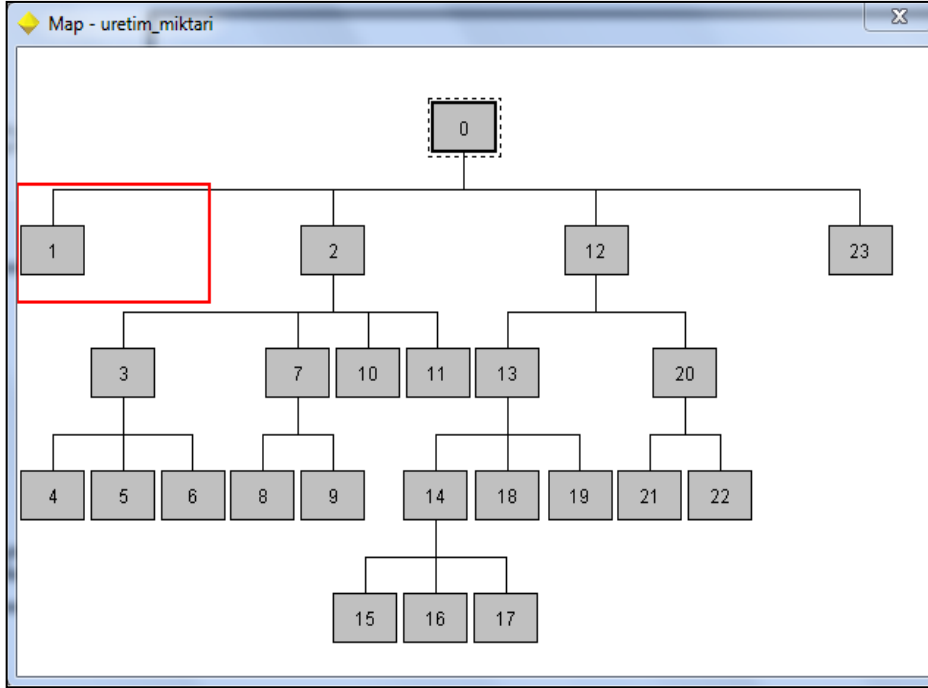


Şekil 3.3. Karar Ağacının Başlangıç Düzümü



Şekil 3.4. Düzüm 0: Tahmini Üretim Miktarı

Üretim miktarına etki eden faktörler; ithalat miktarı, ihracat miktarı, yılsonu stok miktarı, çalışan sayısı ve satış miktarı incelendiğinde; üretim miktarına etki eden en önemli faktörün, satış miktarı olduğu gözlenmiştir.

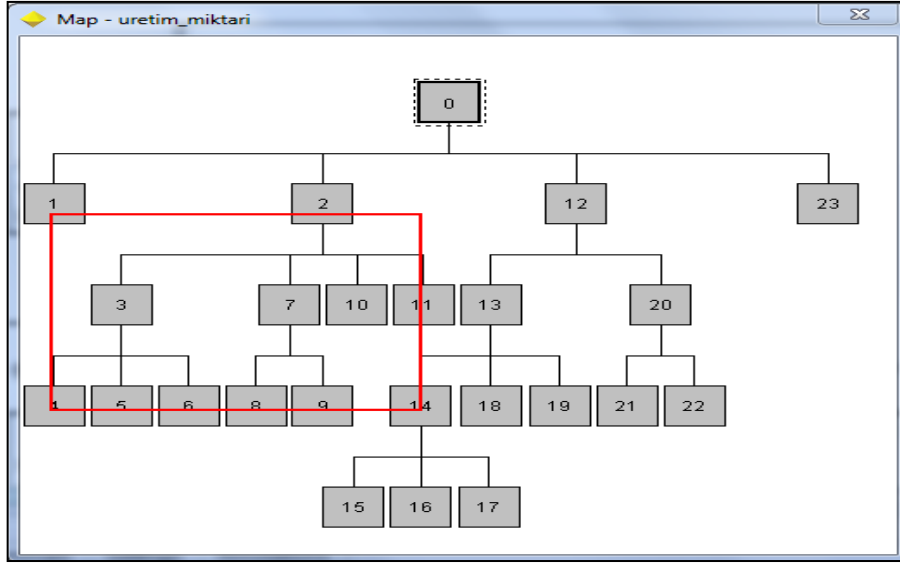


Şekil 3.5. Karar Ağacının Birinci Düzümü

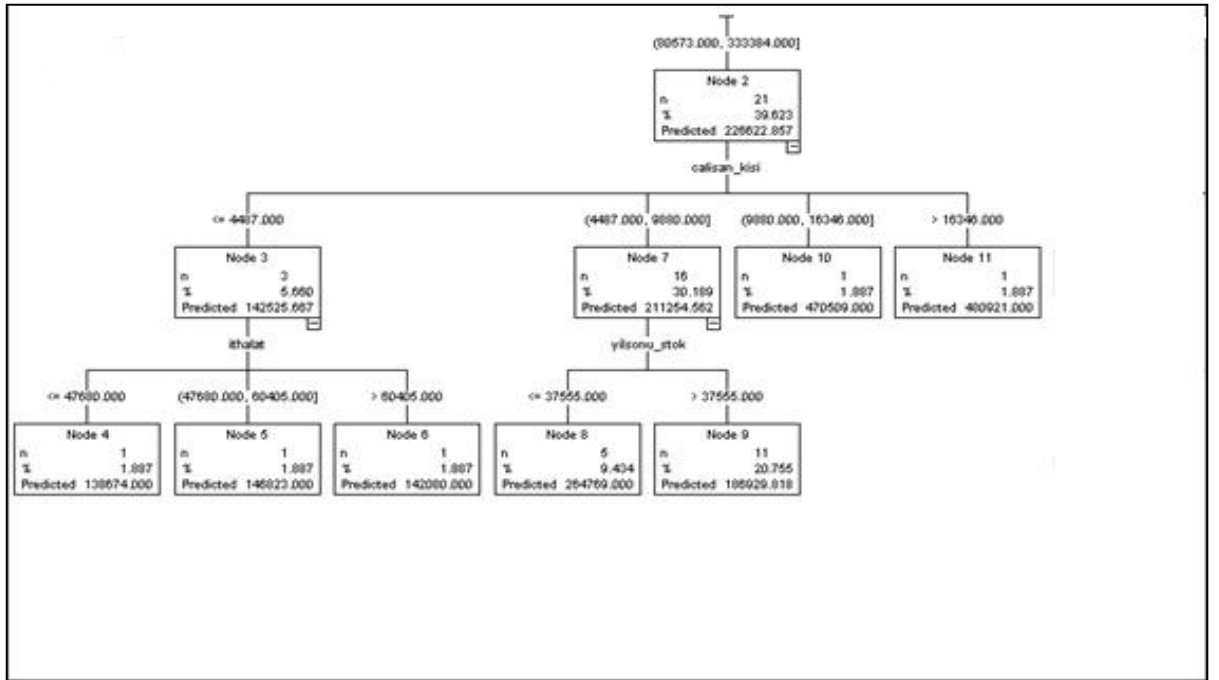
≤ 80573.000	
Node 1	
n	5
%	9.434
Predicted	98833.600

Şekil 3.6. Düzüm 1: Tahmini Üretim Miktarı

Satış miktarının $\leq 80\ 573$ ton olduğu durumda üretim miktarının 98 833 ton olarak gerçekleşeceği görülmüştür.



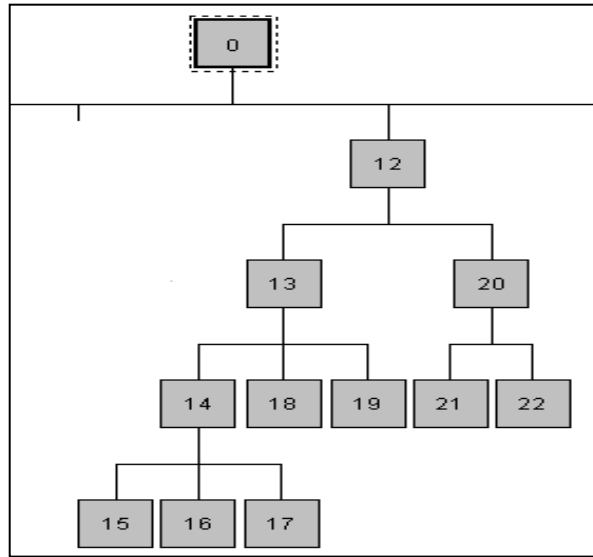
Şekil 3.7. Karar Ağacının 2-11 Düzümleri



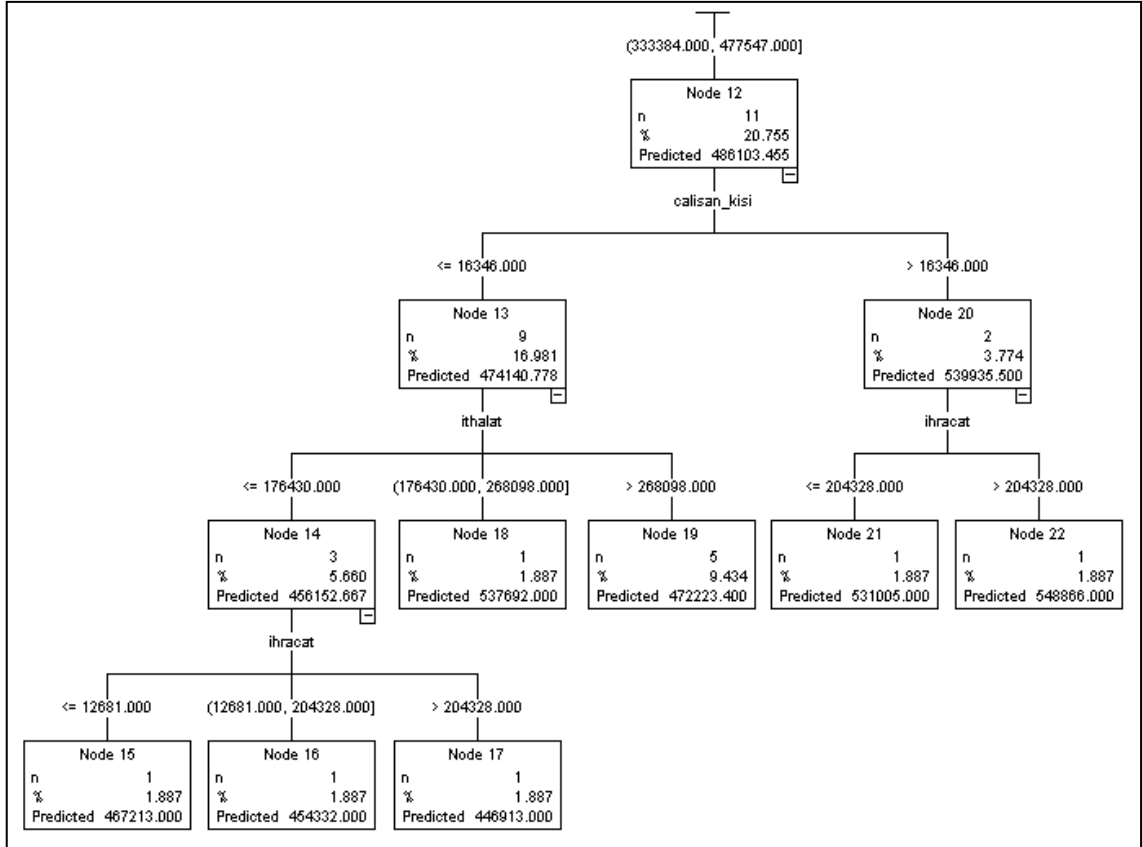
Şekil 3.8. Düzüm 2-11 Tahmini Üretim Miktarları

- Satış miktarının $> 80\,573$ ton ve $\leq 333\,384$ ton, çalışan kişi sayısının $\leq 4\,487$, ithalat miktarının $\leq 47\,680$ ton olduğu durumda üretim miktarının $138\,674$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 80\,573$ ton ve $\leq 333\,384$ ton, çalışan kişi sayısının $\leq 4\,487$, ithalat miktarının $> 47\,680$ ton ve $\leq 60\,405$ ton olduğu durumda üretim miktarının $146\,823$ ton olarak gerçekleşeceği görülmüştür.

- Satış miktarının $> 80\,573$ ton ve $\leq 333\,384$ ton, çalışan kişi sayısının $\leq 4\,487$, ithalat miktarının $> 60\,405$ ton olduğu durumda üretim miktarının $142\,080$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 80\,573$ ton ve $\leq 333\,384$ ton, çalışan kişi sayısının $> 4\,487$ ve $\leq 9\,880$, yılsonu stok miktarının $\leq 37\,555$ ton olduğu durumda üretim miktarının $264\,769$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 80\,573$ ton ve $\leq 333\,384$ ton, çalışan kişi sayısının $> 4\,487$ ve $\leq 9\,880$, yılsonu stok miktarının $> 37\,555$ ton olduğu durumda üretim miktarının $186\,930$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 80\,573$ ton ve $\leq 333\,384$ ton, çalışan kişi sayısının $> 9\,880$ ve $\leq 16\,346$ olduğu durumda üretim miktarının $470\,509$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 80\,573$ ton ve $\leq 333\,384$ ton, çalışan kişi sayısının $> 16\,346$ olduğu durumda üretim miktarının $480\,921$ ton olarak gerçekleşeceği görülmüştür.



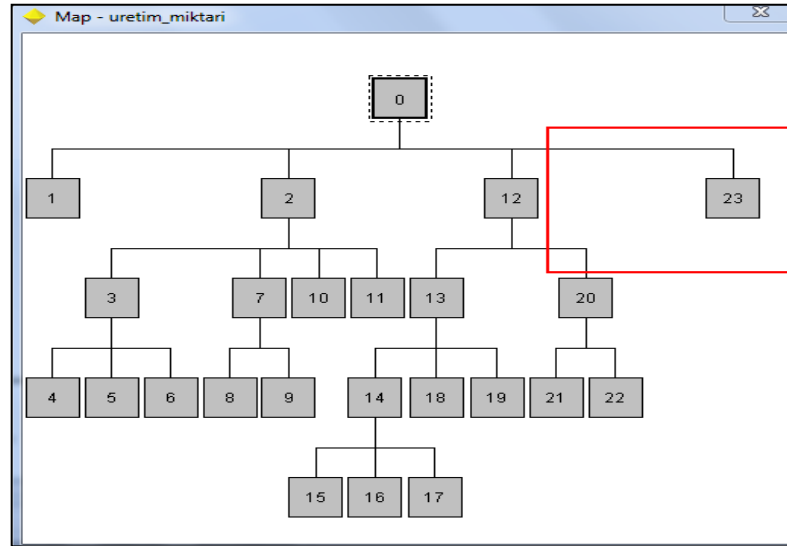
Şekil 3.9. Karar Ağacının 12-22 Dğümleri



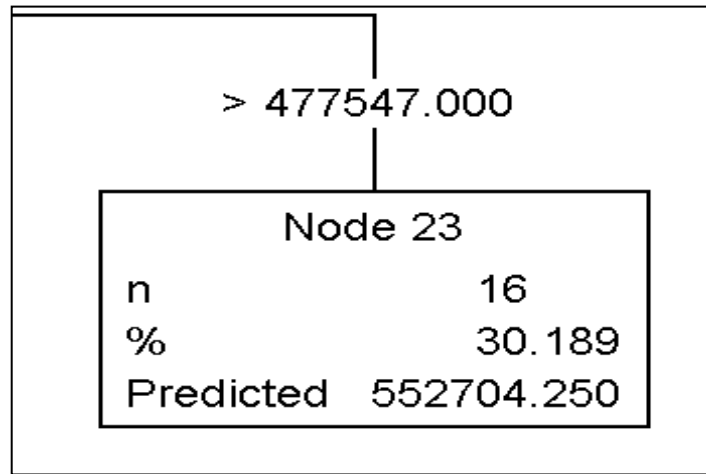
Şekil 3.10. Düzüm 12-22 Tahmini Üretim Miktarları

- Satış miktarının $> 333\ 384$ ton ve $\leq 477\ 547$ ton, çalışan kişi sayısının $\leq 16\ 346$, ithalat miktarının $\leq 176\ 430$ ton, ihracat miktarının ise $\leq 12\ 681$ ton olduğu durumda üretim miktarının $467\ 213$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 333\ 384$ ton ve $\leq 477\ 547$ ton, çalışan kişi sayısının $\leq 16\ 346$, ithalat miktarının $\leq 176\ 430$ ton, ihracat miktarının ise $> 12\ 681$ ton ve $\leq 204\ 328$ ton olduğu durumda üretim miktarının $454\ 332$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 333\ 384$ ton ve $\leq 477\ 547$ ton, çalışan kişi sayısının $\leq 16\ 346$, ithalat miktarının $\leq 176\ 430$ ton, ihracat miktarının ise $> 204\ 328$ ton olduğu durumda üretim miktarının $446\ 913$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 333\ 384$ ton ve $\leq 477\ 547$ ton, çalışan kişi sayısının $\leq 16\ 346$, ithalat miktarının $> 176\ 430$ ton ve $\leq 268\ 098$ ton olduğu durumda üretim miktarının $532\ 692$ ton olarak gerçekleşeceği görülmüştür.

- Satış miktarının $> 333\,384$ ton ve $\leq 477\,547$ ton, çalışan kişi sayısının $\leq 16\,346$, ithalat miktarının $> 268\,098$ ton olduğu durumda üretim miktarının $472\,223$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 333\,384$ ton ve $\leq 477\,547$ ton, çalışan kişi sayısının $> 16\,346$, ihracat miktarının $\leq 204\,328$ ton ve olduğu durumda üretim miktarının $531\,005$ ton olarak gerçekleşeceği görülmüştür.
- Satış miktarının $> 333\,384$ ton ve $\leq 477\,547$ ton, çalışan kişi sayısının $> 16\,346$, ihracat miktarının $> 204\,328$ ton ve olduğu durumda üretim miktarının $548\,866$ ton olarak gerçekleşeceği görülmüştür.



Şekil 3.11. Karar Ağacının 23. Düzümü



Şekil 3.12. Düzüm 23: Tahmini Üretim Miktarı

- Satış miktarının $> 477\,547$ ton olduğu durumda üretim miktarının $552\,704$ ton olarak gerçekleşeceği görülmüştür.

Karar ağacı incelendiğinde, üretim miktarı kararını etkileyen öncelikli değişkenin satış miktarı olduğu ve ağacın ilk dallanmasını bu değişken yardımı ile yaptığı görülebilir. Örneğin, üretim miktarını etkileyen bağımsız değişkenlerden olan yılsonu stok miktarı ile ilgilendiğimizi düşünelim. Bu durumda takip edilmesi gereken dal, satış miktarının $> 80\,573$ ton ve $\leq 333\,384$ ton, çalışan kişi sayısının $> 4\,487$ ve $\leq 9\,880$, yılsonu stok miktarının $> 37\,555$ ton durumu belirten daldır.

- **Sonuçların Doğruluğunun Belirlenmesi**

Üretim yapılan seneler arasındaki yıllık gerçek üretim miktarları (üretim miktarı) ile tahmini üretim miktarları ($\$R$ -üretim_miktarı) ve tahminin güvenilirlik değerleri ($\$RI$ -üretim_miktarı) Tablo 3.10'da gösterilmiştir.

Tablo 3.10. Tahmini Üretim Miktarı (Ton) ve Tahminin Güvenilirlik Değerleri

Table (12 fields, 53 records)											
File Edit Generate											
	uretim_miktari	...	i...	i...	\$R-uretim_miktari	\$RI-uretim_miktari
1	74968.00	84098.667	1
2	86661.00	84098.667	1
3	90667.00	84098.667	1
4	110481.00	129437.000	5
5	131391.00	230572.000	7
6	146823.00	129437.000	5
7	138674.00	129437.000	5
8	142080.00	142080.000	11
9	171485.00	176134.333	6
10	222670.00	230572.000	7
11	121770.00	129437.000	5
12	168892.00	176134.333	6
13	188026.00	176134.333	6
14	272828.00	230572.000	7
15	262777.00	230572.000	7
16	252993.00	230572.000	7
17	312577.00	230572.000	7
18	219096.00	230572.000	7
19	185203.00	230572.000	7
20	196225.00	230572.000	7
21	249960.00	230572.000	7
22	145293.00	145293.000	10
23	149629.00	149629.000	9
24	454332.00	454332.000	12
25	447632.00	478998.727	24
26	622774.00	622774.000	23
27	537692.00	537692.000	18
28	446913.00	478998.727	24
29	467213.00	478998.727	24
30	530349.00	478998.727	24
31	534296.00	478998.727	24
32	531005.00	531005.000	27
33	548866.00	548866.000	28
34	568412.00	568412.000	26
35	480921.00	480921.000	30
36	470509.00	478998.727	24
37	513683.00	513683.000	29
38	491591.00	478998.727	24
39	479067.00	478998.727	24
40	465286.00	478998.727	24
41	486259.00	478998.727	24
42	419871.00	478998.727	24
43	482873.00	482873.000	16
44	515875.00	515875.000	20
45	519038.00	519038.000	22
46	515782.00	515782.000	21
47	260649.00	260649.000	14
48	554835.00	596932.833	31
49	537062.00	596932.833	31
50	537062.00	596932.833	31
51	543496.00	596932.833	31
52	617546.00	596932.833	31
53	601606.00	606032.933	31

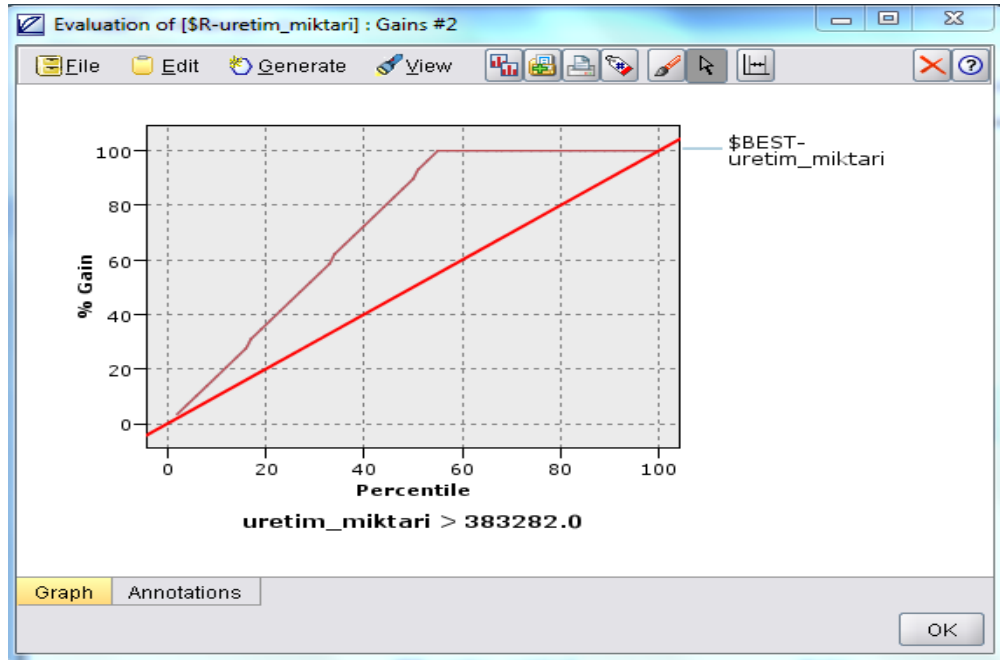
Chaid modelinin analiz sonucu Tablo 3.11’de gösterilmiştir:

Tablo 3.11. Chaid Analizi (Üretim Miktarı Analizi)

	Minimum Hata	Maksimum Hata	Hata Ortalaması	Standart Sapması	Doğrusal Korelasyon
Chaid Modeli (Üretim Miktarı Analizi)	-88 281	84 622	0,00	20 590	0,985

Chaid analizi sonucunda bağımlı değişken olan üretim miktarı ile bağımsız değişkenler arasındaki ilişki düzeyinin belirlendiği doğrusal korelasyon değeri 0,985 hesaplanmıştır. $-1 \leq r \leq 1$ için korelasyon katsayısı anlamlıdır, yani korelasyon katsayısı modelde kullanılabilir. Ayrıca ortalama hata 0 bulunmuştur. Hataların ortalamasının 0 bulunması modelin uygunluğunun göstergesidir (Orhunbilge, 2010:46).

Şekil 3.13’de Chaid analizine ilişkin bağımsız değişkenlerin bağımlı değişkeni üzerindeki etkisini gösteren “Performans Değerlendirme Grafiği (Evaluation düğümü) verilmiştir.



Şekil 3.13. Sınıflandırma Doğruluğu Grafiği

Sınıflandırma doğruluğu grafiklerinde bir tahminin başarısı, merkez çizgiden uzak ve ideal (best) çizgiye yakın olması ile ölçülmektedir. Yüzde başarı oranı ile fikir verilmesine yardımcı olan değerlendirme grafiği (% Gain), modelleri karşılaştırarak uygun modele ilişkin sonuçları ortaya koyar.

Tahmini üretim miktarının grafiği merkez çizgiden uzak ve ideal çizgiye yakın görülmüştür. Bu ölçüye göre Chaid algoritmasının tahminleri başarılıdır.

3.3. SINIFLAMA MODELLERİNİN KARŞILAŞTIRILMASI

Üretim miktarını etkileyen değişkenlerin tespitinde çoklu regresyon ve Chaid analizi bağımlı değişkeni etkileyen en önemli bağımsız değişkenin satış miktarı olduğunu göstermiştir. Hem regresyon analizinde hem de Chaid analizinde bağımsız değişkenler önem sırasına göre; satış miktarı, ihracat miktarı, çalışan kişi sayısı, yılsonu stok miktarı ve ithalat miktarı olarak sıralanmaktadır.

Chaid algoritması sonucunda 53 örneklem üzerinden minimum hata değeri -88 281, maksimum hata değeri 84 622, ortalama hatası 0, standart sapması 31 283 ve doğrusal korelasyon katsayısı 0,985 bulunmuştur.

Çoklu regresyonda 53 örneklem üzerinden tahmin değerlerinin minimum hata değeri -169 772, maksimum hata değeri 132 412, ortalama hatası 0, standart sapması 63 918 ve doğrusal korelasyon değeri 0,937 bulunmuştur.

Tablo 3.12. Üretim Miktarı Analiz Sonuçlarının Karşılaştırması

	Minimum Hata	Maksimum Hata	Hata Ortalaması	Standart Sapması	Doğrusal Korelasyon
Çoklu Regresyon Analizi	-169 772	132 412	0,0	63 978	0,937
Chaid Analizi	-88 281	84 622	0,0	20 590	0,985

Bu sonuçlardan hareketle, uzun ürünlerin üretim miktarlarının tahminine yönelik uygulanan analiz yöntemlerinden Chaid analizi ile oluşturulan modelin çoklu regresyon analizi ile oluşturulan modelden daha güvenilir sonuçlar verdiği görülmüştür.

4. SONUÇ ve ÖNERİLER

Çalışmada, veri madenciliği yöntemleri özetlenerek sınıflandırma yöntemleri üzerinde durulmuş ve eldeki veriler kullanılarak sınıflandırıcı modeller geliştirilmiştir. Çalışmada demir çelik sektöründe uzun ürünlerin üretim miktarlarını etkileyen değişkenler, veri madenciliği ile araştırılmış ve üretim miktarları tahmin edilmiştir.

Çalışmada Başbakanlık Yüksek Denetleme Kurulu (BYDK)'dan alınan veriler kullanılmıştır. Bu bilgiler çeşitli değişkenlerde değerlendirilmiş, aralarındaki ilişkilere bakılarak modellenmesi uygun görülen ve gerekli olan değişkenler saptanmıştır. Sonuçta bu değişkenler ile bağımlı değişken olan üretim miktarının tahminine ilişkin modeller geliştirilmiştir. Bunun için çoklu regresyon ve Chaid algoritması denenmiş ve sonuçları incelenmiştir. Çalışmada tespit edilen modeller, üretim kararı verilirken demir çelik sektörünün özelliklerini yansıttığı görülmektedir. Modellerin geliştirilmesinin esas amacı, sonrasında bağımsız değişkenlerin aynı özelliklerine bakılarak ne kadar üretim meydana geleceğini anlayabilmeye yardımcı olabilecek bir yapı hazırlamaktır.

Her iki modele ait analiz sonuçlarına bakıldığında; üretim miktarını etkileyen değişkenlerin, önem sırasının aynı değişkenlerden oluştuğu, bunların sırasıyla; satış miktarı, ihracat miktarı, çalışan kişi sayısı, yılsonu stok miktarı ve ithalat miktarı olduğu belirlenmiştir.

Üretim sektöründe veri madenciliği uygulamalarının önünde bazı engeller bulunmaktadır. Özellikle demir - çelik sektöründe veriler çoğunlukla düzenli bir şekilde tutulmadığından, çalışmalarda kullanılacak verilerin bulunmasında zorluk yaşanmaktadır. Entegre demir - çelik tesisleri ve özel haddehaneler çeşitli sebeplerden ötürü veri tabanlarından bilgi paylaşımına çok istekli değildirler. Bu gibi durumlarda TÜİK, BYDK vb. gibi kurumlardan bilgi talebinde bulunularak çalışmalar yürütülebilmektedir. Verilerin elde edilmesi, çalışmanın başlaması için her durumda yeterli olmamaktadır. Çoğu kez, çalışmanın yürütülebilmesi için konuyla ilgili bir uzmanın bulunması ve destek alınması gerekebilir. Bu zorlukların en aza indirilmesi, üretim sektöründe veri madenciliği uygulamalarının artmasını sağlayacaktır.

Çalışmada kullanılan verilerin özelleştirme öncesi döneme ait verilerden seçilmiş olmasının sebebi, özelleştirme sonrası sağlıklı verilere ulaşılamaması dolayısıyla analiz edilememesidir. Çalışmada 1943 - 1994 seneleri arasındaki

raporlardan derlenmiş veriler kullanılmış olup, özelleştirme sonrası dönemine ait verilerin düzenlenerek yeni bir sınıflandırma temelli çalışma yapılması önerilmektedir.

Son yıllarda dünya ortalamasının üzerinde bir artış gösteren Türk demir-çelik üretim kapasitesinin, yapılan yeni yatırımlar sayesinde 2015 yılında 55 milyon tonluk, 2023 senesinde ise 85 milyon tonluk bir seviyeye ulaşması hedeflenmektedir. Çelik sektörü üzerindeki yüklerin hafifletilmesi ile kapasite kullanım oranı % 80'lerin üzerine çıkarılabilecektir (TOBB, 2010:18).

İlerleyen dönemlerde, demir çelik sektörünün büyüme performansını sürdürülebilmesi, iç tüketime inşaat sektörünün yaptığı katkıyı, yassı çelik tüketen sanayi kuruluşlarının da yapmasına, Türkiye'de yerli çelik tüketimini teşvik edecek tedbirlerin alınmasına, ülkemize kalitesiz ürünlerin girişinin engellenmesine ve sektörün ihrac piyasalarındaki rekabet gücünün artırılmasına bağlı olduğu değerlendirilmektedir.

Türkiye'nin, yassı-uzun ürün arz-talep dengesizliğinden kaynaklanan sorunları aşması uzun vadede ise vasıflı, paslanmaz ve yapısal çelik gibi katma değeri yüksek ürünlerin, üretim ve tüketim paylarını arttırması öngörülmektedir. Ayrıca, Türkiye'nin deprem bölgesinde olması nedeniyle yapısal çeliğe yönelik tüketim alışkanlıklarının yerleşmesi gerekmektedir.

2009 yılındaki 25.3 milyon tonluk ham çelik üretiminin 20,5 milyon tonu (% 81,2) uzun ürünlere, 4,7 milyon tonu (% 18,8) yassı ürünlere yönelik olarak gerçekleşmiştir. Çelik sektörü, uzun ürünlere Türkiye'nin tüketiminin iki katından daha fazla üretim yaparken, katma değeri yüksek olan yassı ve vasıflı çelik ürünlerinde, tüketiminin sırasıyla % 44 ve % 32'sini gerçekleştirebilmektedir. Bu nedenle, sektör uzun ürünlere net ihracatçı iken, yassı ve vasıflı ürünlerde net ithalatçı konumundadır. Bu sebepten yassı ve vasıflı ürünlerin üretimini arttırarak dışa bağımlılıktan kurtulmak amacıyla altyapı çalışmalarına ağırlık verilmelidir.

Ülkemizde son yıllarda demiryolu altyapıları hızla gelişmektedir. Demiryolu alanında konvansiyonel hatlar ve hızlı tren hatları ile şehirlerarası, Marmaray vb. projeler ile de kentlerimizde metro ve tramvay yatırımlarına önümüzdeki 12 yıl içinde takriben toplam 22- 25 milyar \$ yatırım yapılacağını tahmin edilmektedir. Dolayısıyla uzun ürün grubuna dahil olan ray çeşitliliğinin arttırılması, ihtiyaç analizlerinin yapılarak yeterli ve gerekli üretimin yurtiçi imkanlarla yapılmasına çalışılmalıdır.

Türkiye, 2009 – 2014 yılları arası dönemde inşaat sektöründe % 8,5 oranında büyüme beklentisi ile çok olumlu gözükten ülkeler arasında yer almıştır. Konut talebindeki artış ve artan kentsel dönüşüm projeleri, uzun, yassı ve vasıflı çelik üretimine olan ihtiyacı arttıracaktır.

Türkiye’de demir-çelik sektörünün yeniden yapılandırılması gerekmektedir. Bu amaçla alınacak tedbirler aşağıdaki şekilde sıralanabilir:

- Küresel rekabet ortamındaki gelişmeler doğru olarak algılanmalı, bu bağlamda stratejik işbirlikleri ve ortak projeler geliştirilmelidir.
- Sektördeki üretim faaliyetlerinin çevreye yönelik zararlarının en düşük düzeye indirilmesi amacıyla yasal mevzuattaki eksiklikler giderilmeli ve bunların uygulaması ilgili birimlerce denetlenmelidir.
- Üretim yapısındaki ve kompozisyonundaki dengesizliği gidermeye ve yassı ürünlerde dışa bağımlılığı ortadan kaldırmaya yönelik girişimlerin bir sonucu olarak 2002 yılında Erdemir’e devredilen, İsdemir’deki modernizasyon ve ürün dönüşüm yatırımlarındaki süreç hızlandırılmalıdır.
- Uluslararası demir-çelik ticaretinde, Türk demir-çelik sektörüne yönelik uygulanan tarife dışı engeller ve anti-damping uygulamalarına karşı devletin ilgili birimleri yeni politikalar belirlemeli ve Dünya Ticaret Örgütü (WTO) nezdinde girişimlerde bulunulmalıdır.
- Dünya demir-çelik piyasasındaki ve sektörel yapıdaki gelişmeler takip edilmeli, bu çerçevede yeni stratejiler oluşturulmalıdır.
- Vasıflı çelik ürünleri grubunu oluşturan ve savunma, otomotiv, otomotiv yan sanayi, makine ve yay imalat sanayisinin temel girdisini oluşturan özel yapı, paslanmaz ve ısıya dayanıklı çelik ürünlerinin üretimi teşvik edilmelidir.
- Sektörde faaliyet gösteren KOBİ’lerin ihracata yönlendirilebilmeleri amacıyla, bu alanda firmalara maliyet, kalite, finansman, nitelikli işgücü, verimlilik gibi avantajlar sağlayacak “Sektörel Dış Ticaret Şirketi” (SDS) modeli altında örgütlenmeleri teşvik edilmelidir (Sezgin, 2002:128, Çelik-İş, 2003: 3, Atgür, 2006:54-56).

- Nihai ürün olarak müşteriye hitap eden demir çelik ürünlerinin şekillendiği haddehanelerin iyileştirilmesinin mutlaka sağlanması gerekmektedir.
- Müessese ürünlerinin iç ve dış piyasada rekabet edebilmesi için üretim hatalarının giderilmesi ve alt yapının güçlendirilmesi gerekmektedir.

Yapılan literatür çalışmalarında gerek yurt içinde gerekse de yurt dışında benzer bir çalışmaya fazlaca rastlanamamış olması hem tezin özgünlüğünün hem de veri madenciliği yöntemlerine üretim sektöründe özellikle de demir çelik sektöründe henüz gereken önemin verilmediğinin kanıtı sayılabilir.

Chaid algoritması ve çoklu doğrusal regresyon tekniklerinin kullanıldığı çalışmaya alternatif olarak YSA, CART gibi diğer teknikler kullanılarak yeni modeller kurulabilir.

Ayrıca üretim temelli bir modelleme olan bu çalışmanın yanı sıra uzun ürünlerin tüketimine dönük bir modelleme çalışmasının yapılması büyük önem taşımakta ve önerilmektedir.

KAYNAKLAR

Abhang, L. B., Hameedullah, M. (2012). Modeling and Analysis of Surface Roughness in Steel Turning Using Regression and Neural Networks. (ss. 317-322). India: *IEEE-International Conference on Advances in Engineering, Science and Management CAESM-2012*. March 30-March 31, 2012.

Akın-Koldere, Y. (2008). **Veri Madenciliğinde Kümeleme Algoritmaları ve Kümeleme Analizi**. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü.

Akpınar, H. (1998). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, (29)1:1-22.

Akyalçın, B., Yılmaz Ş. (2007). Çeliklerin Haddelenmesinde Malzeme Modelleme (ss. 231-242). Karabük: *IV. Demir Çelik Kongresi Bildiriler Kitabı*. 1 Kasım-3 Kasım, 2007.

Albayrak, A.S., Yılmaz, Ş.K. (2009). Veri Madenciliği: Karar Ağaçları Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1):31-52.

Alpar, R. (1997). *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1*. Ankara: Detay Yayıncılık.

Alpat, A. (2006). **Web Tabanlı Ortamda Olap Araçlarının Karar Destek Sistemlerinde Kullanılması**. Yayınlanmamış Yüksek Lisans Tezi. Eskişehir: Anadolu Üniversitesi Fen Bilimleri Enstitüsü.

Argüden, Y., Ersahin, B. (2008). *Veri Madenciliği*. İstanbul: ARGE Danışmanlık Yayınları.

Asilkan, Ö. (2008). **Veri Madenciliği Kullanılarak İkinci El Otomobil Pazarında Fiyat Tahmini**. Yayınlanmamış Yüksek Lisans Tezi. Antalya: Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü.

Ataseven, S. (2008). **Üniversitelerin Adaylar Tarafından Tercih Edilme Desenlerini Veri Madenciliği Yöntemleri İle Belirleyen Bir Model Önerisi**. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: İstanbul Kültür Üniversitesi Fen Bilimleri Enstitüsü.

Atgür, M. (2006). **Türkiye’de Demir Çelik Sektörü: Analiz, Sorunlar ve Çözüm Önerileri**. Yayınlanmamış Yüksek Lisans Tezi. Balıkesir: Balıkesir Üniversitesi Sosyal Bilimler Enstitüsü.

Başbakanlık Yüksek Denetleme Kurulu (BYDK). (1995). *Karabük Demir Çelik Fabrikaları Müessesesi*. Ankara: Türkiye Demir Çelik İşletmeleri 1994 Yılı Raporu.

Batı Karadeniz Kalkınma Ajansı-BAKKA. (2009). *Batı Karadeniz Bölgesi Bölge Planı 2010-2013*. Zonguldak.

Belen, M., Yıldırım, M., Çakmak, A. Ç. (2011). *Demir Çelik Uzun Ürünleri Pazar Araştırması Çalışma Grubu Raporu*. Karabük: Karabük Üniversitesi Demir Çelik Enstitüsü Yayınları.

Bilekdemir, G. (2010). **Veri Madenciliği Tekniklerini Kullanarak Üretim Süresi Tahmini Ve Bir Uygulama**. Yayınlanmamış Yüksek Lisans Tezi. İzmir: Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü.

Bozkır, A. S. (2009). **Olap ve Veri Madenciliği Teknolojilerinden Yararlanılarak Web Tabanlı Bir Karar Destek Sisteminin Gerçekleştirilmesi**. Yayınlanmamış Yüksek Lisans Tezi. Ankara: Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T. (2001). Step by Step Data Mining Guide.

Chen, L., Sakaguchi, T., Frolick, M. N. (2000). Data Mining Methods, Applications and Tools. *Information Systems Management*, 17(1):1-6.

Chien, C., Wang, W., Cheng, J. (2006). Data Mining For Yield Enhancement in Semiconductor Manufacturing and an Empirical Study. *Expert Systems with Applications*, 16(1):5-7

Cihan, A. (2009). **Pozitif ve Negatif İlişkilerin Veri Madenciliği ile Belirlenmesine Yönelik Bir Model**. Yayınlanmamış Yüksek Lisans Tezi. Kocaeli: Kocaeli Üniversitesi Fen Bilimleri Enstitüsü.

Corporate & Public Strategy Advisory Group (CPS). (2011). *Çelik Sektörüne İlişkin AB Müktesebat Rehberi*. Brüksel-İstanbul.

Coşkun, C. (2010). **Veri Madenciliği Algoritmaları Karşılaştırılması**. Yayınlanmamış Yüksek Lisans Tezi. Diyarbakır: Dicle Üniversitesi Fen Bilimleri Enstitüsü.

Cross- Industry Standart Process for Data Mining (CRISP – DM).

<http://www.crisp-dm.org>, (01.09.2012).

Çakır, A.M. (2006). **Kok Fırın Gazı Prosesleri ve Ek Tesisleri**. Yayınlanmamış Yüksek Lisans Tezi. Erzurum: Atatürk Üniversitesi Fen Bilimleri Enstitüsü.

Çelik, M. (2009). **Veri Madenciliğinde Kullanılan Sınıflandırma Yöntemleri ve Bir Uygulama**. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Çetin, M. (2009). **Bir Üretim İşletmesinde Veri Madenciliği Uygulaması**. Yayınlanmamış Yüksek Lisans Tezi. Sakarya: Sakarya Üniversitesi Fen Bilimleri Enstitüsü.

Çetinyokuş, T. (2008). **Veri Küplerinin Bütünleşik Kullanımına Yönelik Yeni Bir Olap Mimarisi**. Yayınlanmamış Yüksek Lisans Tezi. Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

Çiflikçi, C., Kahya-Özyirmidokuz, E. (2012). Enhancing Product Quality of a Process. *Industrial Management & Data Systems*, 112(8):1181–1200.

Çil, F. (2010). **Banka Yatırım Fonu Müşteri Hareketlerinin Belirlenmesine Yönelik Bir Veri Madenciliği Uygulaması**. Yayınlanmamış Yüksek Lisans Tezi. Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

Çoban, A. (2006). **İmalat Sanayinde Veri Madenciliği Destekli Tedarikçi Seçimi Uygulaması**. Yayınlanmamış Yüksek Lisans Tezi. Sakarya: Sakarya Üniversitesi Sosyal Bilimler Enstitüsü.

Demir Çelik Store. (2012). *Amerika Çelik Üreticileri Maliyetleri Kısımak İçin Doğalgazı Tercih Ediyor*. İstanbul Demir Çelik Store Dergisi.

Demir Çelik Üreticileri Derneği (DÇÜD). (2011). *Türk Demir Çelik Sektörü 2010 Yılı Raporu*. Ankara: Çelik Dergisi.

Demirel, B. (2010). **Veri Madenciliğinde Chaid Algoritmasının Sosyal Güvenlik Kurumu Veri Tabanına Uygulanması**. Yayınlanmamış Yüksek Lisans Tezi. Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

Deng, B., Liu, X. (2002). Data Mining in Quality Improvement. *SUGI27: Proceedings of the Twenty-Seventh Annual SAS* (ss. 15-25). Orlando, Florida: Users Group International Conference. April 14–April 17, 2002.

Dolgun, M.Ö. (2006). **Büyük Alışveriş Merkezleri İçin Veri Madenciliği Uygulamaları**. Yayınlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi Fen Bilimleri Enstitüsü.

Dunham, M.H. (2003). *Data Mining Introductory and Advanced Topics*. India: Pearson Education Inc.

Efe, M. Ö. (2006). *Yapay Sinir Ağları ve Uygulamaları*. İstanbul: Boğaziçi Üniversitesi Yayınları.

Emel, G. G., Taşkın, Ç., Kılıçarslan, S. (2004). Sinir Ağları Veri Madenciliği İle Çelik Üretim Sürecinde Bir Analiz. *Journal of Faculty of Business*, 5(1):205-225.

Eruz, B. (2003). **Türk Demir Çelik Sektörünün Yeniden Yapılandırılması**. Yayınlanmamış Yüksek Lisans Tezi. Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.

Escribano, D. (2012). Modelling a Skin-Pass Rolling Process by Means of Data Mining Techniques and Finite Element Method. *Journal of Iron and Steel*, 19(5):43-49.

Esen, F. (2009). **Veritabanlarından Bilgi Keşfi: Veri Madenciliği ve Bir Sağlık Uygulaması**. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü.

Eurofer. (2011). *Ekonomi ve Çelik Piyasalarının Görünümü 2011-2012 Raporu, Demir Çelik Sektörü / Rekabet Gücü Raporu*. Ankara: Sektörel Rekabet Gücü Raporları Dizisi.

Fernandez, G. (2003). *Data Mining Using SAS Applications*. London: Chapman & Hall/CRC.

Fung, G., Mangasarian, O. L. (2002). Incremental Support Vector Machine Classification (ss. 4-13). Madison: *Second SIAM International Conference on Data Mining*.

- Giudici, P. (2003). *Applied Data Mining*, London: Wiley.
- Gu, L. (2011). Parameters Optimization of Plasma Hardening Process Using Genetic Algorithm and Neural Network. *Journal of Iron and Steel*, 18(12):57-64.
- Günay, M.H. *Data Mining Introductory and Advanced Topics*. New Jersey: Pearson Education Inc.
- Gürsoy, U. T. Ş. (2009). *Veri Madenciliği ve Bilgi Keşfi*. Ankara: Pegem Akademi.
- Han, J., Kamber, M., Pei, J. (2012). *Data Mining Concepts And Techniques*. USA: Morgan Kaufmann.
- Hong-Bing, W. (2012). Prediction of Endpoint Phosphorus Content of Molten Steel in BOF Using Weighted K-Means and GMDH Neural Network. *Journal of Iron and Steel*, 19(1):11-16.
- Inmon, W. H. (2000). *What is Data Warehouse*. London: Prism.
- Irmak, S. (2009). **Veri Madenciliği Yöntemleri ile Sağlık Sektörü Veritabanlarında Bilgi Keşfi: Tanımlayıcı ve Kestirimci Model Uygulamaları**. Yayınlanmamış Yüksek Lisans Tezi. Antalya: Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü.
- İskenderun Demir Çelik Fabrikaları (İsdemir).
<http://www.isdemir.com>, (05.08.2012).
- İzmir, M. (2011). *Demir Çelik Hurda kalitesi ve Eksperliği*. İstanbul: Bilnet Matbaacılık.
- Kakıcı, A. (19.03.2012). Yapay Sinir Ağlarına Giriş.
<http://www.ahmetkakici.com/yazilim/yapay-sinir-aglarina-giris> (15 Kasım 2012)
- Kalaycıoğlu, O. (2006). **Kardemir’de Ray Üretiminde İyileştirmeler**. Yayınlanmamış Yüksek Lisans Tezi. Sakarya: Sakarya Üniversitesi Fen Bilimleri Enstitüsü.
- Karabük Demir Çelik Fabrikaları (Kardemir).
<http://www.kardemir.com>, (15.08.2012).
- King, R. D., Feng, C., Sutherland, A. (1995). StatLog: Comparison of Classification Algorithms on Large Real-World Problems. *Applied Artificial Intelligence*, 9(3):289-333.

Koyuncugil, A.S. (2007). *Borsa Şirketlerinin Sektörel Risk Profillerinin Veri Madenciliğiyle Belirlenmesi*. Sermaye Piyasası Kurulu Araştırma Raporu. Ankara.

Kurşun, S. (2007). **Tekstil Endüstrisinde Benzetim Tekniği İle Üretim Hattı Modellemesi Ve Uygun İş Akış Stratejisinin Belirlenmesi**. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: İstanbul Üniversitesi Fen Bilimleri Enstitüsü.

Larose, D.T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. John Wiley & Sons Inc.

Li, M., Feng, S., Sethi, I. K., Luciw, J., Wagner, K. (2003). Mining Production Data with Neural Network & CART (ss. 731-734). Melbourne, Florida, USA: *Proceedings of the Third IEEE International Conference on Data Mining*.

Mattison, R. (1997). *Data Warehousing and Mining for Telecommunications*. Norwood: Artech House.

Meyer, D., Cannon, C. (1998). *Building A Better Data Warehouse*. USA: Prentice Hall.

Mieno, F., Sato, T., Shibuya, Y., Odagiri, K., Tsuda, H., Take, R. (1999). Yield Improvement Using Data Mining System (ss. 391-394). USA: *Conference on Semiconductor Manufacturing*.

Milli Eğitim Bakanlığı (MEB) Mesleki Eğitim Ders Kitabı. (1995). *Metal Meslek Bilgisi*. Ankara: MEB Yayınları.

Montgomery, D., Peck, E. (1991). *Introduction to Linear Regression Analysis*. New York: John Wiley.

Oğuzlar, A. (2004). Cart Analizi ile Hanehalkı İşgücü Anketi Sonuçlarının Özetlenmesi. *Uludağ Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 18(3):79-90

Olson, A., Shi, M. (2007). Applying Data Mining to Manufacturing: The Nature and Implications (ss.24-26). Irwin: *International Iron and Steel Symposium*.

Onurlubaş, H. E. (2007). **Türkiye’de Bitkisel Yağ Sanayindeki Gelişmeler ve Geleceğe Yönelik Beklentiler**. Yayınlanmamış Yüksek Lisans Tezi. Ankara: Tarımsal Ekonomi Araştırma Enstitüsü.

- Orhunbilge, N. (2010). *Çok Değişkenli İstatistik Yöntemler*. İstanbul: İstanbul Üniversitesi Yayınları.
- Özçınar, H. (2006). **Kpss Sonuçlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi**. Yayınlanmamış Yüksek Lisans Tezi. Denizli: Pamukkale Üniversitesi Fen Bilimleri Enstitüsü.
- Özden, G. A., Chen, Y. (1999). Design Quality and Robustness with Neural Networks. *IEEE Transactions On Neural Networks*, 10(6):221-229.
- Özkan, Y. (2008). *Veri Madenciliği Yöntemleri*. Ankara: Papatya Yayıncılık.
- Özyirmidokuz, E. K. (2009). **Veri Madenciliği Tekniklerini Kullanarak İmalat Verilerinin Modellenmesi ve Analizi**. Yayınlanmamış Yüksek Lisans Tezi. Kayseri: Erciyes Üniversitesi Sosyal Bilimler Enstitüsü.
- Park, S. C., Selwyn, P., Shaw, M. J. (2001). Dynamic Rule Refinement in Knowledge-Based Data Mining Systems. *Decision Support Systems*, 31:205-222.
- Payas TEML, Hak-İş/Çelik-İş Sendikası. (2011). *Sosyal Diyalog Yoluyla Metal Sektöründe Mesleki Yeterliliklerin Geliştirilmesi Projesi (SODİMGEP)*. Metal Sektörü Yeterlilikler Kılavuzu. Hatay: Payas TEML Yayın.
- Quinlan, J.R. (1993). *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann.
- Rajavarman, V.N., Rajagopalan, S.P. (2007). Comparison Between Traditional Data Mining Techniques and Entropy-Based Adaptive Genetic Algorithm for Learning Classification Rules. *International Journal of Soft Computing*, 2(4):555-561.
- Sanayi ve Ticaret Bakanlığı, Sanayi Genel Müdürlüğü. (2011). *Demir Çelik Sektörü Raporu*. Ankara.
- Serper, Ö. (1986). *Uygulamalı İstatistik 2*. İstanbul: Filiz Kitabevi.
- Silahtaroglu, G. (2008). *Kavram ve Algoritmalarıyla Temel Veri Madenciliği*. İstanbul: Papatya Yayıncılık.
- Skinner, K. R., Montgomery, D. C., Runger, G. C., Fowler, J. W., Mccarville, D. R., Rhoads, T. R. (2002). *Multivariate Statistical Methods For Modeling and Analysis of*

Wafer Probe Test Data. *IEEE Transactions on Semiconductor Manufacturing*, 15(4):523–530.

Söylemez, A. (2009). **Bireysel Müşterilerin Kredi Değerlendirme Sonuçlarını En İyi Tahmin Eden Scorecard Modelinin Oluşturulması**. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: Mimar Sinan Üniversitesi Fen Bilimleri Enstitüsü.

SPSS Türkiye. (2010). *SPSS Modeller Temel Eğitimi*. Ankara. http://pic.dhe.ibm.com/infocenter/spssmodl/v15r0m0/index.jsp?topic=%2Fcom.ibm.sps.s.modeler.help%2Funderstanding_modeltypes.htm (11.12.2012).

Sumathi, S., Sivanandam S.N. (2006). *Data Mining Tasks, Techniques and Applications, Studies in Computational Intelligence (SCI)*. Berlin: Springer-Verlag.

Şekeroğlu, S. (2010). **Hizmet Sektöründe Bir Veri Madenciliği Uygulaması**. Yayınlanmamış Yüksek Lisans Tezi. İstanbul: İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü.

T.C. Bilim, Sanayi ve Teknoloji Bakanlığı. (2012). *Demir Çelik Sektörü Raporu*. Ankara.

T.C. Sanayi ve Ticaret Bakanlığı. (2011). *Demir Çelik Sektörü Raporu*. Ankara.

Tan, H. (2000). *Introduction to Data Mining*. London: Wiley.

Türkiye İstatistik Kurumu (TÜİK). (2010). *İstatistikî Göstergeler 1923-2009*. Ankara.

Türkiye Odalar ve Borsalar Birliği (TOBB). (2010). *Türkiye Demir ve Demir Dışı Metaller Meclisi Sektör Raporu*. Ankara.

Türkiye Taş Kömürü Kurumu (TTK). (2006). *Taşkömürü Çalışma Grubu Raporu*, Zonguldak: Dokuzuncu Kalkınma Planı (2007- 2013).

Ülengin, F., Önsel, Ş., Çekyay, B., Özaydın, Ö., Aktaş, E., Kabak, Ö., TÜSİAD, Sabancı Üniversitesi Rekabet Forumu (REF), Sektörel Dernekler Federasyonu (SEDEFED). (2011). *Demir Çelik Sektörü Rekabet Gücü Raporu 2011*. İstanbul: Sektörel Rekabet Gücü Raporları Dizisi.

Wang, K. (2007). *Applying Data Mining to Manufacturing: The Nature and Implications*. China: Springer Science.

Wong, M.L., Leung, K. S. (1999). *Data Mining Using Gramer Based Genetic Programming and Applications*. London: Kluwer Academic Publishers.

World Steel Association. (2010). Steel Statistical Yearbook 2010. <http://www.worldsteel.org>, (03.07.2012).

Yarımağan, Ü. (2000). *Veri Tabanı Sistemleri*. Ankara: Akademi Kitabevi.

Yavuz, M. A. (2009). **Kaçak Su Kullanımının Tespitinde Veri Madenciliği Yaklaşımı**. Yayınlanmamış Yüksek Lisans Tezi. Sakarya: Sakarya Üniversitesi Fen Bilimleri Enstitüsü.

ÖZGEÇMİŞ

Bülent ORDU 1976 yılında Yozgat' ta doğdu; ilköğrenimini İstanbul' da, orta öğrenimini ise Adana'da tamamladı. Düziçi Anadolu Öğretmen Lisesi'nden mezun oldu. 1995 yılında Abant İzzet Baysal Üniversitesi Fen Edebiyat Fakültesi Matematik Bölümü'nde öğrenime başlayıp 2000 yılında iyi derece ile mezun oldu. 2000 yılından itibaren 10 yıl süre ile çeşitli özel dershanelerde Matematik ve Geometri Öğretmenliği yaptıktan sonra 2010 yılında Karabük Üniversitesi Eskipazar Meslek Yüksekokulu'nda Öğretim Görevlisi olarak göreve başladı. 2010 yılında Karabük Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı'nda başlamış olduğu yüksek lisans programını, aynı yerde tamamladı. Halen Karabük Üniversitesi'nde çalışmaya devam etmektedir.