

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**VERİ MADENCİLİĞİ KULLANARAK BİYOKİMYA
VERİLERİNDEN HASTALIK TEŞHİSİ**

Şengül DOĞAN

Tez Yöneticisi
Yrd. Doç.Dr. İbrahim TÜRKOĞLU

YÜKSEK LİSANS TEZİ
BİYOMÜHENDİSLİK ANABİLİM DALI

ELAZIĞ, 2007

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

VERİ MADENCİLİĞİ KULLANARAK BİYOKİMYA VERİLERİNDEN HASTALIK TEŞHİSİ

Şengül DOĞAN

Yüksek Lisans Tezi
Biyomühendislik Anabilim Dalı

Bu tez, 12/01/2007 tarihinde aşağıda belirtilen jüri tarafından oybirliği /oyçokluğu ile başarılı olarak değerlendirilmiştir.

Danışman: Yrd.Doç.Dr. İbrahim TÜRKOĞLU

Üye: Prof. Dr. Bilal ÜSTÜNDAĞ

Üye: Yrd.Doç.Dr. Arif GÜLTEN

Bu tezin kabulü, Fen Bilimleri Enstitüsü Yönetim Kurulu'nun/...../..... tarih ve sayılı kararıyla onaylanmıştır.

TEŐEKKÖR

Bu tez alıřmasının gerekleřtirilmesinde  zverili bir řekilde gerekli y nlendirici desteęi saęlayan ve yardımlarını esirgemeyen ok deęerli hocam Sayın Yrd. Do. Dr. İbrahim T RK  LU'na, bug nlere gelmemizde emeęi geen t m hocalarıma, belirlenen hastalıklarda hasta verilerini bize saęlayan Dr. Bilal  ST NDAę, Dr. Mustafa F. YAVUZKIR, Dr. Osman Metin İPİ  LU ve Dr. Ebru AęEL'e ve alıřmam boyunca sevgi ve g venlerini her zaman hissettięim aileme teőekk rlerimi ve ř kranlarımı sunarım.

Őeng l DOęAN

İÇİNDEKİLER

Sayfa

TEŞEKKÜR

İÇİNDEKİLER I

ŞEKİLLER LİSTESİ III

TABLolar LİSTESİ V

SİMGELER LİSTESİ VII

KISALTMALAR LİSTESİ VII

ÖZET VIII

ABSTRACT IX

1. GİRİŞ 1

2. VERİ MADENCİLİĞİ 7

2.1. Veri Ambarları ve Veri Madenciliği 7

2.1.1. Veri Ambarının Karakteristik Özellikleri 9

2.1.2. Veri Ambarının Ayırt Edici Özellikleri 10

2.1.3. Veri Ambarları için Veri Modelleme 10

2.1.4. Bir Veri Ambarı Oluşturma 11

2.1.5. Veri Ambarlarının Uygulanmasındaki Zorluklar 12

2.2. Bir Veri Madenciliği Çalışması Örneği 13

2.3. Metodoloji 15

2.4. Kullanılan Teknikler 16

2.5. Sınıflama ve Regresyon Modelleri 17

2.6. Kümeleme Modelleri 18

2.7. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler 18

2.8. Veri Tabanlarında Bilgi Keşfi Süreci 19

2.8.1. Problemin Tanımlanması 19

2.8.2. Verilerin Hazırlanması 19

2.8.3. Modelin Kurulması ve Değerlendirilmesi 21

2.8.4. Modelin Kullanılması 23

2.8.5. Karar Ağaçları 23

2.9. Veri Madenciliği Sistemleri Üzerine Yapılan Çalışmalar 26

2.9.1. Analysis Manager 26

2.9.2. Darwin 26

2.9.3. Clementine 27

2.9.4. DBMiner 27

2.9.5. Data Logic/R 29

2.9.6. INLEN.....	29
2.9.7. KDW (<i>Knowledge Discovery Workbench</i>)	29
2.9.8. SKICAT (<i>Sky Image Classification & Archiving Tool</i>).....	29
2.9.9. R-MINI.....	30
2.9.10. TASA (<i>Telecommunication Network Alarm Sequence Analyzer</i>)	30
2.9.11. GCLUTO (<i>Graphical CLUstering TOolkit</i>)	30
2.9.12. Enterprise Miner	31
2.9.13. Veri Madenciliği Araçlarının Karşılaştırmaları	31
2.10. Uygulama Alanları	36
2.10.1. Biyomedikal için Veri Madenciliği.....	36
2.10.2. Bankacılık ve Finansal Uygulamalarındaki Veri Madenciliği	37
2.10.3. Perakende Sanayii Uygulamalarındaki Veri Madenciliği	39
2.10.4. Telekomünikasyon Sanayii Uygulamalarındaki Veri Madenciliği.....	40
2.10.5. Sağlık Hizmet Uygulamalarındaki Veri Madenciliği.....	41
2.11. Görsel ve Sesli Veri Madenciliği	41
3. VERİ TABANLARINDAN BİLGİ ÇIKARIMI.....	46
3.1 Bilgi Tabanlarının Kurulum Aşamaları.....	46
3.2 Disiplin Alanının Tasviri.....	46
3.2.1 Çözülen Problemlerin Karakterinin Belirlenmesi	48
3.2.2 Disiplin Alanındaki Nesnelerin Belirlenmesi	50
3.2.3 Nesneler Arasındaki İlişkilerin Saptanması	50
3.2.4 Bilginin Sunulması Modelinin Seçilmesi.....	50
3.2.5 Disiplin Alanının Özel Niteliklerinin Meydana Çıkarılması	53
3.3 Bilginin Elde Edilmesi	53
3.4 Bilgi Tabanlarında Bilginin Tamlığının Ve Zıddiyetsizliğinin Kontrol Edilmesi	55
4. GELİŞTİRİLEN HASTALIK TEŞHİS SİSTEMLERİ	58
4.1. Kardiyak Enzim Parametrelerinden Kalp Krizi Teşhisi	62
4.1.1. Kardiyak Enzimler	63
4.1.2 Yöntem.....	64
4.1.3 Uygulama Sonuçları.....	68
4.2. Lipid Parametrelerinden Hiperlipidemi Teşhisi.....	68
4.2.1. Lipidler.....	69
4.2.2. Yöntem.....	70
4.2.3. Uygulama Sonuçları.....	74
4.3. Hematolitik Parametrelerinden Demir Eksikliği Anemisi Teşhisi	74

4.3.1. Hematolik Tablosu Değerleri	75
4.3.2. Yöntem.....	76
4.3.3. Uygulama Sonuçlar	78
4.4. Tiroid Hormon Parametrelerinden Hipotiroidi-Hipertiodi Teşhisi	78
4.4.1.Hormonlar	79
4.4.2. Yöntem.....	80
4.4.3. Uygulama Sonuçları.....	84
5. SONUÇ VE DEĞERLENDİRME	85
5.1.Sonuçların İrdelenmesi.....	85
KAYNAKLAR	86

ŞEKİLLER LİSTESİ

Şekil 2.1. Veri madenciliği yöntemi	7
Şekil 2.2. Günlük veri tabanlarında istenen özet bilginin seçilmesi	8
Şekil 2.3. Veri ambarının yapısı	9
Şekil 2.4. Veri örneği	14
Şekil 2.5. Karar ağacı tarafından tanımlanan sınır	14
Şekil 2.6. Tanımlanan karar ağacının yapısı	15
Şekil 2.7. Veri madenciliği çalışmasında kullanılan metodoloji	16
Şekil 2.8. DBMiner sisteminin yazılım mimarisi	28
Şekil 2.9. GCLUTO arayüzünün görüntüsü	31
Şekil 2.10. NETPROPHET stok tahminleme	38
Şekil 2.11. PILOT Pazara yönelik kampanyalar	39
Şekil 2.12. MapInfo hastaların ve en yakın sağlık ekiplerinin yerinin haritalanması	41
Şekil 2.13. Statsoft'un çok boyutlu uzayda veri dağılımı	42
Şekil 2.14. SAS Emterprise Miner'in scatter plot ile yaptığı çalışma	43
Şekil 2.15. İlişkisel kuralların görüntülenmesi MineSet 3.0	43
Şekil 2.16. MineSet 3.0 karar ağacı	44
Şekil 2.17. IBM Intelligent Miner'dan, küme gruplanması	44
Şekil 2.18. Algısal tabanlı sınıflandırma	45
Şekil 3.1. Herhangi bir disiplin alanında bilgi tabanının oluşturulması safhaları	48
Şekil 3.2. Disiplin alanındaki hiyerarşi tipleri ve onlar arasındaki ilişkiler	51
Şekil 4.1. Örüntü tanıma yapısı	62
Şekil 4.2. Karar destek sistemi algoritması	65
Şekil 4.3. Karar ağacının yapısı	67
Şekil 4.4. Karar destek sistemi algoritması	71
Şekil 4.5. Karar ağacının yapısı	73
Şekil 4.6. Karar destek sistemi algoritması	76
Şekil 4.7. Karar ağacının yapısı	77
Şekil 4.8. Karar destek sistemi algoritması	81
Şekil 4.9. Karar ağacının yapısı	83

TABLolar LİSTESİ

Sayfa

Tablo 2.1. Örnek için standart form	13
Tablo 2.2. Modelin kurulması ve değerlendirilmesi için bir örnek tablo	21
Tablo 2.3. Örnek için standart bir form	22
Tablo 2.4. Kural çıkarımı	24
Tablo 2.5. Veri madenciliği araçlarının anahtarlarının anlamları.....	31
Tablo 2.6. Veri madenciliği araçlarının çalışabildiği platformlar	32
Tablo 2.7. Veri madenciliği araçlarının veri girdisi ve çıktısı açısından karşılaştırılması	33
Tablo 2.8. Veri madenciliği araçlarının desteklediği karar ağaçları.....	33
Tablo 2.9. Veri madenciliği araçlarının kullanılabilirlik karşılaştırılması	34
Tablo 2.10. Veri madenciliği araçlarının görüntülenme açısından karşılaştırılması	34
Tablo 2.11. Veri madenciliği araçlarının otomasyon açısından karşılaştırılması.....	35
Tablo 2.12. Veri madenciliği araçlarının güçlü ve zayıf olduğu alanlar	35
Tablo 3.1. Çözüm tablosunun genel yapısı	56
Tablo 4.1. Kardiyak enzimlerin yükselme zamanları.....	64
Tablo 4.2. Sistemin kural yapısı	66
Tablo 4.3. Kardiyak enzimlerin normal değerleri	66
Tablo 4.4. Önerilen akıllı karar destek sisteminin başarımı	68
Tablo 4.5. Sistemin kural yapısı	72
Tablo 4.6. Lipidlerin normal değerleri	72
Tablo 4.7. Önerilen akıllı karar destek sisteminin başarımı	74
Tablo 4.8. Sistemin kural yapısı	77
Tablo 4.9. Hemalotik parametrelerin normal değerleri	77
Tablo 4.10. Önerilen akıllı karar destek sisteminin başarımı	78
Tablo 4.11. Sistemin kural yapısı	82
Tablo 4.12. Hormonların normal değerleri.....	82
Tablo 4.13. Karar destek sisteminin performansı.....	84

SİMGELER LİSTESİ

$Q1, Q2$	: İki boyutta eşik değerler
Ω	: Karar sınıf elemanlarından herbiri
Ψ	: Sınıf elemanlarından varılacak sonuçlar
C_i	: Koşullar
R_i	: Çözüm kuralları
T_i	: Koşulların karmaşıklığı
A_i	: Eylemler
q_i	: Eylemlerin karmaşıklığı
f_n	: Kuralların sıklığı

KISALTMALAR LİSTESİ

EKG	: ElektroKardiyoGrafı
COM	: Bileşen nesne modeli
OLAM	: Çevirimiçi analitik madencilik
OLAP	: Çok boyutlu analiz
YSA	: Yapay sinir ağları
AID	: Otomatik etkileşimli belirleyici
MI	: Miyokard Infarktüsü
SDBK	: Serum demir bağlama kapasitesi
VTBK	: Veri Tabanlarında Bilgi Keşfi Süreci
MPP	: Büyük Paralel İşlemler
SMP	: Simetrik Çoklu İşlemler
MİY	: Müşteri İlişkileri Yönetimi
ÇT	: Çözüm Tablosu

ÖZET

YÜKSEK LİSANS TEZİ

VERİ MADENCİLİĞİ KULLANARAK BİYOKİMYA VERİLERİ İLE HASTALIK TEŞHİSİ

Şengül DOĞAN

Fırat Üniversitesi

Fen Bilimleri Enstitüsü

Biyomühendislik Anabilim Dalı

2007, Sayfa: 91

Elektronik değişimin yaşandığı günümüzde tıp, sanayi, eğitim gibi pek çok alanda veri hacmi hızla artmaktadır. Bu veriler geniş veri havuzları oluşturmakta ve mevcut verilerden amaçlanan verilere ulaşmak oldukça zorlaşmaktadır. Özellikle tıbbi veri tabanlarında verilerin analizi, karar alma ve yönetim aşamalarında yeni bilgilerin keşfedilmesi ve bu bilgilerin etkili bir şekilde kullanılması işlemlerinin bilgisayar tarafından seri ve hızlı bir şekilde yaptırılması uzmanlara büyük kolaylık sağlamaktadır.

Bu tez çalışmasında, Miyokard infarktüsü, Hiperlipidemi, Demir eksikliği anemisi ve Hipertiroidi-Hipotiroidi hastalıklarının teşhisinde kullanılacak bir karar destek sistemi tasarlanmıştır. Karar destek sisteminde veri madenciliği tekniklerinden karar ağaçları yöntemi kullanılmıştır. Sistemde belirlenen her bir hastalık için doktorların tanıda bulunurken kullandıkları biyokimya verilerinden seçilen parametreler baz alınarak değerlendirme yapılmıştır. Bu amaçla, delphi programlama dili kullanılarak bir paket program geliştirilmiştir. Sonuçta elde edilen veriler uzman bir doktorun verdiği kararlarla karşılaştırılmıştır.

Anahtar kelimeler: Veri madenciliği, Örüntü tanıma, Uzman sistemler, Karar ağaçları, Karar destek sistemleri, Biyokimya parametreleri.

ABSTRACT
MASTER THESIS

AN ANALYSE OF A DESEASES WITH BIOCHEMICAL DATA BY USING DATA
MINING

Şengül DOĞAN

Firat University

Graduate School of Natural and Applied Science

Department of Bioengineering

2007, Page: 91

An electronic revolution is currently occurring; it causes a very fast development of the data volume in many fields like medicines, industry, and education. These data form vast data pools, and it gets more and more difficult to obtain the aimed data from the data available. Especially the analyze of the data which is available in the medical data bases, making decisions, and the discovery of new data in the administrative phases and the impressive application of these data, the fast and practical processes carried out by computers have eased the facilities of the experts.

A decision support system which will be used in analyzing deceases like myocardial infarctus, hyperlipidemics, iron deficiency anaemia, and hypertiroide- hypotiroidy has been planned in this thesis. One of the data mining methods; the decision trees method has been applied in this decision support system. The parameters which have been selected from the biochemical data, which were applied by the doctors while analyzing each decease which was detected in this system, were considered as a base for the evaluation which was carried out. The results of this method were thoroughly in compliance with the decision of an expert.

Key words: Data Mining, Pattern recognition, Expert systems, Decision trees, Decision support systems, Biochemistry parameters.

BÖLÜM 1

1.GİRİŞ

Yeni teknolojiler ve otomatikleşen sistemler sayesinde büyük organizasyonlar verimli ve ucuz bir yolla müşterilerinin, çalışanlarının ve faaliyetlerinin verilerini oluşturur ve depolarlar. Elektronik verinin bu koleksiyonu veritabanlarında tutulur veya çok büyük boyuttaki koleksiyonlar için ise veri ambarları kullanılır. Veritabanları ve veri ambarları sürekli olarak geliştirilir ve güncelleştirilir. Bu veriler gelecekte beklenen akımları tahmin etmeye veya analizler yapmaya yardımcı olurlar [1].

Veri madenciliği; anlamlı modeller ve kurallar ortaya çıkarmak amacıyla, otomatik ve yarı-otomatik yöntemler kullanarak büyük miktarlardaki veriyi analiz etme ve inceleme süreci olarak tanımlanabilir [1].

Veri ambarlarının boyutlarının büyümesi ile birlikte veri madenciliği teknikleri kullanılarak yapılan yüksek seviyeli otomatik işlemlerle bilgi arama işlemi, insan analistler için kolaylaşmaktadır. Özellikle tıp alanında çalışanlar için istatistiksel ve matematiksel işlem yapmadan hemen sonuca ulaşmaları, karar verme sürecinde oldukça faydalı olacaktır [1].

Veri madenciliği ile büyük veri kümeleri üzerinde ilginç ve gizli kalmış örüntüleri tespit etmek mümkündür. Bu noktada veri madenciliği istatistikten ayrılır. İstatistikçiler bilinen faktörler arasındaki ilişkilerin güçlülüğünü araştırır. Veri madenciliğinde ise bilinmeyen faktörler arasında tahmin edilemeyecek ilişkilerin güçlülüğü araştırılır. Örneğin “Çocuk bezi alan müşterilerin %30’u süte satın alır” gibi bir ilişkiyi istatistik yaklaşımları ile bulmak zordur. Çünkü “Çocuk bezi” ve “Süt” bizim için arasında ilişki olduğu düşünülecek ürünler değildir. Ancak veri madenciliği yaklaşımından baktığımızda “Çocuk bezi” ve “Süt” arasındaki ilişkiyi önceden tahmin edemediğimiz halde, bu teknikler ile keşfedilebilecek bir ilişkidir [1].

Veri madenciliği çeşitli araştırma sahalarından farklı teknik ve metotları içermektedir. Yapay zekâ, makine öğrenmesi gibi alanlardaki birçok algorithmadan faydalanılır. Genetik algoritmalar, karar ağaçları, kaba kümeler, bulanık mantık kullanılan algoritmalarından bazılarıdır. Veri madenciliğinin geniş bir uygulama alanı mevcuttur. Bunlarla müşterilerin satın alma eğilimlerinin ve alışkanlıklarının belirlenmesi, web sayfalarına erişimlerin analizi, ilaçların yan etkilerinin ortaya çıkartılması, hastalıklar ve semptomlar arasında bağlantıların kurulması, bankacılıkta sahteciliklerin tespit edilmesi, kredi borcu ödeme tahminleri gibi birçok alanda karar verme işlemlerine yardımcı olacak önemli sonuçlar elde edilmektedir [1].

Elektronik bir devrimin yaşandığı günümüzde her alanda olduğu gibi tıp alanında da potansiyel olarak depolanan veri hacmi hızla artmaktadır. Bununla birlikte bu verilerden faydalı bilgiler elde etmek giderek zorlaşmaktadır. Ancak bu konu göz ardı edilemeyecek kadar

önemlidir. Tıbbi veri tabanlarında yeni bilgilerin keşfedilmesi, veritabanlarında depolanan verinin etkili bir şekilde kullanılması açısından çok önemlidir. Son zamanlarda tıbbi veritabanlarında depolanan veri miktarı büyük bir hızla artmaktadır. Bu büyümeye karşılık, geleneksel veri analiz araçları yetersiz kalmaktadır. Verilerin analizi karar alma ve yönetim aşamalarında çok önemlidir [2].

Büyük boyutlardaki veri ile çalışan birçok sektörde “veri” değil “enformasyon” ve “bilgi” ön plandadır. Bu sektörlerin başında da sağlık sektörü gelmektedir. Çünkü sağlık sektöründe gizli bilgilerin keşfedilmesi hayati önem taşımaktadır. Örneğin semptomlar ve hastalıklar arasındaki gizli kalmış ilişkilerin bulunması, tecrübe ile elde edilebilecek bilgilerin önceden keşfedilmesi, çok faktörlü hastalıklarda önemli faktörler arasındaki ilişkilerin ortaya çıkarılması gibi birçok alanda tıbbi veri madenciliğinin önemi tartışılmazdır. “Veritabanında Bilgi Keşfi” gizli olanın ortaya çıkarılması işlemini tanımlar [2].

Veriden daha önceden bilinmeyen ilginç ve yararlı bilgi elde etmeyi amaçlar. Bu amaçla Veri madenciliği teknikleri kullanılır. Çok büyük miktarda veri toplayan işletmeler ellerindeki veriden geleceğe yönelik tahminler yapmanın önemini kavramışlardır. Bu bağlamda 1990’lu yıllara kadar daha çok, verinin toplanması ve depolanması ile ilgilenilmiş 1990’lu yıllardan itibaren ise veri ambarlarının kullanımının yaygınlaşması ve verinin analizi ile birlikte veri madenciliği ön plana çıkmıştır [2].

Her ne kadar tıbbi veriler ile diğer veriler arasında temelde çok fazla fark olmasa da, özellikle uygulamada önemli farklar ortaya çıkmaktadır. Verinin biçimi ve miktarı yanında, hukuki ve ahlaki yönleri de ciddi önem arz etmektedir. Her geçen yıl daha fazla insanın tıbbi verileri elektronik ortamda tutulmaktadır. Günümüzde artık hemen hemen tüm tıbbi cihazların da veri üretmesi ve bunların da depolama birimlerinde tutulması nedeniyle tıbbi veri miktarı muazzam boyutlara ulaşmıştır [2].

Tıbbi verilerin çok olması, çeşitli ahlaki, kanuni ve sosyal sorunlar ve tıbbi özel bazı durumlar bu alanı diğer alanlardan ayırmaktadır [2].

Tıbbi veriler, her çeşit görüntüleme cihazı, özel görüşmeler, doktor gözlem ve girişimleri, yardımcı sağlık personeli gözlem ve girişimleri, laboratuvar verileri gibi çok çeşitli biçimlerde olabilir. Tıbbi verilerin çeşitliliği konusunu şu başlıklar altında incelemek mümkündür [1]:

- Tıbbi verilerin hacmi ve karmaşıklığı
- Doktorun yorumu
- Hassasiyet ve belirlilik analizi
- Zayıf matematik karakterizasyon
- Doğal biçim

Tutulup işlenecek olan verinin gün geçtikçe artması büyük veri üzerine çalışma probleminin bir an önce halledilmesi gerektiğinin altını çizmektedir. Veri madenciliğinin bu amaç için geliştirildiği düşünüldüğünde gelecekte de veri madenciliğinin çok önemli bir yer teşkil edeceği açıktır. Bu amaçla veri madenciliğinin ortaya çıkması aşağıdaki bazı uygulama alanlarını etkinleştirmiştir. Veri madenciliği uygulamaları, bu güne kadar iş dünyasında rekabete dayalı bir yön kazandırmak için kullanılmıştır. Veri madenciliği popüler olmaya başlayınca, biyomedikal, finansal analiz, perakende satış ve telekomünikasyon gibi diğer alanlarda da kullanılmıştır. Genel veri madenciliği uygulamalarının belirli bir alana özel uygulamalar için kullanılamaması gibi bir kısıtlama olduğundan, gelecekte daha çok, alana özgü uygulamalarının çıkması beklenmektedir [2].

Veri tabanı sistemlerinin artan kullanımı ve hacimlerindeki bu olağanüstü artış, organizasyonları elde toplanan bu verilerden nasıl faydalanılabileceği problemi ile karşı karşıya bırakmıştır. Geleneksel sorgu veya raporlama araçlarının veri yığınları karşısında yetersiz kalması, Veri Tabanlarında Bilgi Keşfi-VTBK adı altında, sürekli ve yeni arayışlara neden olmaktadır. VTBK süreci içerisinde, modelin kurulması ve değerlendirilmesi aşamalarından meydana gelen veri madenciliği en önemli kesimi oluşturmaktadır. Bu önem, birçok araştırmacı tarafından VTBK ile “veri madenciliği” terimlerinin eş anlamlı olarak da kullanılmasına neden olmaktadır. Çeşitli veri kaynaklarından verilerin toplanması ile başlayan VTBK süreci, toplanan verilerin analiz için uygun hale getirilmesi aşaması ile devam etmektedir. Veri madenciliğini, verilerden daha önceden bilinmeyen, muhtemelen faydalı enformasyonun monoton olmayan bir süreçte çıkartılması işlemi olarak tanımlamaktadır. Bu süreç kümeleme, veri özetleme, sınıflama kurallarının öğrenilmesi, bağımlılık ağlarının bulunması, değişikliklerin analizi ve anomali tespiti gibi farklı bir çok teknik yaklaşımı kapsamaktadır [3].

Veri madenciliği, istatistik ve matematik tekniklerle birlikte örüntü tanıma teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni korelasyon, örüntü ve eğilimlerin keşfedilmesi sürecidir [3].

Örüntü tanıma ve sınıflama problemleri üzerinde yoğunlaşan yapay zekâ ve istatistik disiplinlerindeki gelişmeler veri madenciliğinin temellerini oluşturmaktadır. Ayrıca veri madenciliği, yapay zeka çalışmalarının uzantısı olan makine öğrenmesi ve uzman sistemlerin yanı sıra, veri tabanları, optimizasyon, görselleştirme, yüksek performanslı paralel işlemciler (Massively Parallel Processing – MPP- ve Symmetric Multiprocessing -SMP-) gibi çeşitli disiplin ve teknolojilerdeki gelişmelerden de etkilenmektedir [3].

Veri madenciliği astronomi, biyoloji, finans, pazarlama, sigorta, tıp ve birçok başka dalda uygulanmaktadır. Son 20 yıldır Amerika Birleşik Devletleri’nde çeşitli veri madenciliği algoritmalarının gizli dinlemeden, vergi kaçakçılıklarının ortaya çıkartılmasına kadar çeşitli

uygulamalarda kullanıldığı bilinmektedir. Bununla birlikte günümüzde veri madenciliği teknikleri özellikle işletmelerde çeşitli alanlarda başarı ile kullanılmaktadır. Bu uygulamaların başlıcaları ilgili alanlara göre aşağıda özetlenmiştir [3]. Günümüzde muhtemelen, biyolojide en önemli veritabanları diziliş veritabanlarıdır. Biyomoleküllerin dizilişleri üzerinde çalışmalar "insulin B-zinciri" dizilişi (30 peptid parçasından oluşur) ile başladı. İlk tam protein dizilişi 1956 yılında 51 kalıntıdan oluşan sığır insülini için bildirildi. Yaklaşık 14 yıl sonra ilk nükleik asit dizilişi 77 tabanlı maya tRNA'sında keşfedildi [4].

İlk nükleik asit dizilişinin keşfinden bir yıl sonra, ilk Biyoinformatik veritabanını oluşturmak üzere tüm yararlanılabilir diziliş verilerini toplandı ve 1966 yılında yayınlandı. Bunu 1972 yılında, 10 X-ışını kristalografik protein yapısından oluşan bir koleksiyonla Protein Veri Bankası izledi ve 15 yıl sonra 1987 yılında SWISSPROT protein diziliş veritabanı projesi başlatıldı. Bugün, bunun üzerinden 25-30 yıl bile geçmeden her yıl on milyonlarca diziliş veritabanı elde edilmekte, belli organizmalara ait dizilişler tümüyle çözülebilmektedir. Diziliş veritabanlarında kesintisiz bir üssel artış gözlenmektedir. EMBL nükleotid veritabanının kapasitesi Mart 1995'te 26.200.000 tabana ulaşmış durumdadır. Günümüzde çok sayıda ve tipte veri kaynağı hem kamu hem de özel sektör tarafından kullanıma sokulmuştur [4].

Orijinal veritabanlarının tümü basit dosyalarda ya da her seferinde bir kayıt veya bir dosyanın tümünü tamamıyla elde edecek şekilde son derece basit bir organizasyon ile sunulmaktadır. Veritabanlarının oluşturulmasından sonra diziliş veritabanlarında ilk aşamada basit anahtar sözcük ve kısa diziliş sözcükleri karşılaştırma esasına dayanan arama yeteneklerine sahip araçlar da eklenmiştir. Bunun için birçok algoritma geliştirilmiştir. Daha yavaş ancak daha özenli olan FASTA yanında, hızlı karşılaştırma algoritmalarınca tümlenen, ancak hızlı olmasına karşın daha az özenli olan BLAST algoritması on yıldır diziliş veritabanı araştırmalarında kullanılan temel algoritmalar olmuştur [4].

Bu öncü gayretler diziliş bilgisi koleksiyonunun otomasyonunda önemli ilerlemelere neden olmuştur. Biyokimya ve enstrümantasyonun da hızla yapılan keşifler bizi, temelde mikrobiyal patojenler olmak üzere, en azından 20 organizmanın genomik dizilişinin tamamını bildiğimiz bir noktaya ulaştırmıştır. En azından 100 veya daha fazla prokaryotik ve ökaryotik genom keşfedilmek üzere çalışma altında bulunmaktadır [4].

Bugün tüm dünya için önemli bir çalışma olan ve sonradan çok uluslu bir proje haline dönüştürülen İnsan Genom Projesi 1987'de başlatılan dev projelerden biridir. Kısmi veya tamamlanmış diziliş şeklinde 46.000 gen şu anda kamuya açık veritabanlarında hizmete sunulmuş durumdadır. İnsan genomunun tamamının muhtemelen 4 milyarın üstünde DNA diziliş tabanından oluşacağı belirtilerek moleküler biyoloji, matematik, istatistik ve bilgisayar bilimlerinde eğitim görmüş yeni bir tip bilim adamına gereksinim olduğu bildirilmektedir.

Büyük bir kısmı tamamlanan Genom projesinin 5 yıl içinde bitirilmiştir. Bunun birinci anlamı, gelecek yüzyılın başında geliştirilecek tüm ilaçların bu veritabanı üzerindeki çalışmalara dayanabilir olmasıdır. İkincisi, yeni teknolojiler ile, hastalık durumunda canlı hücrelerdeki hem mRNA ve hem de proteinlerin tezahür düzeylerindeki değişimleri doğrudan gözlenebilecek ve bir dokudaki etkileşim etki mekanizmasının anlaşılmasını sağlayacak bir yöntem olarak, hücrelerdeki tepki desenlerini tanımaya yardımcı olacaktır [5].

İnsan Genom Projesi sonucu hazırlanacak "insan genom atlası" ise 21. Yüzyıl'da tıp eğitiminde yeni bir döneme girilmesine neden olacaktır. Yeni yüzyılın ilk çeyreğinde kalıtsal hastalıkların tedavisinde önemli adımlar atılacak, hayvanlar ve bitkiler üzerinde yürütülen genom araştırmalarının ise daha yüksek verim için moleküler düzeyde ıslah anlamına geleceği açıktır [5].

Biyolojik bilgiyi yönetmek, analiz etmek ve anlamak için ne yapmak gerekiyor? Yanıt, geçmişte yapılmış ve gelecekte yapılacak çalışmaların çoğunda bilgi işlem araçlarını kullanmak şeklinde olacaktır. İşte bu noktada iki yönlü bir deneyim ve bilgi birikimi gerektiği, biyolojik bilginin elde edilmesinde matematik, istatistik, genetik, biyokimya ve bilgisayar bilimlerini kapsayan disiplinler arası bir bilim dalına gereksinim olduğu açıkça ortaya çıkmaktadır. Gerçekten de biyolojik bilimler sahasında son 20 yıl içinde altı çizilmesi gereken çok önemli nitel ve nicel kökten değişimler/gelişmeler bu bilimler arası işbirliğinin bir ürünü durumundadır [6].

Karmaşık biyolojik sistemlerdeki bağlantılar ve sinyaller arası ilişkiler gün geçtikçe çözülerek, sır olmaktan çıkmaktadır. İster transkripsiyonun regülasyonu olsun, ister hücresel aktivasyon olsun, isterse gelişmenin organizasyonu veya hücresel iletişim olsun; hepsi çok karmaşık sayılabilecek tüm bu olayların işleyişlerinin çözülmesi niceliksel bakımdan insan beyni tarafından pratik şekilde çözülmesi zor ve hatta olanaksız olaylardır. İşte, son birkaç yıl içinde aniden kullanılmaya başlayan Biyoinformatik, yukarıda sorunlara bilgi işlem çözümleri aramak için gelişen ve 21. Yüzyıla özel yeni bir bilim alanıdır [6].

Biyoinformatik;

- Hızla gelişen ve artan bilgiyi toplamak, yönetmek, dağıtmak,
- Bilgiye en hızlı ve kolay biçimde ulaşmak,
- Çok karmaşık yapıda olan biyolojik sistemleri çözebilmek ya da incelemek için çok

güçlü ve karmaşık araçları kullanma gerçeği ile karşı karşıya kalınmaktadır [6].

Araştırmacılar biyolojik verilerin:

- Toplanması,
- Yönetimi ve erişimi ve
- Analizi ve yorumu

hususlarında Biyoinformatik bilimini ciddi şekilde düşünme gerçeği ve gereği gelişmelerin bir zorlaması olarak görülmektedir [6].

Daha fazla miktarda biyolojik olgu açıklandıkça bilgi ile ifade edilmeye çalışılan kavramlar ve bilgiden ne anlaşıldığı hususu dahi değişmek zorunda kalacaktır. Bu yüzden bu tür bilgileri yönetmek ve anlamlı bulguları çıkarmak üzere analiz etmek için güçlü araçlara, bilgisayarlara gereksinim vardır ve zamanla bu gereksinim çok hızlı şekilde artacaktır. Bu nedenle, bilgi toplama, depolama, yönetim ve analizini ve edinimi sağlayabilecek birtakım modeller üzerinde çalışma yapılması gereği, kendiliğinden ortaya çıkmaktadır. Böyle bir model altında, yaşam bilimlerindeki:

- Homoloji ve analogi,
- Filogeni ve yaşamın kökeni,
- Protein yapısı, işlevleri ve sentezi
- Kendi kendini-yönetim gibi sıralanabilecek problemlerinin çözümü ve analizlerinin

yapılmasında, matematik ve bilgisayar bilimlerindeki:

- Veritabanı yönetimi (dağıtık bilgi işlem ve ağ ortamlarında),
- Sinir ağları,
- Birlikte ya da bağlantılı karmaşıklık ve en iyileme,
- Genetik algoritmalar ve
- Evrimsel görüntüler

gibi çözüm araçlarından yararlanılması gerekmektedir [6].

Bu tez, biyokimya test sonuçlarını kullanarak uzman ve pratisyen hekimler için bir karar destek sistemi oluşturup çeşitli hastalıkları teşhis etmeyi amaçlamaktadır. Bu amaçla hazırlanan tez aşağıdaki bölümlerden oluşmuştur:

İkinci bölümde, karar destek sistemi tasarımı için kullanılan veri madenciliği ve biyoinformatik teknikleri incelenmiştir.

Üçüncü bölümde, bilgi çıkarımı yapısı ele alınarak, alt bileşenlerinin işlevleri açıklanmıştır.

Dördüncü bölümde, Kalp Krizi, Hiperlipidemi, Hipotiroidi-Hipertiroidi ve Demir Eksikliği Anemisinde uzman ve pratisyen hekimlere yardımcı olabilecek geliştirilen karar destek sistemleri sunulmuştur.

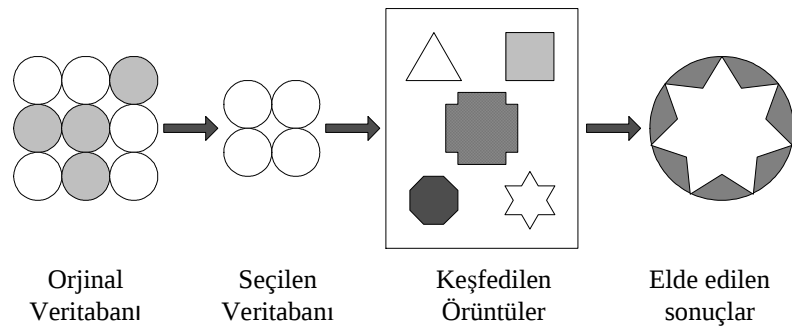
Son bölümde ise, elde edilen sonuçlar irdelenerek katkılar vurgulanmıştır. Ayrıca ileride yapılabilecek benzeri çalışmalar ve uygulama alanları için öneriler tartışılmıştır.

BÖLÜM 2

2. VERİ MADENCİLİĞİ

2.1. Veri Ambarları ve Veri Madenciliği

Veri madenciliği, büyük miktardaki veriler arasından gelecek ile ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilgisayar programı kullanarak aranması sağlayan bilgiye ulaşma işidir. Yani veri işlenerek amaçlanan bilgi elde edilir. Veri madenciliği biyotıp, gen fonksiyonları ve DNA sıralama desenlerinin veri analizlerinde, hastalık tanısında, perakende satış verilerinde, telekomünikasyon endüstrisinde, satış tahmini, finans analizi, astronomi ve birçok alanda uygulanmaktadır [7]. Örneğin bir tıbbi teşhis sisteminde kullanılan standart yazılımlar verinin dağınık olması, kontrol edilememesi, çok farklı çeşitlerinin kullanılması, örnekleri karşılaştırma, kritik farkları tespit etme, veritabanları üzerindeki sistematik ve tutarlı analizler yapılamaması gibi nedenlerle sorunlu olup, doktorlar tarafından eksik görülmüştür. Bu bağlamda hem sayısal parametrelere hem de morfolojilere ilişkin eğrilerin bir uzman bilgi sistemi ile temel analizini sağlayan, gerekli bilgileri depolayan, işleyen, yöneten ve karar için referans gibi son derece yararlı bir yapıyı içeren veri madenciliği devreye girmiştir. Ayrıca veri madenciliği ve teknikleri uzman doktorlara kolaylık, pratisyenlere de yardım sağlaması nedeniyle tıp alanında kabul görmüştür [8]. Veri madenciliği aslında tıp ve diğer alanlarda bilgi keşfi sürecinin bir parçası olarak değerlendirilmektedir [9]. Veri madenciliği aşamaları Şekil 2.1’de gösterilmiştir [10].



Şekil 2.1. Veri madenciliği yöntemi.

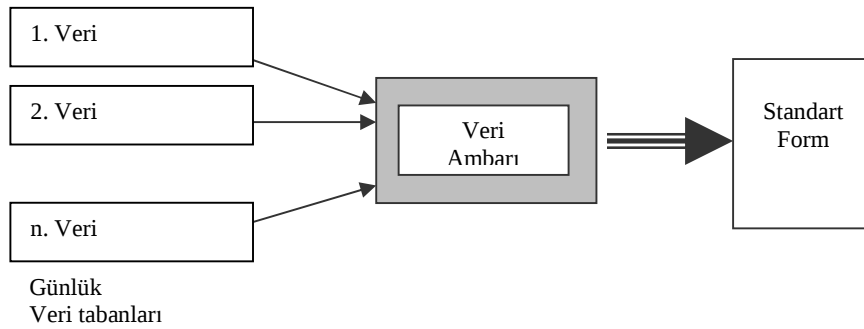
Veri madenciliği, kullanıcı ve bilgi tabanı ile etkileşim halindedir. İlginç veri örüntüleri kullanıcıya gösterilir. Ayrıca, istenirse bilgi tabanına da kaydedilebilir. Buna göre, veri madenciliği işlemi, gizli kalmış veri örüntülerini ve ilişkilerini bulana kadar devam eder [11].

Veri tabanlarındaki veriden anlamlı bilgiler çıkarılırken öncelikle veri yığınları arasından gereken veriler alınır, sınıflandırılır ve daha sonrada işlenir. Sınıflandırma, örüntü tanıma gibi birkaç araştırma alanında çalışan veri madenciliği ve makine öğrenmesinde verilen bir hastanın durumunu kaydetme, tüm hastaların laboratuar test sonuçları, bulgular ve sinyaller gibi verilerinin niteliğini önceden söyleyebilme önemli bir problemidir [12]. Veri madenciliğinde sınıflama metotları, büyük verilerde ilginç nesnelerin otomatik tanımlanmasında ve mali piyasalardaki eğilimin sınıflamalarını içeren uygulamalarda bilgi keşfi için kullanılır [13]. Bu verileri sınıflandırırken kullanılan çeşitli yöntemler mevcuttur. Bu yöntemler içerisinde tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, veri madenciliğinde

- Kuruluşlarının ucuz olması,
- Yorumlanmalarının kolay olması,
- Veri tabanı sistemleri ile kolayca entegre edilebilmeleri,
- Güvenilirliklerinin daha iyi olması

nedenleri ile sınıflama modelleri içerisinde yaygın kullanıma sahiptir [14].

Son zamanlarda bilgi teknolojilerinde dikkati çekecek ölçülerde yaşanan gelişmeler ile depolanan verilerin boyutları çok büyük boyutlara ulaşmaya başlamıştır. Günümüzde yaygın olarak kullanılmaya başlanan veri ambarları günlük kullanılan veri tabanlarının birleştirilmiş ve işlemeye daha uygun bir özetini saklamayı amaçlar. Bu yapı Şekil 2.2.'de gösterilmiştir [15].

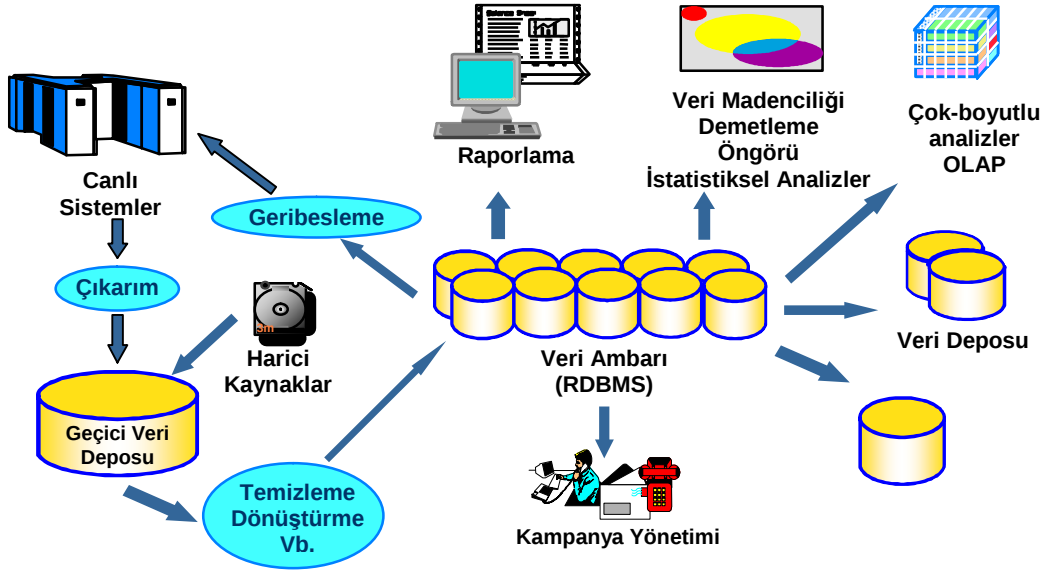


Şekil 2.2. Günlük veri tabanlarındaki verinin veri madenciliği çalışması için standart bir forma çevrilmesi.

Veri ambarında veri oluşturulduktan sonra bu verinin elle veya gözle analizi yapılabilir. Bunun için OLAP programları kullanılır. Bu programlar veriye her boyutu veride bir alana karşılık gelen çok boyutlu bir küp olarak bakmayı ve incelemeyi sağlar. Böylece boyut bazında gruplama, boyutlar arasındaki korelasyonları inceleme ve sonuçları grafik veya rapor olarak sunma olanağı sağlar.

Veri madenciliğinde amaç, kullanıcının bilgi çıkarma sürecinde katkısının olabildiğince az tutulması, işin olabildiğince otomatik olarak yapılabilmesidir. Çünkü OLAP programlarını

kullanırken bulunabilecek sonuçlar kullanıcının sormayı düşündüğü sorgularla sınırlıdır. Ama veri içinde örneklerdeki bağıntı gibi, kullanıcının hiç aklına gelmeyecek bilgiler de olabilir. Zaten veri madenciliğinde esas amaç, bu tip bilgileri bulabilmektir [15].



Şekil 2.3. Veri ambarı yapısı

Şekil 2.3’de yer aldığı gibi veri ambarı üzerinde bir çok işlem yapılabilmektedir. Bunlardan başlıcaları veri madenciliği, çok boyutlu analizler, müşteri ilişkileri yönetimi (MİY), kampanya yönetimi, istatistiksel analizler ve raporlamadır. OLAP ile veri ambarı içerisinde yer alan kayıtlar yöneticiler tarafından istenilen boyutlarda ve biçimlerde raporlar haline getirilebilir ve çok boyutlu analizler yapılabilmektedir [15].

2.1.1. Veri Ambarının Karakteristik Özellikleri

- Çok boyutlu bir modellemede depolama için işlem gören, çoklu kaynaklardan gelen entegre verinin deposudur.
- Zaman serileri ve eğilim analizi gibi tarihsel bilgi gerektiren yapıları destekler.
- İşlemsel veri tabanları gibi uçucu değildir. (Bilgi çok daha az sıklıkta değişir ve periyodik güncelleme yapılır.)
- Karar mekanizmasındaki kişinin daha iyi ve daha hızlı kararlar vermesini sağlayan karar destek teknolojileri topluluğu olarak da tanımlanabilir [16].

2.1.2. Veri Ambarının Ayırt Edici Özellikleri

- Çok boyutlu kavramsal bakış
- Genel boyutsallık
- Sınırsız boyutlar ve toplanma seviyeleri
- Dinamik seyrek matris kontrolü
- İstemci-sunucu mimarisi
- Çoklu kullanıcı desteği
- Erişilebilirlik
- Saydamlık
- Tutarlı raporlama performansı
- Esnek raporlama [16]

2.1.3. Veri Ambarları için Veri Modelleme

Çok Boyutlu Modellemeler: Veriyi “veri küpleri” denilen çok boyutlu bir matrise yerleştirerek, verideki kalıtsal ilişkilerden yararlanılır. Üç boyuttan fazla olanlarına “hiper küpler” denir. Örneğin; ürünleri satırlarla, bölgelerdeki satışları sütunlarla gösterelim. Buna bir de şirketin çeyrek mali dönemlerini gösteren bir zaman boyutu eklersek ortaya 3 boyutlu bir matris çıkar. Sorgulama performansı çok boyutlu matrislerde çok daha iyidir [16].

Bir eksen etrafında döndürme (Pivoting): Çok boyutlu veri küplerinde bir boyuttan diğerine kolayca geçebilmek için kullanılan tekniktir. Bu teknikte veri küpünün eksenlerinin farklı yönlerini gösterecek şekilde döndüğü kabul edilir [16].

Yukarıya Toplanan Görüntü (Roll-up Display): Büyük üniteleri tek bir boyutta toplayarak hiyerarşinin bir üst seviyesine çıkar. Örneğin, haftalık veriyi çeyrek dönemlere ayırıp toplama amaçlanır [16].

Aşağıya Açılan Görüntü (Drill-down Display): Tek bir boyutta toplanmış bilgiyi ünitelere parçalar. Örneğin, ülkedeki satışları bölgelere, bölgelerdeki satışları alt bölgelere ayırarak bilgiler işlenir [16].

Birleştirme İndekslemesi: Bir yıldız şemasında, ana anahtar (primary key) ve dış anahtar (foreign key) değerleri arasındaki ilişkilere bakan tekniktir. Yıldız şemasındaki bir boyutun değerleriyle doğruluk tablosundaki satırlar arasında ilişki kurar. Örneğin; boyutları şehir ve çeyrek mali dönem olan bir satış doğruluk tablosu olsun. Eğer şehir üzerinde bir birleştirme indeksi varsa, birleştirme indeksi her şehir için o şehri içeren sıraların sıra “id”lerine bakar. Birleştirme indeksleri çok boyutlu ortamlarda da çalışır [16].

Bu yöntemler kullanılarak veriler oluşturulduktan sonra bu verilerin Veri ambarlarında depolanması gerekir. Veri ambarı depolaması, veri ambarlarının değişmezliği ve yapılan analizlerin önceden tahmin edilebilirlik derecesi sayesinde özet veriye ulaşmada kolaylık sağlar [16].

Veri ambarı depolamasında 2 temel yaklaşım kullanılır:

- 1) Özet veriyi kapsayan daha küçük tablolar (Çeyreklik satışlar gibi).
- 2) Seviyenin (haftalık, çeyreklik, yıllık gibi) mevcut tablolara kodlanması [16].

2.1.4. Bir Veri Ambarı Oluşturma

Veri ambarının tasarım aşamasında olası tüm sorgulamaları ya da analizleri öngörmek mümkün olmayabilir. Fakat en azından geçici sorgulama desteklenmelidir. Veri ambarı için verinin elde edilmesi şu aşamaları kapsar [16]:

- Veri, birden çok heterojen kaynaktan elde edilmelidir.
- Veri, veri ambarı içerisinde tutarlı olacak şekilde düzenlenmelidir.
- Veri, geçerliliği garantilemek için silinmelidir. Veri silme, veri ambarında en çok iş gücü gerektiren kapsamlı ve karmaşık bir işlemdir. Geri püskürtme (Backflushing) terimi ise silinmiş verinin geri çağırılması demektir.
- Veri, veri ambarının veri modeline uygun olmalıdır.
- Veri, veri ambarına yüklenmelidir. Büyük hacimli veri için artımlı güncelleme en makul olanıdır.

Veri ambarında verinin depolanması işlemi şu süreçleri kapsar:

- Veriyi, veri ambarının veri modeline uygun olarak saklamak.
- Gereken veri yapılarını oluşturmak ve kontrolünü sağlamak.
- Uygun erişim yollarını oluşturmak ve kontrolünü sağlamak.
- Yeni veri eklendikçe, zaman-değişimli veriyi sağlamak.
- Veri ambarındaki verinin güncellenmesini sağlamak.
- Veriyi temizlemek.

Veri ambarı tasarım etmenleri şunları kapsamalıdır:

- Veri ambarını kim, ne şekilde kullanacak vb. kullanım alanları
- Veri modelinin kullanılacağı alanı destekleme uygunluğu
- Eldeki kaynakların temel özellikleri
- “Meta-veri” (metadata) bileşeninin tasarımı (Metadata: şema tanımını da içerecek şekilde veritabanının temel tanımıdır.)

- Modüler bileşen tasarımı (veri ambarının organizasyonla ve bilgi ortamıyla birlikte evrim geçirmesini sağlamak adına pratik bir gereksinimdir)
- Yönetilebilirlik ve değişiklik tasarımı (veri ambarı yöneticilerinin etkin planlama yapabilmeleri ve kullanıcılara en iyi desteği sağlarken diğer yandan değişiklikleri de idare edebilmeleri için gereklidir)
- Dağıtık ve paralel mimari etmenleri.

Veri ambarının tasarımını belirleyen en önemli etmenlerden biri de organizasyonun kullandığı dağıtık mimaridir. İki temel dağıtık mimari çeşidi vardır: Dağıtık veri ambarı ve birleşik veri ambarı [16].

Dağıtık veri ambarı: Dağıtık veritabanlarının tüm önemli noktaları (kopyalama, bölümlere ayırma, haberleşme, tutarlılık gibi) birbirleriyle ilişkilidir. Dağıtık bir mimari, veri ambarının performansında çok önemli rol oynayabilir (gelişmiş yük dengeleme, performansın ölçeklenebilirliği, yüksek erişilebilirlik gibi). Tek kopyadan oluşan bir metadata deposu her iki dağıtım alanında da bulunmalıdır [16].

Birleşik veri ambarı: Merkezden birkaç yere dağıtılmış, özerk veri ambarları birliğidir. Her birinin kendi metadata deposu vardır. Bu tür birleşimler, veri marketleri gibi daha küçük ölçekli elemanlardan oluşur [16].

2.1.5. Veri Ambarlarının Uygulanmasındaki Zorluklar

Veri ambarında önemli işlemsel noktalar şunlardır: Oluşturma, yönetim ve kalite kontrol. Proje yönetimi – veri ambarının tasarımı, oluşturulması ve uygulanması – önemli ve uğraştırıcı bir etmendir. Büyük bir organizasyon için veri ambarının oluşturulmasıyla uygulamaya geçirilmesi arasındaki süre yılları bulabilir. Zorlukları ve harcanan zaman dikkate alındığında, gösterdiği gelişim ve yaygın kullanımı nedeniyle veri marketleri OLAP ve veri madenciliği desteğine acil ihtiyaç duyan organizasyonlar için iyi bir alternatif olabilir [16].

- Veri ambarının yönetimi, büyüklüğüyle ve karmaşıklığıyla orantılı olarak ciddi bir yatırım gerektirir. Veri ambarını yönetmek isteyen bir organizasyon, onun karmaşık doğasını anlayabilmelidir. Çünkü veri ambarı sağladığı bilgi kaynaklarının dışında statik bir yapıdadır. Kaynak veritabanlarının gelişmesi gerekmektedir.
- Veri ambarlamadaki önemli diğer bir nokta da verinin kalite kontrolüdür. Veri bir temizleme fonksiyonundan geçse de, kalite ve tutarlılık gibi noktalar veritabanı yöneticisine kalmaktadır. Veriyi heterojen ve tamamen farklı kaynaklardan almak; isimlendirme, kimlik numaraları gibi konularda sorun yaratabilir. Ne zaman kaynak veritabanında bir değişiklik gündeme gelse, veri ambarı yöneticisi bu değişikliğin ambarın diğer elemanlarıyla olan etkileşimini düşünmek durumundadır.

- Veri ambarı öyle tasarlanmalıdır ki, veri kaynaklarının artması ya da azalması, büyük bir tasarım değişikliğine neden olmamalıdır. Kaynaklar ve kaynak verideki bu değişikliğe veri ambarı yanıt verebilmelidir.
- Veri ambarı yönetimi bir ekip işidir. Veri ambarı yönetim fonksiyonunun tasarımı ve yönetim ekibinin seçimi çok önemlidir. Ekipteki elemanlar; teknik yetenekleri olan, işbirliği içinde çalışabilen ve etkin liderlik vasfı olan kişiler arasından seçilmelidir.

2.2. Bir Veri Madenciliği Çalışması Örneği

Bir finans kurumunda insanlara kredi verilebileceği durumları tasarlayan bir sistemde amaç; geçmişte kredi verilen insanların bazıları kredilerini normal olarak geri ödemiş, bazılarıysa ödeyememiş ve kanuni takibe düşmüş olduğu düşünülürse bu kişilerin özelliklerine bakarak ödemelerini düzenli yapan müşterileri, kanuni takibe düşmüş müşterilerden ayırabilmektir. Bu uygulama için standart formda veri her kolonunun bir alan, her satırının da bir kişi ile ilgili bilgilerin olduğu bir tablodur (Tablo 2.1). Amaç, alanlardan birinin değerini ilgili diğer alan bilgileri cinsinden açıklayan bir kural bulabilmektir. Veri analizini karmaşık yapan satır (kayıt) sayısından çok, sütun sayısı, yani boyut sayısı veya alan sayısıdır [17].

Tablo 2.1 Örnek için standart form.

Adı	Yıllık gelir	Birikim	Ev sahibi	Medeni hali	Kredi durumu
Ali	15,000	30,000	Evet	Evli	Normal ödeme
Veli	10,000	20,000	Evet	Evli	Kanuni takip
...					

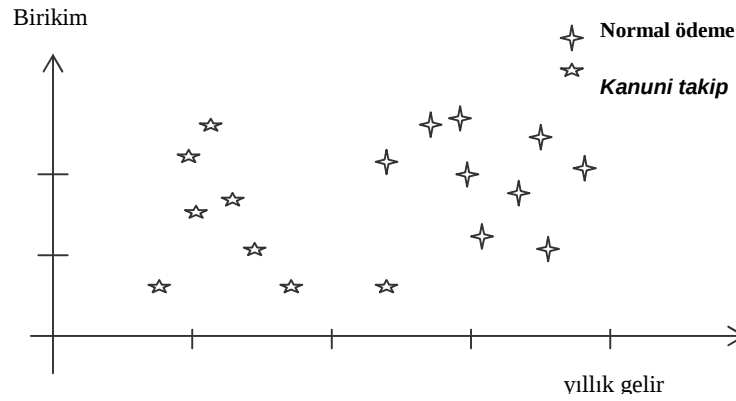
Veriyi görselleştirebilmek için sadece iki alan olduğunu varsayılırsa, bunlar yıllık gelir ve birikim olsun. Örnek veri bu iki boyutlu uzayda noktalar olarak çizilebilir. Amaç; bu iki boyutlu uzayda kredisini normal ödemiş ve kanuni takibe düşmüş iki grupta kişilere karşılık gelen noktaları birbirinden ayıran bir sınır bulabilmektir. Bu bir sınıflandırma problemidir. Eğer böyle bir sınır bulabilirsek yeni bir kredi başvurusunda bu kişinin birikim ve yıllık gelirin bakarak iki boyutlu uzayda sınırın hangi tarafında kaldığına bakarak, hangi gruba ait olduğuna dair karar verilebilir. Sınıflandırma için kullanılabilecek birçok teknik vardır. Bunlardan bir tanesi de karar ağaçlarıdır. Bu verinin iki boyutlu uzayda dağılımı Şekil 2.4'te, karar ağacı yapısı Şekil 2.6'da gösterilmiştir. Bu karar ağacı şu kurala karşılık gelir:

EĞER yıllık gelir > Q₁ (VE) birikim > Q₂

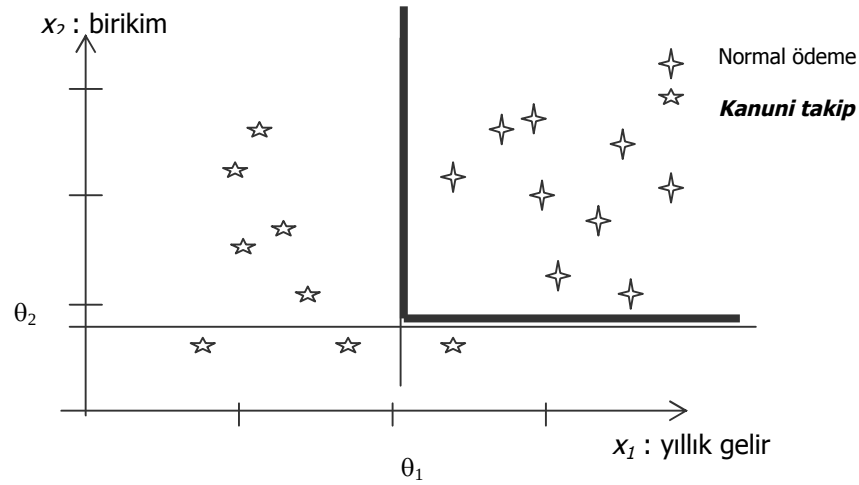
İSE Normal Ödeme

DEĞİLSE Kanuni Takip

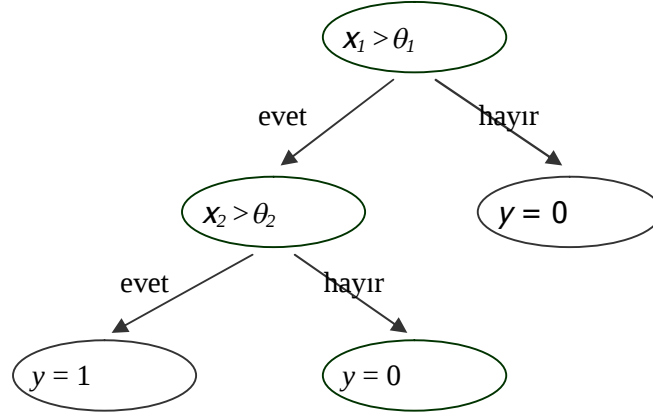
Karar ağaçlarının en büyük yararı veriden öğrenilen kuralın anlaşılır bir şekilde yazılabilesidir. Q_1 ve Q_2 iki boyuttaki eşik değerleridir. Karar ağacının yapısı ve bu eşik değerleri karar ağacı öğrenme algoritması tarafından veriden otomatik hesaplanır. Kullanılan iki alanın tanımladığı iki boyutlu uzayda standart formdaki her kişi ile ilgili bilgi bir noktaya karşılık gelir. Sınıflandırmada amaç bu uzayda iki sınıfa ait (normal ödenmiş ve kanuni takibe gitmiş) noktaları birbirinden ayıran bir sınır bulabilmektir [17].



Şekil 2.4. Veri örneği.



Şekil 2.5. Karar ağacı tarafından tanımlanan sınır.



Şekil 2.6. Tanımlanan sistemin karar ağacının yapısı.

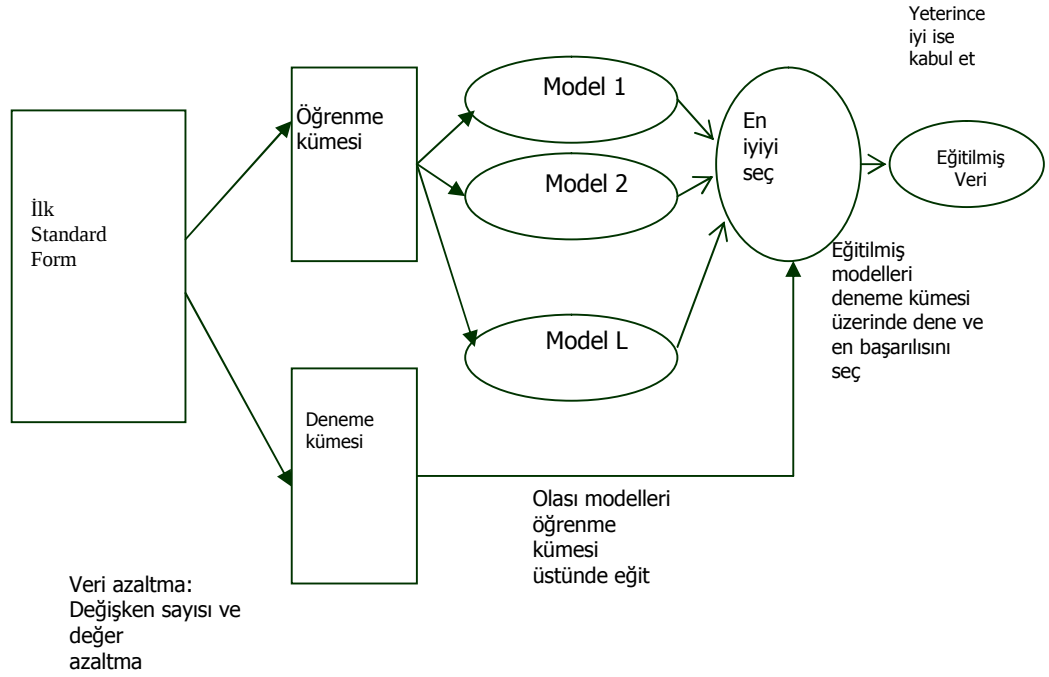
Burada x_1 :yıllık gelir, x_2 :birikim, $y=0$:Kanuni takip, $y = 1$:Normal ödeme olarak tanımlanmıştır [17].

2.3. Veri Madenciliğinde Metodoloji

Bir veri madenciliği çalışmasında kullanılan metodoloji Şekil 2.7.'de verilmiştir. Standart form içinde verilen veri, öğrenme ve deneme olmak üzere ikiye ayrılır.

Her uygulamada kullanılabilecek birden çok teknik vardır ve önceden hangisinin en başarılı olacağını kestirmek olası değildir. Bu yüzden öğrenme kümesi üzerinde L değişik teknik kullanılarak L tane model oluşturulur. Sonra bu L model deneme kümesi üzerinde denenerek en başarılı olanı, yani deneme kümesi üzerindeki tahmin başarısı en yüksek olanı seçilir.

Eğer bu en iyi model yeterince başarılıysa kullanılır, aksi takdirde başa dönerek çalışma tekrarlanır. Tekrar sırasında başarısız olan örnekler incelenerek bunlar üzerindeki başarının nasıl artırılabilceği araştırılır. Örneğin standart forma yeni alanlar ekleyerek programa verilen bilgi artırılabilir veya eldeki mevcut bilgi değişik bir şekilde kodlanabilir veya amaç daha değişik bir şekilde tanımlanabilir [17].



Şekil 2.7. Veri madenciliği çalışmasında kullanılan metodoloji.

2.4. Kullanılan Teknikler

Veri madenciliğinde kullanılan modeller, tahmin edici ve tanımlayıcı olmak üzere iki ana başlık altında incelenmektedir. Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçları bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır. Örneğin, bir banka önceki dönemlerde vermiş olduğu kredilere ilişkin gerekli tüm verilere sahip olabilir. Bu verilerde bağımsız değişkenler kredi alan müşterinin özellikleri, bağımlı değişken değeri ise kredinin geri ödenip ödenmediğidir. Bu verilere uygun olarak kurulan model, daha sonraki kredi taleplerinde müşteri özelliklerine göre verilecek olan kredinin geri ödenip ödenmeyeceğinin tahmininde kullanılmaktadır.

Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile, çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir [17].

Veri madenciliği modellerini gördükleri işlevlere göre,

- Sınıflama ve Regresyon
- Kümeleme
- Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler,

olmak üzere üç ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme, birliktelik kuralları ve ardışık zamanlı örüntü modelleri tanımlayıcı modellerdir [17].

İstatistiksel Yöntemler: Veri madenciliği çalışması esas olarak bir istatistik uygulamasıdır. Verilen bir örnek kümesine bir kestirici oturtmayı amaçlar. İstatistik literatüründe son elli yılda bu amaç için değişik teknikler önerilmiştir. Bu teknikler istatistik literatüründe çok boyutlu analiz başlığı altında toplanır ve genelde verinin parametrik bir modelden (çoğunlukla çokboyutlu bir Gauss dağılımından) geldiğini varsayar. Bu varsayım altında sınıflandırma, regresyon, öbekleme, boyut azaltma, hipotez testi, varyans analizi, bağıntı kurma için teknikler istatistikte uzun yıllardır kullanılmaktadır [17].

Bellek Tabanlı Yöntemler: Bellek tabanlı veya örnek tabanlı bu yöntemler istatistikte 1950’li yıllarda önerilmiş olmasına rağmen o yıllarda gerektirdiği hesaplama ve bellek yüzünden kullanılamamış ama günümüzde bilgisayarların ucuzlaması ve kapasitelerinin artmasıyla, özellikle de çok işlemcili sistemlerin yaygınlaşmasıyla, kullanılabilir olmuştur. Bu yönüme en iyi örnek en yakın k komşu algoritmasıdır [17].

Yapay Sinir Ağları: 1980’lerden sonra yaygınlaşan yapay sinir ağlarında amaç fonksiyon birbirine bağlı basit işlemci ünitelerinden oluşan bir ağ üzerine dağıtılmıştır. Yapay sinir ağlarında kullanılan öğrenme algoritmaları veriden, üniteler arasındaki bağlantı ağırlıklarını hesaplar. YSA istatistiksel yöntemler gibi veri hakkında parametrik bir model varsaymaz yani uygulama alanı daha geniştir ve bellek tabanlı yöntemler kadar yüksek işlem ve bellek gerektirmez [17].

Karar Ağaçları: İstatistiksel yöntemlerde veya yapay sinir ağlarında veriden bir fonksiyon öğrenildikten sonra bu fonksiyonun insanlar tarafından anlaşılabilir bir kural olarak yorumlanması zordur. Karar ağaçları ise veriler oluşturulduktan sonra yukarıdaki örnekte de olduğu gibi ağaç kökten yaprağa doğru inilerek kurallar yazılabilir. Bu şekilde kural çıkarma, veri madenciliği çalışmasının sonucunun geçerlenmesini sağlar. Bu kurallar uygulama konusunda uzman bir kişiye gösterilerek sonucun anlamlı olup olmadığı denetlenebilir. Sonradan başka bir teknik kullanılsa bile karar ağacı ile önce bir kısa çalışma yapmak, önemli değişkenler ve yaklaşık kurallar konusunda bilgi verdiğinden tavsiye edilir [17].

2.5. Sınıflama ve Regresyon Modelleri

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanan 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 gibi kategorik değerlerin de tahmin edilmesine olanak sağlayan tekniklerle, her iki model giderek birbirine yaklaşmakta ve bunun bir sonucu olarak aynı tekniklerden yararlanılması mümkün olmaktadır. Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler [17],

- Karar Ağaçları,
- Yapay Sinir Ağları,
- Genetik Algoritmalar,
- K-En Yakın Komşu,
- Bellek Temelli Nedenleme
- Lojistik Regresyondur.

2.6. Kümeleme Modelleri

Kümeleme modellerinde amaç, küme üyelerinin birbirlerine çok benzediği, ancak özellikleri birbirlerinden çok farklı olan kümelerin bulunması ve veri tabanındaki kayıtların bu farklı kümelerle bölünmesidir. Başlangıç aşamasında veri tabanındaki kayıtların hangi kümelerle ayrılacağı veya kümelemenin hangi değişken özelliklerine göre yapılacağı bilinmemekte, konunun uzmanı olan bir kişi tarafından kümelerin neler olacağı tahmin edilmektedir [17].

2.7. Birliktelik Kuralları ve Ardışık Zamanlı Örüntüler

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

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

- Müşteriler kola satın aldığı anda, % 75 ihtimalle patates cipsi de alırlar,
- Düşük yağlı peynir ve yağsız yoğurt alan müşteriler, %85 ihtimalle diet süt de satın alırlar.

Ardışık zamanlı örüntüler ise aşağıda sunulan örneklerde görüldüğü gibi birbirleri ile ilişkisi olan, ancak birbirini izleyen dönemlerde gerçekleşen ilişkilerin tanımlanmasında kullanılır [17].

- X ameliyatı yapıldığında, 15 gün içinde % 45 ihtimalle Y enfeksiyonu oluşacaktır,
- İMKB endeksi düşerken A hisse senedinin değeri % 15'den daha fazla artacak olursa, üç iş günü içerisinde B hisse senedinin değeri % 60 ihtimalle artacaktır.

2.8. Veri Tabanlarında Bilgi Keşfi Süreci

Ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda fayda sağlaması mümkün değildir. Bu nedenle aşağıda tanımlanan tüm aşamalardan önce, iş ve veri özelliklerinin öğrenilmesi / anlaşılması başarının ilk şartı olacaktır [18].

- Problemin Tanımlanması,
- Verilerin Hazırlanması,
- Modelin Kurulması ve Değerlendirilmesi,
- Modelin Kullanılması,
- Modelin İzlenmesi veri tabanlarında bilgi keşfi sürecinde izlenmesi gereken temel aşamalardır.

2.8.1. Problemin Tanımlanması

Veri madenciliği çalışmalarında başarılı olmanın ilk şartı, uygulamanın hangi işletme amacı için yapılacağının açık bir şekilde tanımlanmasıdır. İlgili işletme amacı işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır. Ayrıca yanlış tahminlerde katlanılacak olan maliyetlere ve doğru tahminlerde kazanılacak faydalara ilişkin tahminlere de bu aşamada yer verilmelidir [18].

2.8.2. Verilerin Hazırlanması

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

hazırlanması aşaması kendi içerisinde toplama, değer biçme, birleştirme ve temizleme, seçme ve dönüştürme adımlarından meydana gelmektedir [18].

Toplama: Tanımlanan problem için gerekli olduğu düşünülen verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adımıdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı, hava durumu, merkez bankası kara listesi gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından faydalanılabilir [18].

Değer Biçme: Veri madenciliğinde kullanılacak verilerin farklı kaynaklardan toplanması, doğal olarak veri uyumsuzluklarına neden olacaktır. Bu uyumsuzlukların başlıcaları farklı zamanlara ait olmaları, kodlama farklılıkları (örneğin bir veri tabanında cinsiyet özelliğinin e/k, diğer bir veri tabanında 0/1 olarak kodlanması), farklı ölçü birimleridir. Bu nedenlerle, iyi sonuç alınacak modeller ancak iyi verilerin üzerine kurulabileceği için, toplanan verilerin ne ölçüde uyumlu oldukları bu adımda incelenerek değerlendirilmelidir [18].

Birleştirme ve Temizleme: Bu adımda farklı kaynaklardan toplanan verilerde bulunan ve bir önceki adımda belirlenen sorunlar mümkün olduğu ölçüde giderilerek veriler tek bir veri tabanında toplanır. Ancak basit yöntemlerle ve baştan savma olarak yapılacak sorun giderme işlemlerinin, ileriki aşamalarda daha büyük sorunların kaynağı olacağı unutulmamalıdır [18].

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

Sıra numarası, kimlik numarası gibi anlamlı olmayan ve diğer değişkenlerin modeldeki ağırlığının azalmasına da neden olabilecek değişkenlerin modele girmemesi gerekmektedir. Bazı veri madenciliği algoritmaları konu ile ilgisi olmayan bu tip değişkenleri otomatik olarak eelse de, pratikte bu işlemin kullanılan yazılıma bırakılmaması daha akılcı olacaktır.

Verilerin görselleştirilmesine olanak sağlayan grafik araçlar ve bunların sunduğu ilişkiler, bağımsız değişkenlerin seçilmesinde önemli yararlar sağlayabilir.

Genellikle yanlış veri girişinden veya bir kereye özgü bir olayın gerçekleşmesinden kaynaklanan verilerin, önemli bir uyarıcı enformasyon içerip içermediği kontrol edildikten sonra veri kümesinden atılması tercih edilir [18].

Dönüştürme: Kredi riskinin tahmini için geliştirilen bir modelde, borç/gelir gibi önceden hesaplanmış bir oran yerine, ayrı ayrı borç ve gelir verilerinin kullanılması tercih edilebilir. Ayrıca modelde kullanılan algoritma, verilerin gösteriminde önemli rol oynayacaktır. Örneğin bir uygulamada bir yapay sinir ağı algoritmasının kullanılması durumunda kategorik değişken değerlerinin evet/hayır olması; bir karar ağacı algoritmasının kullanılması durumunda ise örneğin gelir değişken değerlerinin yüksek/orta/düşük olarak gruplanmış olması modelin etkinliğini artıracaktır [18].

2.8.3. Modelin Kurulması ve Değerlendirilmesi

Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılncaya kadar yinelenen bir süreçtir. Model kuruluş süreci denetimli ve denetimsiz öğrenimin kullanıldığı modellere göre farklılık göstermektedir. Örnekten öğrenme olarak da isimlendirilen denetimli öğrenimde, bir denetçi tarafından ilgili sınıflar önceden belirlenen bir kritere göre ayrılarak, her sınıf için çeşitli örnekler verilir. Sistemin amacı verilen örneklerden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir. Bu kurallara göre hazırlanan örnek tabloda borç, gelir ve evli/bekar durumlarına göre kişilerin kredi durumları Tablo 2.2’de gösterilmiştir [18].

Tablo 2.2. Modelin kurulması ve değerlendirilmesi için bir örnek tablo

Adı	Borç	Gelir	Evli ?	Risk
A	Yüksek	Yüksek	Evet	İyi
B	Düşük	Yüksek	Evet	İyi
C	Düşük	Yüksek	Hayır	Kötü
D	Yüksek	Düşük	Hayır	Kötü
E	Düşük	Düşük	Evet	Kötü
F	Yüksek	Düşük	Hayır	İyi
G	Yüksek	Yüksek	Evet	İyi
H	Düşük	Düşük	Evet	Kötü

Adı	Borç	Gelir	Evli ?	Risk
A	Yüksek	Yüksek	Hayır	?
B	Düşük	Düşük	Hayır	?
C	Yüksek	Düşük	Evet	?

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

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

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

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik testidir. Bu yöntemde tipik olarak verilerin % 5 ile % 33 arasındaki bir kısmı test verileri olarak

ayrılır ve kalan kısım üzerinde modelin öğrenimi gerçekleştirildikten sonra, bu veriler üzerinde test işlemi yapılır. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile doğruluk oranı hesaplanır [18].

$$\text{Doğruluk Oranı} = 1 - \text{Hata Oranı}$$

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

Model kuruluşu çalışmalarının sonucuna bağlı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiği değişik modeller kurulabilir. Model kuruluş çalışmalarına başlamadan önce, imkânsız olmasa da hangi tekniğin en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere sayısız deneme yapılmasında yarar bulunmaktadır.

Özellikle sınıflama problemleri için kurulan modellerin doğruluk derecelerinin değerlendirilmesinde basit ancak faydalı bir araç olan risk matrisi kullanılmaktadır. Aşağıda bir örneği görülen bu matriste sütunlarda fiili, satırlarda ise tahmini sınıflama değerleri yer almaktadır. Örneğin fiilen B sınıfına ait olması gereken 46 elemanın, kurulan model tarafından 2’sinin A, 38’inin B, 6’sının ise C olarak sınıflandırıldığı matriste kolayca görülebilmektedir [18].

Tablo 2.3. Sınıflama problemleri için örnek tablo.

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

Önemli diğer bir değerlendirme kriteri modelin anlaşılabilirliğidir. Bazı uygulamalarda doğruluk oranlarındaki küçük artışlar çok önemli olsa da, birçok işletme uygulamasında ilgili kararın niçin verildiğinin yorumlanabilmesi çok daha büyük önem taşıyabilir. Çok ender olarak

yorumlanamayacak kadar karmaşıksalar da, genel olarak karar ağacı ve kural temelli sistemler model tahmininin altında yatan nedenleri çok iyi ortaya koyabilmektedir.

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

Kurulan modelin doğruluk derecesi ne denli yüksek olursa olsun, gerçek dünyayı tam anlamı ile modellediğini garanti edebilmek mümkün değildir. Yapılan testler sonucunda geçerli bir modelin doğru olmamasındaki başlıca nedenler, model kuruluşunda kabul edilen varsayımlar ve modelde kullanılan verilerin doğru olmamasıdır. Örneğin modelin kurulması sırasında varsayılan enflasyon oranının zaman içerisinde değişmesi, bireyin satın alma davranışını belirgin olarak etkileyecektir [18].

2.8.4. Modelin Kullanılması ve İzlenmesi

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, kredi değerlendirme, dolandırıcılık tespiti gibi işletme uygulamalarında doğrudan kullanılabileceği gibi, promosyon planlaması simülasyonuna entegre edilebilir veya tahmin edilen envanter düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine gömülebilir [18].

Modelin İzlenmesi: Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve gerekiyorsa yeniden düzenlenmesini gerektirecektir. Tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler model sonuçlarının izlenmesinde kullanılan yararlı bir yöntemdir [18].

2.8.5. Karar Ağaçları

Karar ağacı, ağaç yapısında olan bir akış şeması şeklindedir. Her düğüm, bir nitelik üzerindeki test işlemini temsil eder. Her dallanma test işleminin sonucunu temsil eder ve sonuç olarak ağaç sınıflar ile son bulur. Karar ağaçları kolayca sınıflandırma kurallarına dönüştürülebilirler. Karar ağaçları, üç bölümden oluşan bir modeldir:

1. Tanımdaki gibi bir karar ağacı
2. Ağacı oluşturacak bir algoritma
3. Ağacı veriye uygulayacak ve söz konusu sorunu çözecek bir algoritma

Karar ağaçları, eğitici örnekteki veriyi sınavan bir algoritma aracılığıyla gerçekleştirilir ya da alanın bir uzmanı tarafından oluşturulur. Karar ağacı tekniklerinin çoğu, birbirlerinden ağacın nasıl oluşturulduğuyla ayrılır [3].

Karar ağaçlarında, veriler oluşturulduktan sonra ağaca kökten yaprağa doğru inilerek kurallar (*IF-THEN rules*) yazılabilir. Bu şekilde kural çıkarma, veri madenciliği çalışmasının sonucunun geçerli olmasını sağlar. Bu kurallar uygulama konusunda uzman bir kişiye gösterilerek sonucun anlamlı olup olmadığı denetlenebilir [18]. Karar ağaçlarının kullanıldığı analizler genellikle, çeşitli vakaların yüksek, orta, düşük risk grupları gibi çeşitli kategorilere ayrılması, gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulmasında, belirli alt gruplara özgü olan ilişkilerin tanımlanması, kategorilerin birleştirilmesi, tıbbi gözlem verilerinden yararlanarak en etkin kararların verilmesi gibi uygulamalarda kullanılmaktadır [19,20]. Bu yapı Tablo 2.4.'de verilmiştir [21].

Tablo 2.4. Kural Çıkarması.

If (Kural1) then (Karar1)
Else if (Kural2) then (Karar2)
.
Kural= <a1> and ... and <an>
Karar= Ψ , $\Psi \in [\Omega_1, \dots, \Omega_m]$

Burada kurallara göre alınacak karar sınıfı elemanlarından her biri Ω ile bu sınıf elemanlarından varılacak sonuçlar ise Ψ ile temsil edilmektedir. Böyle bir karar listesinin hazırlanması herhangi bir tıp uygulamasında önemli bir adımı oluşturur ve bu kurallara göre kolayca bir karar ağacı çizilebilir [22]. Bu şekilde hazırlanan kurallara göre bir karar ağacı algoritması, bilgi kazanımı anlamında bir karar ağacını oluşturur [23].

Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları, veri madenciliğinde

- Kuruluşlarının ucuz olması,
- Yorumlanmalarının kolay olması,
- Veri tabanı sistemleri ile kolayca entegre edilebilmeleri,
- Güvenilirliklerinin daha iyi olması

nedenleri ile sınıflama modelleri içerisinde en yaygın kullanıma sahiptir.

Karar ağacı temelli analizlerin yaygın olarak kullanıldığı sahalar,

- Belirli bir sınıfın muhtemel üyesi olacak elemanların belirlenmesi,
- Çeşitli vakaların yüksek, orta, düşük risk grupları gibi çeşitli kategorilere ayrılması,
- Gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması,

- Parametrik modellerin kurulmasında kullanılmak üzere çok miktardaki değişken ve veri kümesinden faydalı olacakların seçilmesi,
 - Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması,
 - Kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikliye dönüştürülmesidir.
- Karar ağacı temelli tipik uygulamalar ise:
- Hangi demografik grupların mektupla yapılan pazarlama uygulamalarında yüksek cevaplama oranına sahip olduğunun belirlenmesi,
 - Bireylerin kredi geçmişlerini kullanarak kredi kararlarının verilmesi,
 - Geçmişte işletmeye en faydalı olan bireylerin özelliklerini kullanarak işe alma süreçlerinin belirlenmesi,
 - Tıbbi gözlem verilerinden yararlanarak en etkin kararların verilmesi,
 - Hangi değişkenlerin satışları etkilediğinin belirlenmesi,
 - Üretim verilerini inceleyerek ürün hatalarına yol açan değişkenlerin belirlenmesidir.

Gerçek dünyanın sosyal ve ekonomik olaylarını daha güvenilir bir şekilde gösterebilmek için standart istatistik tekniklerin dışında yeni analiz tekniklerinin geliştirilmesi ile ilgilenen Morgan ve Sonquist tarafından University of Michigan’da 1970’li yılların başlarında kullanıma alınan **Automatic Interaction Detector – AID** (Otomatik etkileşimli belirleyici) karar ağacı temelli ilk algoritma ve yazılımdır. AID tekniği en kuvvetli ve en iyi tahmini gerçekleştirebilmek için bağımlı ve bağımsız değişkenler arasındaki mümkün bütün ilişkilerin incelenmesine dayanmaktadır. En kuvvetli ilişkiye sahip bağımsız değişken bulunduğu, veri kümesi bu bağımsız değişken değerlerine göre ikiye ayrılmakta ve süreç mümkün bölünmeler tamamlanıncaya kadar devam etmektedir. Karar ağacı tekniğinin sağladığı kuruluş ve yorumlama kolaylıkları, AID yazılımının başlangıçta istatistikçi ve veri analistleri tarafından büyük coşku ile karşılanmasına neden olmuştur.

Ancak AID’in bağımlı ve bağımsız değişkenler arasındaki ilişkilerin tanımlanmasında aşırı saldırgan davrandığı ve bunun sonucunda anlamlı ve anlamsız ilişkileri ayırt edemediği yönünde Einhorn başta olmak üzere bir çok araştırmacı tarafından yayınlar yapılmıştır.

İlk temelleri AID yöntemi ile atılan karar ağacı modelleri, çeşitli algoritmalar ile sürdürülmüştür. Geliştirilen bu algoritmalar içerisinde CHAID, C&RT, ID3, Exhaustive CHAID, C4.5, MARS, QUEST, C5.0, SLIQ, SPRINT başlıcalarıdır [3].

2.9. Veri Madenciliği Sistemleri Üzerine Yapılan Çalışmalar

Veri madenciliği tekniklerinin birçok alanda gerekli olan bilgiye erişmek için uygulanabilir olması veri madenciliği teknikleriyle hem genel hem de özel amaçlı birçok uygulamanın geliştirilmesi sağlanmıştır [24].

A. Özel Amaçlı Sistemler: Veri madenciliği algoritmalarının belirli problem çözümleri için kullanılmasıdır. Bu uygulamaların çıkış amacı Veri madenciliği'nin kullanıcıdan bağımsız bir şekilde çalıştırılarak kullanıcının istediği bilgilerin keşfedilmesi ve/veya keşfedilen bilgilerin gömülü bir uygulama içinde doğrudan karar alınmasında faydalanılmasını sağlamaktır.

Veri madenciliği algoritmalarının özel amaçlı uygulandığı yerlerden ilk göze çarpanlar: astronomi, işletmelerdeki satış analizleri, pazarlama, borsa, sigorta vb. alanlardır [24].

B. Genel Amaçlı Sistemler: Bu tür sistemlerde amaçlanan veri madenciliği sorgularının problemten bağımsız olarak tanımlanması ve bu özelliğinden dolayı istenen problemde bu sorguların kullanılabilmesidir [24].

2.9.1. Analysis Manager

Analysis Manager Microsoft firmasının veri madenciliği için üretmiş olduğu üründür. Kümeleme analizi ve karar ağaçları için hazırlanmıştır. Analysis Manager OLAP (çevrim içi analitik işlem) küp desteği sunmaktadır. Analysis Manager'ın güçlü olduğu taraf kullanıcı-dostu bir ara yüze sahip olması ve uygulama kolaylığıdır. Aracın SQL sonucu ile bütünleşik çalışabilmesi bu aracı etkin hale getirmektedir. Analysis Manager'ın bir veri madenciliği sorgusu için farklı algoritmaları desteklememesi en büyük eksikliğidir. Kaynak kodun açık olmaması uygulama geliştiriciler için büyük zorluklar oluşturmaktadır. Kaynak kod yerine, Microsoft kümeleme ve karar ağacı için COM desteği sunsa da bu destek bir çok gömülü sistem uygulamalarında geliştiriciler için eksik bir hizmet olarak görülmektedir. Analysis Manager, üretilen sonuçları farklı birçok gösterim şekliyle kullanıcıya sunulabilmektedir. Mesela karar ağaçları için, karar ağacını gösterebildiği gibi sonuçları kural tabloları şeklinde yorumlama imkânı vermektedir [24].

2.9.2. Darwin

Darwin Oracle firmasının veri madenciliği aracıdır. Darwin regresyon ağaçları, karar ağaçları, kümeleme, yapay sinir ağları, Bayesian öğrenme, k-yakınlığında komşuluk gibi birçok algoritmayı destekleyen bir veri madenciliği aracıdır. Paralel sunucular için geliştirilmiş bir veri

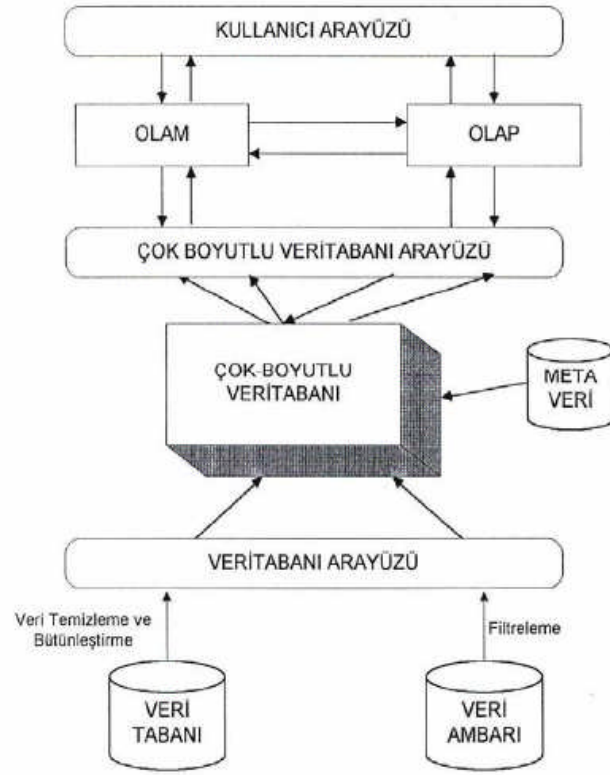
madenciliği sistemidir. Darwin kullanımı kolay bir ara yüze sahiptir. Darwin veri madenciliği algoritmalarından CART, StarTree, StarNet ve StarMatch'i kullanır [24].

2.9.3. Clementine

Clementine SPSS firmasının veri madenciliği için geliştirmiş olduğu bir modüldür. SPSS istatistiksel bir araçtır. Clementine'nin SPSS içinde bir modül olarak kullanılması kullanıcıların SPSS'in istatistiksel fonksiyonlarından faydalanmasına imkan verir. Yapay sinir ağları ve kural tümevarım yöntemlerini kullanır. Clementine müşteri hizmetleri yönetimi, kimya sektöründe maddelerin aşındırıcılık tahmininde ve bankacılık alanında kredi kartı dolandırıcılıkları gibi konularda kendine uygulama alanı bulmuştur [24].

2.9.4. DBMiner

Kanada Simon Fraser Üniversitesi tarafından geliştirilen bir sistemdir. DBMiner sınıflama, kümeleme, eşleştirme ve sıra örüntüleri sorgularını yapabilecek veri madenciliği algoritmalarını kullanır. DBMiner çevrimiçi analitik işleme özelliğiyle veri madenciliği algoritmalarının bütünleşik çalışabilme özelliği sayesinde ön plana çıkmaktadır. Bu özellik OLAM (*Online Analytical Mining*) olarak anılır. DBMiner OLAP ve veri madenciliği yöntemlerini dinamik bir şekilde seçebilme imkânına sahiptir. Kullanıcının kolay kullanabileceği bir ara yüze sahiptir. Bu ara yüz sayesinde elde edilen sonuçlar çok yönlü bir soyutlama kullanılarak gösterilebilmektedir. DBMiner sisteminin mimarisi Şekil 2.8'de verilmiştir [24].



Şekil 2.8 DBMiner sisteminin yazılım mimarisi

Şekil 2.8’de de görüldüğü üzere DBMiner verilerini ilişkisel veri tabanından ve/veya veri ambarından alarak veri küpleriyle bütünleştirerek çok boyutlu veri tabanına aktarır. Bu aktarım kaynaktan, ya verilerin bir bütün olarak çekilmesiyle ya da belli bir bölümünün çekilmesiyle gerçekleşir.

DBMiner’ın diğer sistemlere göre en büyük avantajı geliştirilen DMQL’i (*data mining query language*) kullanmasıdır. DMQL, SQL benzeri bir veri madenciliği sorgu dilidir. DMQL sayesinde çevrimiçi sorgular OLAM veya OLAP modülüne yönlendirilerek işlenir.

DBMiner’ın Veri tabanı ara yüzü çok boyutlu veri tabanına temizlenmiş, filtrelenmiş ve bütünleştirilmiş verileri aktarmaya yarar. Veri aktarımı için ODBC ve OLE DB(*Object Linking and Embedding Database*) gibi bağlantılar da kullanılabilir.

OLAP ve OLAM modülleri arasındaki ilişkinin varlığı iki modülün birbirlerinin sonuçlarını kullanılabilmesine imkân tanır.

DBMiner ürettiği sonuçları farklı birçok şekilde gösterebilme imkânına sahiptir. Mesela karar ağaçları için, karar ağacı şeklinde, kural tabloları şeklinde eşleştirme sorgusu için kural tablosu ve grafikleri üretebilmektedir.

DBMiner ne kadar genel amaçlı bir sistemse de DBMiner'ı kullanarak ortaya çıkarılan özel amaçlı sistemler de mevcuttur. Bunlar arasında MultiMediaMiner, GeoMiner ve WeblogMiner sayılabilir [24].

2.9.5. Data Logic/R

DataLogic/R kümeleme ve sınıflama analizi için kullanılan ticari bir veri madenciliği aracıdır. DataLogic/R artık nitelik ve verilerin temizlenmesi işlemlerini yapabilmektedir. Sistemin en güçlü olduğu taraf, üretilen kuralların öğrenme-test geçerliliği ve güvenlik gibi kriterlerde değerler üretmesidir. Bu değerler üretilen kuralların kalitesini belirlemek için kullanılabilir. Bu araç, kimya ve ticaret sektöründeki çeşitli uygulamalarda kullanılmaktadır [24].

2.9.6. INLEN

İlişkisel veri tabanından aldığı verileri makine öğrenimi teknikleriyle işledikten sonra ortaya çıkan sonuçları Veri tabanına yazmaktadır. Üretilen bilgi kesimi, basit ya da bileşik olabilmektedir [24].

INLEN aracında dört işleç vardır:

1. Veri tabanı yönetim işleci: Veri tabanı sorgularını yazmak için geliştirilen bir işleçtir.
2. Bilgi yönetim işleci: Üretilen bilgiyi yönetmek için kullanılır.
3. Bilgi üretim işleci: Veri tabanından bilgi almak ve makine öğrenimi algoritmalarını çağırmak için kullanılır.
4. Makrolar: INLEN işleçlerini bir sırada tanımlamayı ve tek bir işleç gibi kullanabilmeyi sağlar.

2.9.7. KDW (*Knowledge Discovery Workbench*)

Knowledge discovery workbench; kümeleme, sınıflama, bağımlılık analizi algoritmalarını kullanan bir araçtır. Etkileşimli veri analizine imkan vermektedir. INLEN sistemiyle birçok ortak özelliği bulunmaktadır [24].

2.9.8. SKICAT (*Sky Image Classification & Archiving Tool*)

Sky image classification & archiving tool, özel amaçlı bir veri madenciliği sistemidir. Özelleştiği konu astronomidir. Bu araç astronomik verileri indirgemek ve karar ağacı analizi için ID3, GID3, O-Btree algoritmalarını kullanmaktadır. Görüntü işleme, veri sınıflama ve

VTYS metodlarını kullanır. SKICAT adından da anlaşılacağı gibi gökyüzü fotoğraflarındaki gök cisimlerini tanımlamak, bunları sınıflandırmak, kataloglamak için kullanılan bir araçtır.

Sayısal gökyüzü fotoğraflarındaki gök cisimlerinin parlaklık, alan, çekirdek büyüklüğü gibi özelliklerini kullanarak sınıflandırma sorgusunu gerçekleştirmektedir.

SKICAT'ın deneysel testlerle fotoğraftan cisimleri tanıma ve sınıflandırma performansı %94 olarak saptanmıştır [24].

2.9.9. R-MINI

R-MINI, SKICAT gibi özel amaçlı bir veri madenciliği sistemidir. Finansal konularda özelleşen R-MINI sınıflama ve sapma tespiti yapmak için kullanılır. R-MINI Veri tabanından çektiği gürültü içerikli verileri kullanarak tamlık ve tutarlılık kriterlerini sağlayan en küçük kural kümesini bulur [24].

2.9.10. TASA (*Telecommunication Network Alarm Sequence Analyzer*)

Telecommunication network alarm sequence analyzer, telekomünikasyonda kullanılan özel amaçlı bir veri madenciliği sistemidir. Telekomünikasyon hatlarında oluşabilecek bir hatanın önceden tahmini için kullanılır. Zaman serileri arası bağımlılıklarda kullanılan veri madenciliği algoritmaları, hata tahmini için kullanılmaktadır. Hatlarda olağandışı bir olay meydana geldiğinde bu sistem tetiklenir. Tetikleme sayısının, kontrol edilebilecek sayının çok üzerinde olması böyle bir sisteme ihtiyaç doğurur [24].

2.9.11. GCLUTO (*Graphical CLustering Toolkit*)

Graphical CLustering Toolkit Minnesota Üniversitesi tarafından gerçekleştirilmiş bir araçtır. Bu araç kümeleme algoritmaları için geliştirilmiştir. Girdi kütüğünden aldığı verileri istenen kümeleme algoritmasına göre işleyip sonuçları çıktı kütüğüne yazmaktadır. Kolay kullanılabilir arayüze sahip olması ve görüntüleme problemlerinin iyi çözülmüş olması, üretilen sonuçların farklı gösterimleri ile GCLUTO kümeleme analizi için güçlü bir araçtır. Şekil 2.9'da GCLUTO arayüzünün bir görüntüsü yer almaktadır [24].



Şekil 2.9 GCLUTO arayüzünün görüntüsü

2.9.12. Enterprise Miner

SAS firmasının veri madenciliği aracıdır. SAS'ın Veri ambarı ve ÇAI (çevrimiçi analitik işleme) araçlarıyla bütünleşik çalışabilmektedir. Enterprise Miner karar ağaçları, yapay sinir ağları, regresyon analizi, 2-aşama modelleri (*two-stage models*), kümeleme, zaman serileri, ilişkilendirme, vb. veri madenciliği sorgularını ele alabilmektedir. Grafikselleştirme sayesinde kullanım kolaylığı sağlar ve kullanıcılar uygulamanın karmaşıklığından habersiz bir şekilde sadece girdi ve çıktıya yoğunlaşabilirler. 2 katmanlı mimariyi kullanır. İstemci bilgisayardaki yazılım gereksinimi Windows 98, 2000 ve NT'dir. Sunucu bilgisayardaki yazılım gereksinimi Windows 98, 2000 ve NT ile Linux'dır [24].

2.9.13. Veri Madenciliği Araçlarının Karşılaştırmaları

Bu bölümdeki veri madenciliği araçlarının karşılaştırmaları, Elder ve Abbott (1998)'un "A Comparison of Leading Data Mining Tools" isimli sunum sonuçlarından faydalanılarak oluşturulmuştur. Bu bölümdeki çizelgeler için, Tablo 2.5.'deki anahtarlar kullanılmıştır [24].

Tablo 2.5. Tablo (2.6-2.12)'de kullanılan anahtarların anlamları

Anahtar	Anlamı
Boşluk	Sıfır kapasite
X	Düşük kapasite
X-	Normal kapasite
X+	Yüksek kapasite

Tablo 2.7’de veri madenciliği araçlarının veri girdisi ve çıktısı açısından karşılaştırmaları, Tablo 2.8’de veri madenciliği araçlarının desteklediği karar ağaçları, Tablo 2.9’de veri madenciliği araçlarının kullanılabilirlik karşılaştırmaları, Tablo 2.10’da veri madenciliği araçlarının görüntüleme açısından karşılaştırması, Tablo 2.11’de veri madenciliği araçlarının otomasyon açısından karşılaştırması, Tablo 2.12’de veri madenciliği araçlarının güçlü ve zayıf olduğu alanlar verilmiştir [24].

Tablo 2.6 Veri madenciliği araçlarının çalışabildiği platformlar

Platformlar	PC Standalone (95 / NT)	Unix Standalone	Unix Sunucu /PC İstemci	NT Sunucu PC İstemci	VT Bağlanabilirlik
Clementine	X	X +			X
Darwin			X		X
DataCruncher	X		X		X
Enterprise Miner	X		X +	X	X
GainSmarts	X	X			X
Intelligent Miner			X		X
MineSet		X			X
Model 1	X		X	X	X
ModelQuest	X	X			X
PRW	X				X
CART	X	X +			
Scenario	X				X
NeuroShell	X				
OLPARS	X	X			
See5	X	X +			
S-Plus	X				X -
WizWhy	X				

Tablo 2.7 Veri madenciliği araçlarının veri girdisi ve çıktısı açısından karşılaştırmaları

Veri Girdisi & Model Çıktısı	Otomatik Başlık	Veri Formatını Kaydetme	ODBC	Yerel Veri Tabanı Sürücüler	Özet Raporlar	Çıktı aynak Kodu
Clementine	X		X			X
Darwin		X	X			X
DataCruncher	X	X	X	X	X	
Enterprise Miner	X-		X	X	X-	X
GainSmarts	X	X		X	X	X
Intelligent Miner				X-		X
MineSet		X		X		
Model 1	X	X	X	X	X	X
ModelQuest	X			X	X	X
PRW	X	X	X		X	X
CART						
Scenario	X				X	
NeuroShell	X					
OLPARS		X				
See5	X-					
S-Plus	X		X		X	X
WizWhy	X				X	

Tablo 2.8 Veri madenciliği araçlarının desteklediği karar ağaçları

Karar Ağaçları	Sınıflandırılmış Maliyetler	Budama	Görsel Ağaçlar	"Çizelge"	C5 veya C4. 5	CHAID	Diğerleri
Clementine	X	X	X-		X		
Darwin	X			X			
Enterprise Miner	X	X	X	X	X-	X	
GainSmarts			X	X		X	X
Intelligent Miner			X				X
MineSet	X	X	X	X		X	
Model 1				X		X	
ModelQuest		X			X-		
CART	X		X	X+			
Scenario							X
S-Plus		X	X	X			
See5	X	X			X+		

Tablo 2.9 Veri madenciliği araçlarının kullanılabilirlik karşılaştırmaları

Kullanılabilirlik	Veri Yükleme ve İşleme	Model Kurma	Model Anlama	Teknik Destek	Genel
Clementine	X+	X+	X +	X +	X +
Darwin	X	X	X +	X	X
DataCruncher	X+	X +	X	X	X
Enterprise Miner	X	X	X	X	X
GainSmarts	X+	X	X	X	X
Intelligent Miner	X	X	X	X	X
MineSet	X	X +	X +	X	X +
Model 1	X+	X +	X +	X +	X +
ModelQuest Enterprise	X	X +	X +	X +	X +
PRW	X+	X +	X +	X +	X +
CART	X-	X	X	X	X
Scenario	X	X +	X +	X	X +
NeuroShell	X	X	X	X	X
OLPARS	X-	X	X	X	X
See5	X	X	X	X	X
S-Plus	X	X	X +	X	X
WizWhy	X	X	X +	X	X

Tablo 2.10 Veri madenciliği araçlarının görüntüleme açısından karşılaştırması

Görüntüleme	Histogramlar	Pasta Grafiği	Dağılım/ çizgi çizimleri	Dönen dağılım	Şartlı çizimler	Sınıflandırma karar bölgeleri	Korelasyon çizimleri
Clementine	X		X		X	X-	X
Darwin	X-	X-	X-				
DataCruncher	X	X	X		X		
Enterprise Miner	X	X	X	X-	X		X
GainSmarts	X		X-				
Intelligent Miner	X-	X	X		X		
MineSet	X	X	X	X	X		
Model 1	X		X	X			
ModelQuest Enterprise	X		X				
PRW	X		X	X			
CART							
Scenario							X
NeuroShell			X				
OLPARS	X	X	X	X-	X	X	
See5	X						
S-Plus	X	X	X		X		X
WizWhy							

Tablo 2.11 Veri madenciliği araçlarının otomasyon açısından karşılaştırması

Otomasyon	Otomasyon Metodları	Özgür Metin Açıklamasının Adımları
Clementine	Görsel programlama , Programlama dili	X
Darwin	Programlama dili	X
DataCruncher	(İş Yöneticisi)	
Enterprise Miner	Görsel programlama, Programlama dili	X
GainSmarts	Makro Dili , Sihirbazlar	X-
Intelligent Miner	(Sihirbazlar)	
MineSet	Veri Geçmişi , Log	
Model 1	Model Sihirbazları	
ModelQuest	Toplu Ajanda	
PRW	Deney Yöneticisi ; Makrolar	X
CART	<i>Built-in Basic Scripting</i>	
Scenario		
NeuroShell		
OLPARS		
See5		
S-Plus	<i>Scripting (S) ; C / C ++</i>	
WizWhy		

Tablo 2.12 Veri madenciliği araçlarının güçlü ve zayıf olduğu alanlar

Otomasyon	Otomasyon Metodları	Özgür Metin Açıklamasının Adımları
Clementine	Görsel programlama , Programlama dili	X
Darwin	Programlama dili	X
DataCruncher	(İş Yöneticisi)	
Enterprise Miner	Görsel programlama, Programlama dili	X
GainSmarts	Makro Dili , Sihirbazlar	X-
Intelligent Miner	(Sihirbazlar)	
MineSet	Veri Geçmişi , Log	
Model 1	Model Sihirbazları	
ModelQuest	Toplu Ajanda	
PRW	Deney Yöneticisi ; Makrolar	X
CART	<i>Built-in Basic Scripting</i>	
Scenario		
NeuroShell		
OLPARS		
See5		
S-Plus	<i>Scripting (S) ; C / C ++</i>	
WizWhy		

2.10. Uygulama Alanları

Bu bölümde veri madenciliğinin uygulama alanları incelenerek ve bu uygulama alanları için geliştirilmesi önerilen veri madencilik araçlarının nasıl özelleştirilmesi gerektiği anlatılmıştır. Veri madencilik alanları genel olarak şu şekilde gruplandırılabilir [2]:

- Biyoloji
- DNA sıra(veri) analizi
- Ticari
- Bankacılık ve finans
- Perakende satış
- Telekomünikasyon
- Sağlık hizmetleri

Şimdi bu alanlarda yapılmış çalışmalara bakalım.

2.10.1. Biyomedikal için Veri Madenciliği

Gen fonksiyonları ve DNA sıralama örüntülerinin keşfedilmesiyle insan gen yapısı üzerine çalışmalar hız kazanmıştır. Biyomedikal araştırmalar DNA veri analizi üzerinde yoğunlaşmıştır.

DNA dizilimlerinin tüm yaşayan organizmalardaki genetik kodların temelini oluşturmasından dolayı “DNA dizilimi” gen araştırmalarında da önemlidir. DNA dizilimi şu 4 temel bloktan (nükleotit) oluşur: Adenin (A), Sitozin (S), Guanin (G), Timin (T).

Bu nükleotitler birbirine sarılı bir yangın merdivenini andıran bir görüntü oluştururlar. Belirli bir sırada dizilen nükleotitlerin oluşturduğu yapıya “Gen” adı verilir. Nükleotitler sonsuz sayıda dizilip sıralanarak farklı genler oluştururlar. İnsanda yaklaşık 100.000 gen vardır. Hastalıklara yol açan gen sıralama örneklerini binlerce gen arasından bulmak, tanımlamak oldukça zor bir iştir. Veri madencilikte, geliştirilen sıralama örnek analizi ve benzerlik arama metodları DNA verisi üzerinde analiz yapmayı kolaylaştırır [2].

Bu konuda yapılan çalışmalara bakıldığında:

- **Heterojen, dağınık gen veritabanının anlamsal bütünlüğü**

DNA verisinin dağınık olması, kontrol edilememesi, çok farklı çeşitlerinin kullanılması, DNA veritabanı üzerinde sistematik ve tutarlı analizler yapılma sorun çıkarır. Bu sorun, birincil ve türetilmiş gen verisinin saklanması ve yönetilmesi için geliştirilen birleşik veri ambarları ve dağınık veritabanlarının çıkmasıyla aşılmıştır. Veri temizleme ve veri birleştirme metodlarının geliştirilmesi, gen verisinin bütünleştirilmesi ve analizinde yardımcı olacaktır.

- **DNA sıralarının karşılaştırılması ve benzerlik arama**

Sağlıklı ve hastalıklı deneklerden alınan gen örnekleri karşılaştırılarak, kritik farklar tespit edilir ve hastalığa neden olan gen sıra deseni bulunmaya çalışılır. Daha önce bahsedilen zaman serilerindeki benzerlik arama metodları bir kaç değişiklik ardından burada da kullanılabilir.

- **Gen sırasının tanımlanması ve İlişkisel çözümleme**

Pek çok hastalık sadece bir genin değil, birden fazla genin beraber çalışmasıyla harekete geçer. İlişkisel çözümleme metodu, birlikte aktif olan genlerin tespitinde kullanılır ve bu sayede gen gruplarının tanımlanması ve birbiriyle etkileşiminin ortaya çıkarılması kolaylaşır.

- **Hastalığın farklı safhalarında genlerin bağlanması**

Hastalığın farklı safhalarında, farklı gen grupları aktif hale geçerek, hastalığın yayılımını sağlar. Eğer bu gen gruplarının aktif olma sırası tespit edilirse, hastalık tedavisinde önemli adımlar atılabilir.

- **Görsel araçlar ve genetik veri çözümleme:**

Genlerin sıra örneklerini grafik, ağaç, kübik ve zincir gibi görsel araçlar ile gösterilmesi örüntünün daha iyi anlaşılmasını ve daha kolay bilgi tespitini sağlar [2].

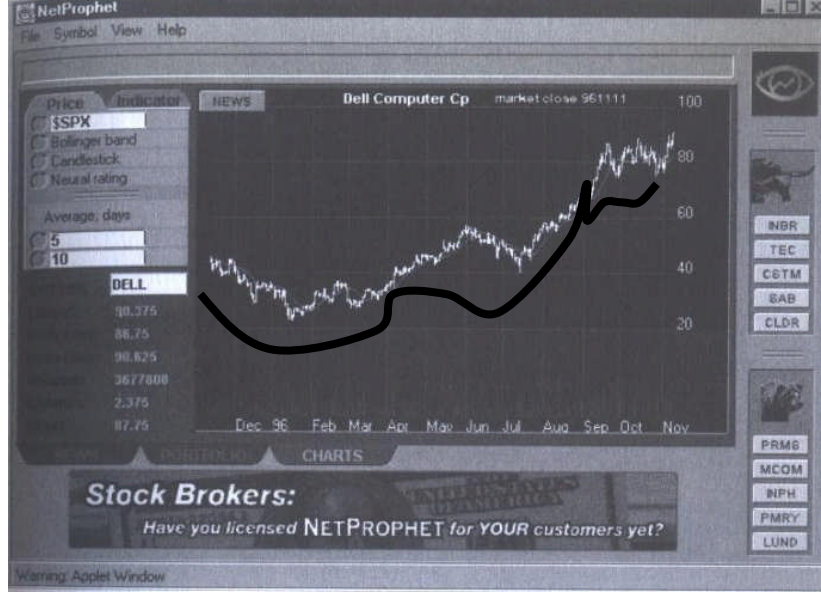
2.10.2. Bankacılık ve Finansal Uygulamalarda Veri Madenciliği

Pek çok banka ve finansal kuruluş hesap açtırma, yatırım yapma gibi banka hizmetleri, otomobil, iş, ev gibi kredi hizmetleri sunmaktadır. Bankacılıkta ve finansal sanayide kullanılan verinin tam, güvenilir ve yüksek kalitede olması sistematik veri analizi uygulanmasını kolaylaştırır.

Bankacılık sektöründe, veri madenciliği, kredi sahtekârlığı tahminleme, risk değerlendirme, eylim analiz değerlendirme, kar analizi gibi alanlarda kullanılır. Finansal kuruluşlarda, stok fiyat tahminleme, gümrük ölçümleme, portföy yönetimi, mal fiyat tahminleme gibi alanlarda yapay sinir ağları kullanılır [2].

Bu alanda yapılan çalışmalar:

- **NETPROPHET** adlı yazılım şu an pek çok firma tarafından kullanılan stok tahmin uygulamasıdır. Şekil 2.10'deki grafikte bulunan iki çizgi gerçek ve tahminlenen stok değerlerini gösterir.



Şekil 2.10. NETPROPHET stok tahminleme

- Bankacılıkta ise, daha çok sahtekârlık saptama ve yönetiminde veri madencilik uygulamaları kullanılmaktadır. HNC adlı şirketin **FALCON** adlı ürünü şu an Amerikada pek çok banka tarafından kullanılmaktadır. Bu ürün sayesinde banka sahtekârlık girişimlerini engelleyerek yatırımlarının %50 arttığını tespit etmiştir. Bu yazılım aşağıdaki araçları kullanarak veri madenciliğine başvurur.
- Görsel Araçlar: “Zaman-İnsan” ekseninden oluşan grafikte işlerin dökümünü sunar.
- İnsanların ve aktivitelerin eşleştirildiği çözümleme araçları
- Sınıflama Araçları: Gereksiz özelliklerden kaçınıp ilgili olanları çalışmalarda kullanır.
- Farklı davranışları gruplayan araçlar
- Beklenmeyen miktarda yapılan para transferini inceleyen araçlar
- Beklenmeyen yerlere erişim yapanları karakterize eden sıralı örüntü çözümleme araçları
- Amerika Bankası kendi ürünlerini kullanan banka müşterilerinin tesbitinde veri madenciliği kullanır ve müşteri ihtiyaçlarını karşılamak için ürün ve hizmetlerden oluşan paketler sunar.
- Farmer Group şirketinin “Risk Analiz Yönetim” paketi, sigorta oranı belirlenmesi, yatırım portföyü yönetimi, iyi ve kötü kredi riskleri taşıyan şirketlerin ve müşterilerinin belirlenmesinde veri madenciliği kullanır.
- Twentieth Century Fox adlı film şirketi ise fatura bilgilerini analiz ederek hangi aktörün, hangi filmin, hangi bölgede daha çok izlendiğini tespit ederek yeni film projelerini başlatmış ve bölge bazıda gösterimler sunmuştur [2].

2.10.3. Perakende Sanayii Uygulamalarındaki Veri Madenciliği

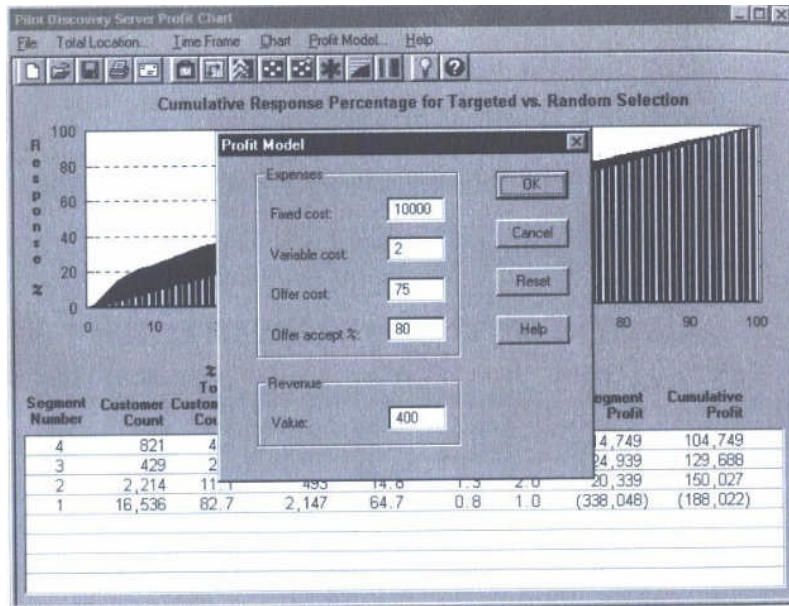
Perakende sanayii, satış bilgileri, müşteri profili, ürün nakil bilgileri, servis kayıtları gibi pek çok veriyi işlediğinden veri madenciliğini kullanır. Ayrıca günümüzde çoğalan elektronik ticaret uygulamaları ile perakende sanayii veri madenciliği için zengin bir kaynak sunmaktadır.

Perakende veri madenciliği şunları sunmaktadır:

- Müşteri alış davranışı belirlenmesi
- Müşteri alışveriş örnekleme ve eğilimin belirlenmesi
- Müşteri servis kalitesinin artırılması
- Müşteri hatırlama ve memnuniyetinin sağlanması
- Ürün tüketim oranının artırılması
- Ürün nakil ve dağıtım düzeninin tasarlanması
- İş maliyetinin düşürülmesi

Bu alanda yapılan çalışmalar:

- Şekil 2.11’de görülen **PILOT** yazılımı pazara yönelik kampanya belirleme çalışmalarında, yapılan alışverişe göre sınıflandırma algoritmalarını kullanarak, müşteri sınıflandırma sonuçlarını ve kar oranlarını göstermektedir.



Şekil 2.11. PILOT Pazara yönelik kampanyalar

- **IBM** veri madenciliği, POS makineler üzerindeki satış bilgilerinin çözümlemek ve öneriler sunmak için kullanır. 2 milyar dolar yatırımı olan 15 eyalette 129 mağazası olan bir firma üzerinde yapılan çalışmalar sonucunda şöyle bir sonuca varılmıştır: “Mağazaya gelen müşteriler promosyonlu ürünlerin bulunduğu sol taraftaki raflara

yönelmekte ve alışveriş için diğer bölümleri gezmeye ihtiyaç duymamaktadır.” Bu sonuç ardından mağazanın dekorunda yapılan değişikliklerle satışlarda artışlar gözlemlenmiştir.

- “Müşteri Hatırlama Yönetim” paketi SABRE şirketinin havayolu sanayisi için geliştirdiği veri madencilik uygulamasıdır.

Bu uygulamalar dışında perakende sanayisinde veri madenciliği şu alanlarda da kullanılır:

- Veri madencilik, müşteri ihtiyaçları, ürün satış eğilimi gibi değerlerin, kalite, maliyet, kar, ürün servisi gibi ölçütlere göre değişimi görüntülemek için görsel araçları kullanarak grafiksel sonuçlar verir.
- Müşteriyi etkilemek ve ürünü geliştirmek için satış kampanyalarına başvurur. Dikkatli analizler yapılarak sonuçlandırılan etkileyici satış kampanyaları şirket kazancını artırır. Çok boyutlu çözümlemeler satış öncesi ve sonrası satış miktarlarını karşılaştırarak kampanyanın başarısı hakkında bilgi verir.
- İlişkilendirme madenciliği ile, satış kayıtlarından bir ürünü alan müşterilerin başka bir ürünü de aldığı sonucu çıkarılır ve paket ürün satışı kararı verilebilir. Bu hem müşteri servisini iyileştirir hem de satışları artırır [2].

2.10.4. Telekomünikasyon Sanayii Uygulamalarındaki Veri Madenciliği

Telekomünikasyon sanayisinde, ses, faks, cep telefonu, resim, e-posta, Internet gibi kapsamlı iletişim servisleri hem uzak hem de yakın mesafelere sunulduğundan büyük miktarda veri transferi söz konusudur. Bilgisayar ve Internet teknolojilerinin gelişimiyle bu sanayide hızla gelişmekte ve rekabet artmaktadır. Rekabetin arttığı durumda müşteri memnuniyeti önem kazanır bu nedenle, iş içeriğini daha iyi anlamak, telekomünikasyon örnekleri belirlemek, kaynakları daha iyi kullanmak, servisin kalitesini artırmak ve sahtekârlık girişimlerini engellemek gibi hizmetlerde veri madenciliği kullanılır [2].

Bu alanda yapılan çalışmalar:

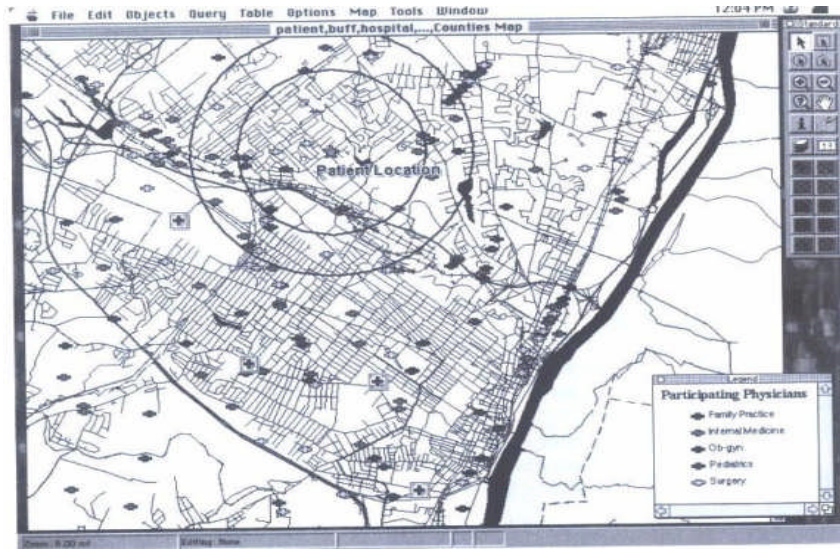
- Coloroda şirketinin FraudBuster adlı ürünü veri madencilik tekniklerini, üyelerin iletişim malzeme kullanım örneklerinin modellenmesi ve sahtekârlıktan şüphelenilen bir çağrının tahminlenmesi için kullanır.
- MapInfo şirketinin ürünü, görsel araçlar ile kablosuz iletişim ağlarındaki sorunlu bölgeleri tespit etmekte ve yeni stratejiler belirlenmesini sağlamaktadır.

Bu uygulamalar dışında telekomünikasyon sanayisinde veri madenciliği şu alanlarda da kullanılır:

- İletişim bilgileri genellikle arama zamanı, süresi, arayanın ve aranılanın yeri, arama çeşidi gibi boyutlara sahiptir. Bu verilerin çözümlenmesi ile, veri trafiği, sistem çalışma zamanı, kaynak kullanımı, kullanıcı davranışı, kar gibi bilgiler belirlenip, karşılaştırılır.
- Telekomünikasyon servisleri, müşteri grubu, ay ve gün bazında örneklendirerek kullanım grafiği çıkarır. Bunu için kayıtları müşterilere tutarak bu kayıtlar üzerinde “Drill up“ ve “Drill down” işlemleri yapılarak, hangi meslek gruplarının ya da hangi şehirlerin daha çok iletişim kurduğu tespit edilip meslek grubu veya şehir tabanlı servisler sunulabilir [2].

2.10.5. Sağlık Hizmet Uygulamalarındaki Veri Madenciliği

Sağlık alanında da veri madenciliği çok kullanılmaktadır. Vysis, ilaç sanayisi için protein analizini yapay sinir ağlar algoritmasını kullanarak yapar. Rochester Kanser Merkezi bölümü araştırmalarında KnowledgeSEEKER adlı karar ağacı tekniğini kullanır. California Hastanesi veriyi madenlemek için “Information Discovery” adlı ürünü kullanır. Hastanede çalışan bir doktor, bu program sayesinde hastalarını hiçbir fiziksel teste tabii tutmadan teşhis koyabildiğini açıklamıştır. Şekil 2.12’de görüldüğü üzere MapInfo şirketi haritalama teknolojisini kullanarak hastaların yoğun olarak bulunduğu bölgeleri işaretlemektedir bu sayede bölgeye daha iyi hizmet sunulabilir [2].



Şekil 2.12. MapInfo Hastaların ve en yakın sağlık ekiplerinin yerinin haritalanması

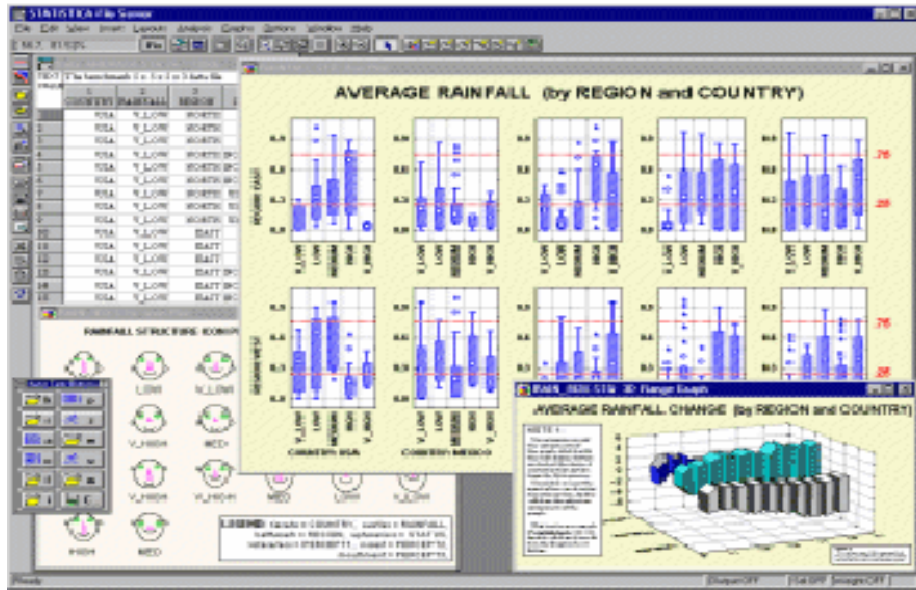
2.11. Görsel ve Sesli Veri Madenciliği

Görsel veri madenciliği, saklı ve yararlı bilgileri görsel veri ve bilgi tekniklerini kullanarak geniş veri kümelerinden ortaya çıkarır. İnsan görsel sistemi, güçlü, yüksek paralel süreçleme yapabilen ve büyük bir bilgi tabanını içeren sonlandırma mekanizması olan, göz ve

beyin tarafından kontrol edilir. Görsel veri madenciliği aslında, verideki veri dağılımları, örüntüler, kümeler, ve sınırdışı öğelerinin gücünü daha çekici ve etkin bir araç yapmak için birleştirir.

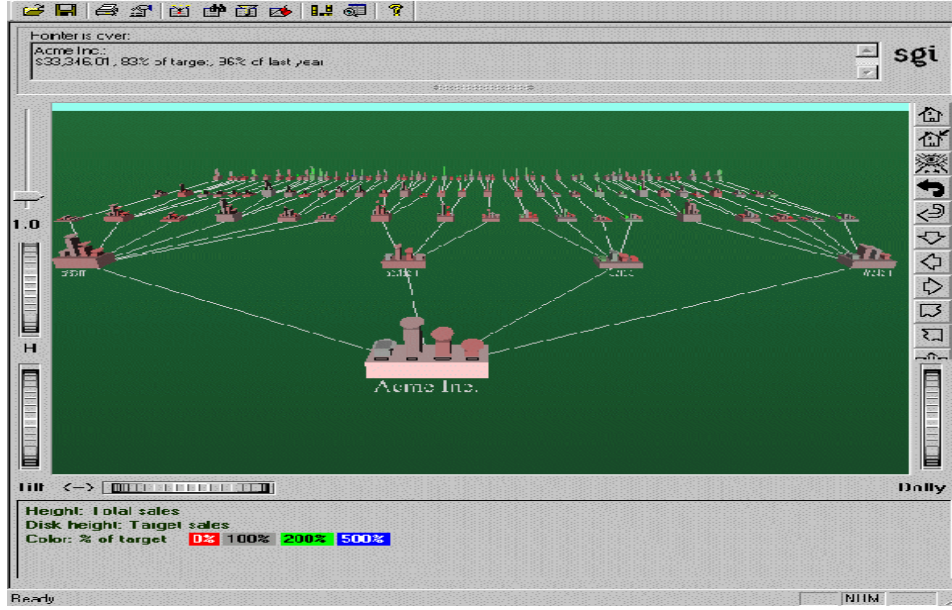
Görsel veri madenciliği, iki disiplinin kombinasyonu olarak görülebilir: veri görselleme ve veri madenciliği. Ayrıca, bilgisayar grafikleri, çoklu ortam sistemleri, örüntü tanıma ve yüksek performans hesaplama konularıyla da yakından ilgilidir. Genel olarak veri görselleme ve veri madenciliği aşağıdaki yollarla birleştirilebilir [2].

- **Veri Görselleme:** Veri tabanındaki ya da ambardaki veri, değişik soyutlamanın ya da tanelemenin değişik seviyelerinde ya da niteliklerin ya da boyutların değişik kombinasyonlarında izlenebilir. Veriler boxplots, 3-D cubes, veri dağılım haritaları, eğriler, yüzeyler, bağlı grafikler ve vb. çeşitli görsel formlarda gösterilebilir. Şekil 2.13’de Statsoft’un çok boyutlu uzayda veri dağılımı görülebilir. Görsel görüntüler kullanıcılara, veri tabanındaki veri karakteristikleri için açık bir izlenim vermek ve genel bakış kazandırmak açısından yardımcı olur [2].

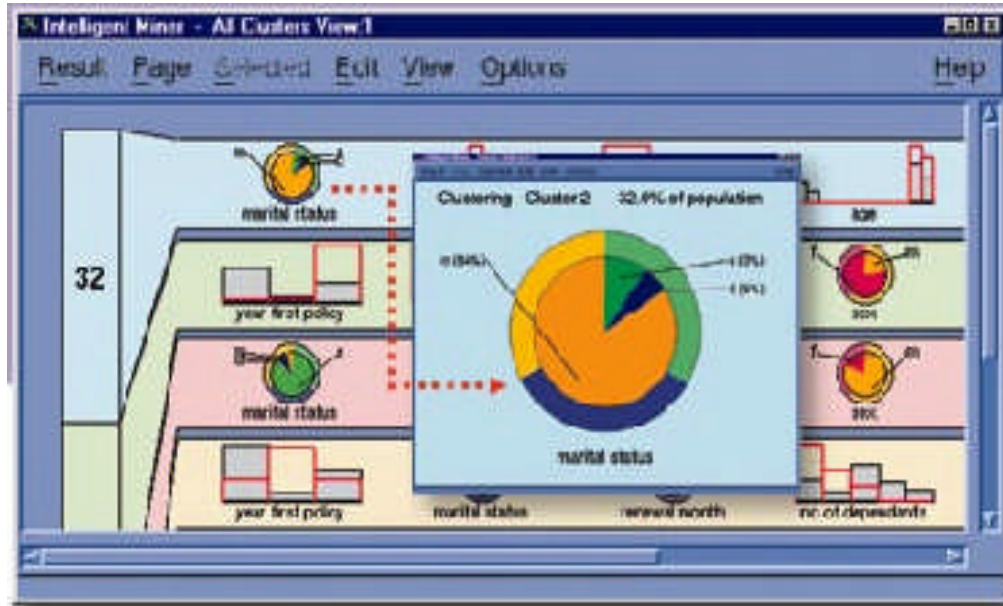


Şekil 2.13. Statsoft’un çok boyutlu uzayda veri dağılımı

- **Veri madenciliği sonuç görselleme:** Veri madenciliği sonuçlarının görsellenmesi, veri madenciliğinden elde edilen bilgi ve sonucun görsel formlarla sunulmasıdır. Bu formlar scatter plot, boxplot, decision tree, association rule, cluster, outlier, generalized rule vb. Örneğin: SAS Enterprise Miner’in scatter plot ile yaptığı çalışma Şekil 2.14 de görülebilir. Şekil 2.15 MineSet 3.0, veri tabanından çıkartılmış verilerin ilişkileri çıkartmak için kolonları kümesini bir düzlemde ilişkilendirir. Şekil 2.16 yine Mine Set 3.0’dan karar ağacını gösterir.



Şekil 2.16. Mine Set 3.0 karar ağacı

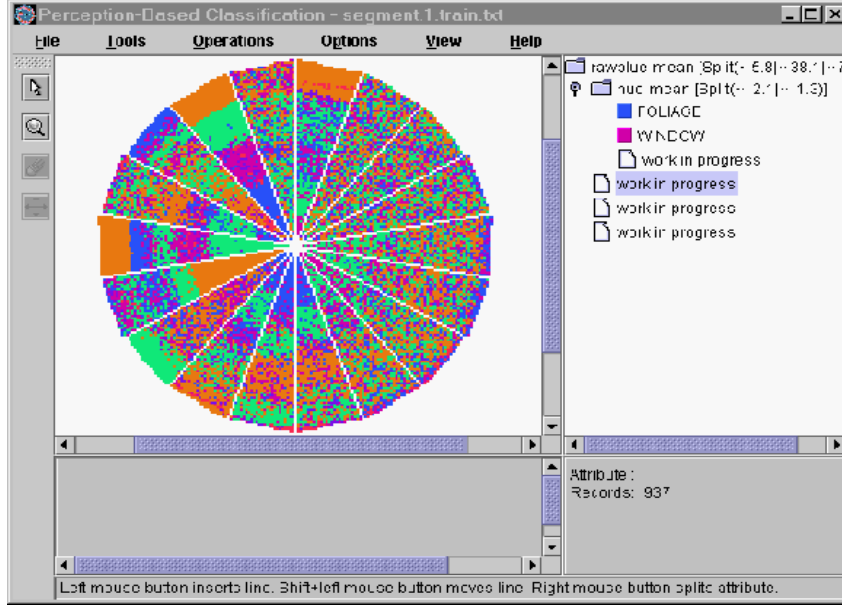


Şekil 2.17. IBM Intelligent Miner'dan, küme gruplaması

Veri madenciliği ile süreç görselleme: Bu tip görselleme, verinin nasıl çıkarıldığını, hangi veri tabanı ya da veri ambarından çıkarıldığını, seçilen verinin nasıl temizlendiğini, birleştirildiğini ve ön sürecini gösterir. Ayrıca, veri madenciliği için hangi metodun seçildiğini, sonucun nerede saklandığını ve nasıl görülebileceğini gösterir [2].

- **Etkileşimli görsel veri madenciliği:** Etkileşimli görsel veri madenciliğinde, görsel araçlar veri madenciliği sürecinde kullanıcılara akıllıca veri madenciliği kararları almasında

yardımcı almada kullanılabilir. Örneğin nitelikler kümesindeki dağılım bütün yüzeyin bir daire ya da kolon olarak gösterilmesine bağlı olarak renkli dilimler ya da kolonlarla görüntülenebilir. Bu görüntü, belki kullanıcılara hangi sektörün sınıflandırma için önce seçilmesi ve bu sektör için güzel bir ayırma noktasının nerede olabileceğini belirlemede yardım edebilir. Şekil 2.18, Münih Üniversitesi tarafından geliştirilen PBC, algısal-tabanlı sınıflandırma sisteminin çıktılarını gösterir [2].



Şekil 2.18. Algısal-tabanlı sınıflandırma

Sesli veri madenciliği, veri alanlarının ya da veri madenciliği sonuçlarını özelliklerini göstermede radyo sinyallerini kullanır. Görsel veri madenciliği, grafik görüntüleri kullanarak ilginç alanlar ortaya çıkarmasına karşın, kullanıcıları, alanları izlerken ilginç ve yeni özellikler bulmaya konsantre etmeye gereksinim duyar. Bu bazen çok yorucu olabilir. Eğer alanlar ses ve müziğe transforme edilebilirse, resimleri izlemek yerine perdeleri, ritimleri, tonları ve melodiyi dinleyebiliriz. Bu belki görsel konsantrasyon sağlamada sıkıcı olabilecek bir çok alan için daha rahatlatıcı bir alternatif olabilir [2].

BÖLÜM 3

3. VERİ TABANLARINDAN BİLGİ ÇIKARIMI

3.1. Bilgi Tabanlarının Kurulum Aşamaları

Yapay zekâ sistemlerinde veri tabanlarında bilginin işlenilmesi temel problemlerden biridir. Fakat burada, uzman insandan bilginin nasıl alınacağı ve yapay sisteme nasıl iletileceği inceleme ve araştırma konusudur. Adeta uzman insan, bir problemin çözümü için çoğu zaman prosedür ve niceliksel yöntemlerden değil, benzetimden, tecrübeden, içgüdüden ve soyutlaştırmadan yararlanmaktadır. Ayrıca bazı durumlarda uzman, çözümü nasıl bulduğunu anlatamamaktadır. Uzmanla yapay sistem arasında arabuluculuk yapan, uzmanın bilgilerini sisteme aktaran ikinci bir uzman daha var ki, o da bilgi mühendisi olarak adlandırılır. Bilgi mühendisi uzmanın problem ve onun çözümü hakkında görüşlerini kesin, tam ve zıt olmayan bir sisteme dönüştürmeli ve bu sistem, problem için uzmanın kendisinden daha kötü çözüm getirmemelidir. Bilgi tabanının oluşturulması süreci uzmandan alınan bilgiler temelinde üç aşamada oluşur [25].

1. Disiplin alanının tasviri
2. Bilginin elde edilmesi
3. Bilginin sunulması yöntemi ve modelinin seçilmesi

Bilgi tabanının oluşturulması süreci yapılandırılmış ve lineer değildir. Aslında bu süreç, kötü yapılandırılmıştır ve döngülüdür. Bilgilerle uğraşan, yöntem ve sistemlerin tamamı son yıllarda bilgi mühendisliği olarak adlandırılır. Bu disiplin dalı, bilimin birçok alanında (örneğin kavram psikolojisi, enformasyon teorisi, enformatik ve yapay zekâ) aktif olarak kullanılmaktadır [25].

3.2. Disiplin Alanının Tasviri

Bilgi sistemi kurulduğunda, ilk adım bu sistemin yöneldiği problemlerin çözümüne ait disiplin alanının belirtilmesidir. Yani bilgi mühendisi sistemin uygulanacağı alanın sınırlarını ve bu alanda çözülecek olan problemin sınıfını belirlemelidir. Bu iş beş aşamada yerine getirilir.

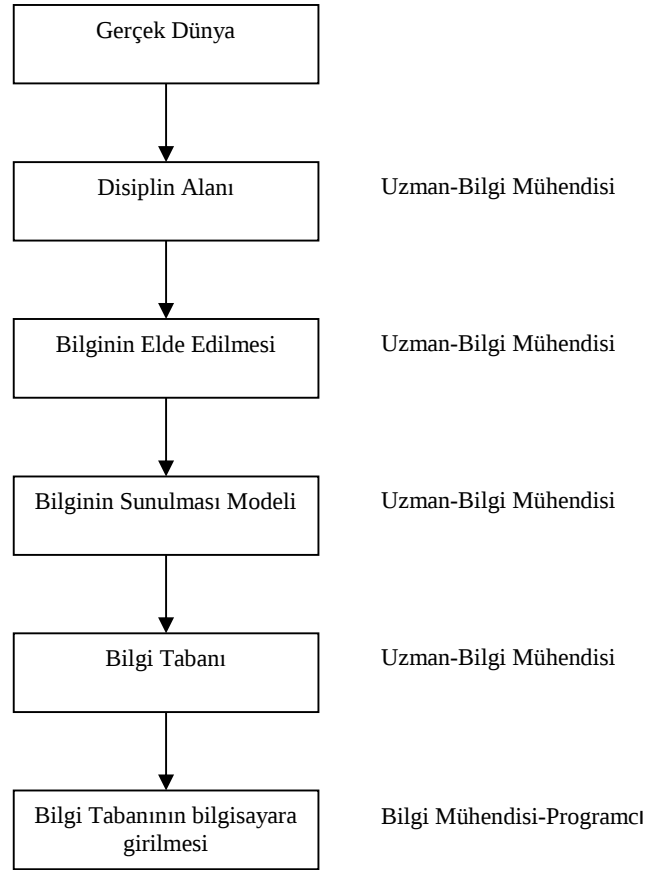
1. Çözülen problemlerin karakterlerinin belirlenmesi,
2. Disiplin alanının nesnelerinin seçilerek birbirlerinden ayırt edilmesi,
3. Nesneler arasındaki ilişkilerin tespit edilmesi
4. Bilginin sunulma modelinin seçilmesi
5. Disiplin alanının özel niteliklerinin meydana çıkarılması

Bilgi mühendisliği problemi, düzgün olarak ortaya koymayı becermelidir. Problem yapılaşmamış ise o bunu görmeli ve bu halde onu formal bir hale getirmekten veya sistematik çözüm yöntemlerini uygulamaktan kaçınmalıdır.

Bilgi tabanı kurulmasının ilk aşaması, çeşitli soyut seviyelerde disiplin alanı tasvirinin nasıl olacağını belirlemektir. Bilgi tabanının soyutlaştırılmasının esas adımları Şekil 3.1’de görülmektedir. Bilgi sistemleri, bunun içinde Uzman Sistemler de gerçek dünyanın belirli özlerinin soyutlaştırılması sonucu alınan herhangi bir disiplin alanının formelleştirilmesi yolu ile oluşturulur. Örneğin; tıp disiplin alanından bir örneğe bakalım. Herhangi bir hastalığın belirtileri sadeleştirilmiş şekilde şöyle gelişir. İnsan belirli bir genetik karakteristikle doğar. Çevredeki çeşitli faktörler de (beslenme rejimi, iklim ve ekolojik koşullar vb.) onun organizmasının durumunu etkilemektedir. Bu faktörlerin etkisi ile organizmanın görevlerinin bozulması belirtileri oluşur. Bu belirtiler ise hastalığın başladığını haber verir. Pratikte çeşitli uzman hekimler bir insanın aynı hastalığına kendi disiplin alanına göre çeşitli açılardan bakmaktadır. Teşhisçi hekim, organizmada kök salmış hastalığı görmektedir. O, hastanın şikâyetlerine dayanarak onda bir hastalığın olduğu kanaatine varır ve böylece bir hipotez ileri sürmüş olur. Sonra bu hipoteze dayanarak hastayı tedavi etmeye ve onu sağlamlaştırmaya çalışır. Bir koruyucu hekim ise, hastalığın ilk aşamasında onun belirtilerini ortaya çıkarmaya ve hastalığı erken tedavi etmeye çalışır. Fakat hastalık aktif fazda olursa zorluk çekecektir. Böylece her bir hekim uzman olmasına rağmen aynı organizma hakkında, hastalığın teşhisi ve tedavisi hakkında kendi özel disiplin alanına sahiptir. Yani disiplin alanının belirlenmesi ve başka yakın alanlardan ayrılması gerçek dünyanın soyutlaştırılmasının ilk adımıdır.

Disiplin alanı belirlendikten sonra, bilgi mühendisi onu formel olarak tanımlamalıdır. Bunun için o bilgilerin sunulması yöntemlerinden birini seçmelidir. Halen herhangi bir disiplin alanı, formelleştirmeye imkân veren genel bir yöntem mevcut değildir. Buna göre bilgi mühendisi disiplin alanının özelliklerini en iyi şekilde sunabilen modeli bulmalı ve bu modeli kullanmalıdır. Bilginin sunulmasının genel teorisi oluşturulursa bu durumda herhangi bir disiplin alanının özelliklerini dikkate almadan böyle genel yöntemi kullanmak mümkün olur.

Disiplin alanının formelleştirilmesinden sonra, alınmış bilgi tabanı bu alanın soyutlaştırılması sonucunda elde edilir. Bu durumda disiplin alanı gerçek dünyanın soyutlaştırılması neticesinde ondan ayrılmıştır. Bilgisayardan farklı olarak insanın çeşitli tip bilgi sunma modellerini kullanması, gerçek dünya kavramlarına çeşitli bilgileri kıyaslaması, problemin çözümü için çeşitli yollara başvurusu mümkündür. Çeşitli disiplin alanlarına yönelmiş bilgi tabanlarının, birlikte kullanılmasının ayrı ayrı örnekleri vardır, fakat çağdaş sistemlerin çoğu yalnızca bir disiplin alanına ait olan problemlerle çalışmaktadır [25].



Şekil 3.1 Herhangi bir disiplin alanında bilgi tabanının oluşturulması safhaları

3.2.1. Çözülen Problemlerin Karakterinin Belirlenmesi

Örneğin tıptan verilen örneği devam ederek, laboratuarda yerine getirilen kimyasal kan analiz sonuçlarını işlemek için bir uzman sistem kuralım. Bilgi mühendisi her şeyden önce bu alandaki uzmanı sorgulamalı ve sonra sistemi kurmaya girişmelidir. Önce sistemin amacı ve görevleri belirlenmeli, sistemin hangi problemleri çözebileceği kesinleştirilmelidir. Sistem tasarlanacağı zaman amacının tam, kesin olarak ve çelişkisiz belirlenmesi çok önemlidir. Örneğin, bir teşhis sistemi için şu sorulara cevap aranabilir:

1. Müşteri sağlıklı mıdır? Eğer değilse, onun hastalığı nedir? Eğer müşterinin birkaç hastalığı var ise onlardan en tehlikelisi hangisidir?
2. Yemekte ve beslenme rejiminde hangi değişiklikler tavsiye edilmeli ve onlardan en çok hangisine dikkat edilmeli?
3. Hangi laboratuvar araştırmaları yapılmalı ve onlardan hangisi acildir?
4. Müşterinin yaşam tarzını veya yaşadığı yerin hava koşullarını nasıl değiştirmeli?

5. Müşterinin muayene için diğer uzman hekimlere ihtiyacı var mı, var ise hangilerine?

Sistemin tasarım amacı belirlendikten sonra, bilgi mühendisi alt amaçları biçimlendirmeye girişir. Bu ona sistemin hiyerarşik yapısını belirlemeye ve modüllere bölmeye imkân verecektir. Bilginin ayrı ayrı kısımları arasındaki ilişkilerin olması, problemi daha az karmaşık olan iki veya daha çok alt probleme bölüp sonrasında onların çözümünün aranmasına getirilir. Tabi ki gerektiğinde problemlerin bölünmesi devam edebilir. Aynı yöntemi farklı alanlarda da uygulayabiliriz. Örneğin DEC şirketinin tasarladığı XCON uzman sistemi kullanıcıların siparişi ile bilgisayar sistemlerinin konfigürasyonunu belirler. Esas problem altı alt probleme bölünmektedir [25].

1. Donanımda ciddi bozukluklar olduğunda alış siparişini yoklamak,
2. Sistemin gereken bileşenlerini elde ederek montaj için işyerine götürülmesini sağlamak,
3. Montaj için sistemin gereken bloklarını elde ederek işyerine götürülmesini sağlamak,
4. Gereken kartları elde ederek montaj işlerinin yapılmasını sağlamak,
5. Sistemin bileşenlerinin birleştirilmesi ardışıklığını belirlemek,
6. Sistemin bileşenlerinin kablo ile birleştirilmesini sağlamak.

Çoğu zaman bilgi sistemleri önce daha zor alt problemlere, sonra daha basitlerine geçer. Hiyerarşinin bir sonraki seviyesine, o anki seviyedeki bütün problemlerin çözülmesi bittikten sonra geçilir. Örneğin tıp teşhisi örneğinde, ilk problemin çözümü aşağıdaki sorulara cevap verebilen alt problemlere bölünmektedir [25].

1. Kan dolaşım sisteminde aksama var mı?
2. Sindirim sisteminde aksama var mı?
3. Sinir sistemi düzenli mi?

Bütün bu alt problemler teşhis sisteminin karşısına koyulan daha yukarıda gösterilmiş beş sorudan birincisi ile ilişkilendirilebilir. Yani, bu alt problemler müşterinin sağlıklı olup olmadığı sorusunu detaylaştırır veya onun için alt amaçlar rolünü yerine getirmektedir. Denilenlerden anlaşılmaktadır ki, sistem problemin çözümünü aksi yönde aramaktadır ve bu zaman alt problemler daha da detaylaştırılmış olmaktadır. Detaylaştırma süreci o zamana kadar devam ettirilir ki ölçme veya laboratuvar analizleri sonuçlarını içeren gerçekler bulunsun, yani mevcut problem için ilk veriler aydınlatılmış olsun. Örneğin,

- Kimyasal kan analizi sonuçları,
- Hastalığın görünen belirtileri,
- Hava koşulları ve irsi faktörler [25].

3.2.2. Disiplin Alanındaki Nesnelerin Belirlenmesi

Bilgi tabanı tasarlanmasında bir sonraki adım disiplin alanı nesnelerinin belirlenerek ayrılması veya sistem teorisi dili ile söylenirse, sistemin sınırlarının saptanmasıdır. Her bir formel sistem gibi ayrılmış kavramlar tam, kesin ve çelişkisiz olmalıdır. Böylece, somut olarak “hangi laboratuvar analizleri yapılmalıdır?” sorusuna cevap aranacaktır. Müşterinin hastalık tarihine bakılmalı mı ve eğer bakılacak ise orada en önemli olanlar hangi verilerdir? Müşteri hakkında hangi bilgiler gerekmektedir? Müşteri hakkında hangi bilgiler önemli olabilir? Önceden hastanın kabul ettiği ilacı ve hekimlerin daha önce tayin ettiği tedaviyi dikkate almalı mı? Hastanın yaşam tarzı, işinin türü, hava koşulları ve beslenme rejimi herhangi bir rol oynamakta mıdır? Şu anda hangi belirtiler gözlenmektedir?

Bu soruların cevabı ilk verilerin sınırlarını çizmeyi sağlar. Çözümün aranması uzayını kurmak için, genel problemin amaçları hiyerarşisinin her bir seviyede alt amaçları belirlemek gerekir. Hiyerarşinin en başında genelliği ile sistemin imkanlarını ve amaçlarını belirleyen problem gelmelidir [25].

3.2.3. Nesneler Arasındaki İlişkilerin Saptanması

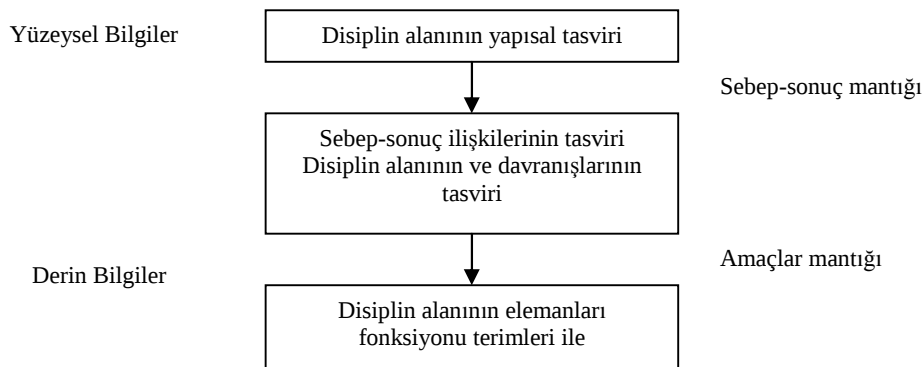
Disiplin alanının nesnelerinin belirlenmesinden sonra, onlar arasında hangi ilişkilerin olduğu tespit edilmelidir. Örneğin, kandaki tireotrop hormonunun düşük seviyede olması pankreasın fazla aktifliğine işaret edilebilir. Fakat başka bir şeyi de gösterebilir. Mümkün olduğu kadarı ile daha çok ilişki tespit edilmesine, idealde ise disiplin alanında mevcut bütün ilişkileri belirlemeye çalışmak gerekir. Diğer taraftan disiplin alanının elde edilmiş niteliksel tasviri, sistemin bilgi tabanına yerleştirmek için herhangi bir formel dilde sunulmalıdır. Bunun için bilginin sunulmasının öyle modeli seçilmelidir ki, disiplin alanı hakkındaki tüm bilgileri formel olarak ifade etmek mümkün olsun. Bazı bilgi sunma modellerinde nesneler arası ilişkiler dilsel, niteliksel, eylemsel, mantıksal, küme vb. şekilde yapılmaktadır. Dilsel ilişkiler; nesne, koşul, yer, araç, amaç, zaman vb. özellikleri, fiil ilişkiler; fiillerin zamanı, sayısı vb. biçimlerini, mantıksal ilişkiler; söylemler hesabı işlemlerini, kuvantiye ilişkileri; genellik ve mevcutluk kuvantorlarını, küme ilişkileri; kümenin elemanı, altküme, süper küme vb. kavramlarını içermektedir [25].

3.2.4. Bilginin Sunulma Modelinin Seçilmesi

Disiplin alanı nesnelerinin oluşturulması yönteminin seçilmesi ve bunlar arasındaki ilişkilerin belirlenmesi, bilginin sunulmasının herhangi bir modelinde yapılmalıdır. Bu aşama

çatı ve seramik şebeke yardımıyla bilgi sunulması sürecinde çok önem taşır ve üretici sistemler için de belirli bir rol oynamaktadır. Bazı disiplin alanlarının tasviri sınıflı yapıya sahiptir, yani bütün nesneler SINIFIN_ELEMANI_OLMAK akrabalık ilişkisi ile bağlıdır. Mesela, biyolojide kuş ile güllerin sınıflandırılması hiyerarşik prensibe göre yapılabilir. Fakat bütün disiplin alanları sınıflandırılmış yapıya sahip değildirler. Tıp alanından baktığımız örnek bunu doğrular. Bu durumlarda bilgi sisteminin tasarlanması süreci bütün üretimler kümesini sınıflara bölmekle basitleştirilebilir. Böyle bölünme bir sınıfa dahil olan kuralların grup grup dahil edilmesini ve yeniden düzenlenebilmesini sağlar. Teşhis sistemi örneğinde birinci alt amaçla ilgili kurallar grubunu ayırmak, yani kan dolaşımı sisteminde bozulmanın olup olmadığına cevap vermeye imkân sağlamaktadır. Bir sonraki detaylaştırma seviyesinin alt amacı daha az sayıda kural içeren diğer bir grupla ilgili olacaktır. Disiplin alanında bilginin teşkili yöntemleri, sınıfların terkihi, yapı türü ve ayrılmış nesnelerin listesi uzmanın ve bilgi mühendisinin bu disiplin alanına hangi açıdan baktığına bağlıdır. Disiplin alanı gibi bilgisayar tasarımı örneğini inceleyelim. Yapı hiyerarşisini tasvir etmek için bilgisayara hiyerarşik olarak yapılandırılmış fiziksel bileşenlerin toplamı gibi bakılır. Hiyerarşinin zirvesinde bilgisayarın kendisi durmaktadır. Bir sonraki seviyede o, bloklardan oluşmuş şekildedir. Üçüncü seviyede her bir bloğun yerleştiği kartlardan, dördüncü seviyede ise bileşenler mikro şemaya kadar detaylaştırılır.

Bilgisayar bozulduğunda (hiyerarşinin en üst düzeyi) önce her bir bloğun çalışıp çalışmadığı kontrol edilir. Bozulmuş blok bulunduğundan sonra, bu bloğu oluşturan kartlar belirlenerek bozuk mikro şema aranır. Arızanın bulunmasının bu yöntemi, uygulamayı yapan kişinin bilgisayar ve elektronik hakkında bilgilerinin olmasını talep etmemektedir. Fonksiyonel bloklar hakkında veriler de gerekli olmamaktadır. Bakılan bozukluk aranması için gereken bilgiler yüzeysel bilgiler olarak adlandırılır [25].



Şekil 3.2. Disiplin alanındaki hiyerarşi tipleri ve onlar arasındaki ilişkiler

Sebep-sonuç hiyerarşisi sistemin davranış modelini kurmaya imkân verir. Bilgisayarın çalışması sebep-sonuç terimleri ile verilir. Örneğin, hiyerarşinin zirvesinde klavyede bir tuşun basılması ekrana uygun sembolün getirilmesine veya yazıcıya çıkarılmasına neden olur. Eğer böyle bir şey gerçekleşmemişse, hiyerarşinin bir sonraki seviyesine inmek, ekran veya yazıcının çalışıp çalışmadığını yoklamak lazımdır. Yazıcı çalışmadığı takdirde bir sonraki seviyeye inerek voltmetre veya osiloskopun yoklanması lazımdır. Yazıcının bozulmasının diğer sebepleri de olabilir ve bunlara bakarak arıza yeri bölgeselleştirilir. Bu hiyerarşi tipinde daha derin bilgiler talep edilir, fakat onlar da bilgisayar gibi çok karmaşık bir sistem için nispeten yüzeysel sayılabilir.

Üçüncü ve en yüksek hiyerarşi tipi belirli fonksiyonları yürüten modüller ve alt sistemler hiyerarşisidir. Hiyerarşinin zirvesinde yer alan bilgisayarın fonksiyonu, verilmiş algoritma yardımı ile giriş verilerinin çıkış verilerine dönüştürülmesidir. Sonraki seviyede bir sıra fonksiyonlar yığını var; verilerin dâhil edilmesi, onların işlenmesi, depolanması, dış bellekten okunması ve yazılması. Eğer bilgisayar esas görevini yürütmezse o zaman hiyerarşinin bir aşağıdaki seviyesine inmek ve bilgisayarın hangi fonksiyonunun bozulduğunu belirlemek gerekir. Verilerin dış belleğe yazılmadığını, yani disket sürücüsünün çalışmadığını varsayalım. Bundan sonra disk sürücüsünün tasviri olan hiyerarşik seviyesinde yürümeyen fonksiyonun yerel olarak araştırılması yapılır. Örneğin, disket sürücüsüne bakıp motorun çalışıp çalışmadığını görmek mümkündür. Tıp alanındaki örneğe dönülürse, hiyerarşinin zirvesinde insan bedeninin fiziksel yapısı duracaktır. Bu yapı hiyerarşisine tam organizma, ayrı ayrı organlar, dokular, hücreler, moleküller ve genler dâhil olacaktır. Fonksiyonel hiyerarşik zirvede yine organizma duracak, sonraki seviyelere ise organizmanın alt sistemleri dahil olacaktır. Parçalama, istenilen detaylık derecesi alınana kadar yapılabilir. Neden-sonuç hiyerarşisinde hastalıkların belirtileri ve kimyevi elemanların adları durmaktadır. İki olay veya belirtilerin ortaya çıkarılması arasındaki ilişki, onlardan biri diğerine neden olursa, sebebiyet olarak adlandırılır.

Herhangi bir disiplin alanını tasvir ettikçe, yüzeysel bilgilerden derin bilgilere geçilirken, bilgilerin miktarı ve bu bilgilerin ayrı ayrı bileşenleri birleştirilerek bir tam birime getirileceği hakkındaki bilgi artmaktadır. Yapı hiyerarşisinden sistemin davranış tasvirine geçerken sebep-sonuç mantık olarak adlandırılan teori kullanılır. Derin bilgiler somut disiplin alanının esas kanunları, aksiyomları ve gerçeklerin tamamıdır. Derin bilgiler, disiplin alanındaki gelişen süreçler ve onların özelliklerine uygun genel prensipleri yansıtmaktadır. Yüzeysel bilgiler ise, disiplin alanının yapısını onu teşkil eden fiziksel bileşenlerle ifade etmeye imkân veren bilgilerdir. Bu bilgiler genel teori ve yöntemler olmadıkça, kullanılan heuristiklerden ve deneysel yolla elde edilen bazı kurallardan oluşurlar. Düşünce süresinde insan, bilgilerinin derin

veya yüzeysel olduğunun farkına varmadan her iki bilgiden aynı derecede kullanmaktadır. Bir uzman da, çalıştığında kullandığı bilgilerin hangi tip olduğunu bilmemektedir. Fakat bilgi sistemi tasarlandığında hangi tip bileşim gerektiğini ve buraya hangi tip hiyerarşinin dahil edileceği çok iyi bilinmelidir [25].

3.2.5. Disiplin Alanının Özel Niteliklerinin Meydana Çıkarılması

Disiplin alanında uygulamalı problemlerin çözümünü zorlaştıran bir sıra özel niteliklerin ortaya çıkarılması da gereken işlerdendir. Bu niteliklerin bir kısmı:

1. Bilirkişi incelemesi yalnızca bir somut alanda yapılabilir. Örneğin, bilgisayarın konfigürasyonunu belirlemek için kurulmuş sistem tıp teşhisi koyamaz.
2. Bilgi tabanı ve sonuç çıkarma mekanizması çeşitli bileşenlerdir. Hakikaten, bazen bir çıkarım mekanizması diğer veri tabanı ile kullanılabilir. Örneğin, kanda enfeksiyon analizi programı veri tabanı değiştirilerek aynı çıkarım mekanizması ile pulmanolojide de kullanılabilir.
3. Bilgi sistemleri, problemin çözümünde takip edilen yollarını kullanıcı anlayacak derecede izah edebilmelidir. Eğer, uzman “Neden?” sorusuna cevap veremezse onun cevabı kabul edilemez.
4. Çıkış sonuçları nicel olarak değil, niteliksel olarak alınmaktadır.
5. Bilgiye dayalı sistemler onların bilgi tabanını yavaş yavaş artırmaya imkan veren modül prensibi esasına dayalı kurulmaktadır.

Tıp teşhis örneğindeki nitelikleri sıralanırsa;

- Bazı belirtiler (analiz sonuçları) aynı olmaktadır,
- Bazı belirtiler sıçrayışıyla değişir ve onların değerleri gözlem periyoduna veya dış koşullara bağlıdır,
- Bir sıra belirtilere erişmek mümkün olmuyor veya onların elde edilmesi çok büyük masraflara bağlıdır,
- Belirtiler arasında tüm ilişkiler belli olmayabilir [25].

3.3. Bilginin Elde Edilmesi

Bilgi sistemlerinin kurulması aşamasında gereken bilginin elde edilmesi ve onun bilgi tabanına yüklenmesi çok önemli problemlerdendir. Bu aşama problemi mümkün olduğu kadarı ile maksimal sayıda alt problemlere bölmekle başlar. Bu bilgilerden çıkarım yapılabilmesi için bilgi mühendisi uzmanla birlikte çalışmalıdır. Bu aşamada uzmandan bölmenin nasıl yapılacağı ve problemi çözerken hangi yöntemlerden yararlandığı öğrenilmelidir. Uzman, çalıştığı disiplin

alanına uygun bir profesyonel dili kullanır. Fakat bu dil problemin çözülmesi yöntemini tasvir etmek için, çok da yeterli olmayabilir. Buna bakmayarak uzmanı, bilgiyi sunma modellerini, uzman sistemlerini ve yapay zekâ elemanlarını öğrenmeye mecbur etmek gerekmemektedir. Problemi uygun disiplin alanında çözebilmeyi ve nasıl çözdüğünü anlatabilmeyi uzmandan talep edilmektedir.

Bilginin elde edilmesi süreci lineerden daha çok döngülü karakter taşır ve uzman sistemlerin tasarlanmasında yürütülen başka işlemlere göre daha kötü yapılanmıştır. Önce mühendis, uzmandan disiplin alanının herhangi bir kısmı hakkında veya bir alanın nitelikleri hakkında mümkün olduğu kadar çok bilgi alır. Sonra bu bilgiler analiz edilir ve uzlaşmazlıklar var ise bu açıklığa kavuşturulur. Sonra bilgiler genişletilir ve modifikasyona uğratılır. Daha sonra yine uyuşmazlıkların olup olmadığına bakılır. Böylece uzman sistemin imkânları aşama aşama artırılır.

Uzmanla çalıştığında bilgi mühendisinin diyalogdan “Eğer- O Halde” gibi ifadeleri çıkarması gerekir. Çünkü bunların uygulanması mühendisin problem örnekleri kurduğunu, daha sonra ise uzmanın verdiği çözümü depoladığını göstermektedir. Bu usul yüzeysel bilgi almaya olanak verir. Disiplin alanının modelinin kurulması esasına dayanan yöntem daha verimli sayılır. Eğer bu, tasarımcının ilk işi ise, o zaman açık olan problem seçmek gerekir. Ayrıca önceden hazırlanmış benzer bilgi sistemleri incelenmeli, özellikle onların değeri ve zamanı göz önüne alınmalıdır.

Ham bilginin elde edilmesi için farklı yaklaşımlar kullanılmaktadır:

1. Söyleşi veya soru-cevap yöntemi;
2. Gözlem veya izleme yöntemi;
3. Anket yöntemi;
4. Uzman ile birlikte çalışma;
5. Protokol analizi

En yaygın olarak kullanılan söyleşi yönteminde bilgi, uzmanın bilgi mühendisi ile karşılıklı konuşmasıyla ve konu üzerine sorular sormakla cevaplarından yola çıkarak elde edilir. Bu yöntemin kötü tarafı, uzmanın her zaman kendi bilgilerini açık olarak anlatamaması ve bu bilgilerin sınırlarının ortaya koymasının zor olmasıdır. Bu yöntemin dezavantajını aşabilmek için kullanılan gözlem yönteminde, bilgi mühendisi uzmanı işini yaparken onu yakından gözlemlemekte ve elde ettiği bilgilerin bir tür onaylamasını yapmaktadır. Anket yönteminde, bilgi mühendisi önceden tutarlı ve detaylı bir anket hazırlayarak onu uzmanın cevaplamasını istemektedir. Bu cevaba göre edinilen bilgiler daha sonra diğer anketler yardımıyla daha da detaylaştırılabilir. Uzman ile birlikte çalışma yönteminde uzman, disiplin alanındaki bir problemi ele alarak baştan sona bilgi mühendisi ile birlikte sonuca ulaşınca kadar çözer ve

detayları ona anlatır. Daha sonra problemin değişik varyasyonları üzerinde bu çalışma devam eder. Protokol analizi yöntemi birden fazla uzman görüşü oldukça kullanılan bir yöntemdir. Aynı konunun uzmanı olsala bile değişik sorunların çözümü için değişik fikir üretebilirler. Bazen bu fikirler birbirine ters düşmektedir. Bu durumda ya uzmanların çoğunluğunun görüşü tercih ediliyor ya da onlardan en çok güvenilen birine üstünlük verilir. Fakat bu, sonucun daha iyi olduğunu göstermeyebilir. İşin içinden çıkmak için uzmanlarla ayrı ayrı çalışmak, bunun için uzmanlarla bilginin boyutlarını gösterebilecek protokol kartları dağıtmakta ve onların birbirinden habersiz bu kartları doldurması istenmektedir. Elde edilen bilgilerin literatür bilgilerine ters düşmemesi, bilinen gerçekler ile bağdaşması gibi kriterleri göz önüne alarak bilgi mühendisi bir sonuca varmalıdır. Elbette burada bilgi mühendisinin deneyimi ve yeteneği çok önemli olmaktadır. Son yıllarda otomatik bilgi edinme araçları geliştirilmiş ve bu araçlar olumlu sonuçlar vermeye başlamıştır [25].

3.4. Bilgi Tabanlarında Bilginin Tamlığının ve Zıddiyetsizliğinin Kontrol Edilmesi

Bilginin sunulmasının temel yöntemlerinden biri kurallardır. Kuralların bu veya diğer bir şekilde depolanmış olduğu yer kurallar tabanı olarak karşımıza çıkmaktadır. Bu kuralların birbiri ile uyum içinde ve çelişkisiz, aynı zamanda tüm istenen bilginin bu veya diğer bir şekilde bulunabilir olması kurallar tabanının istenilen özelliğidir. Kurallar tabanında uzman kişi tarafından belirlenen ve bilgi mühendisi tarafından biçimlendirilen bilginin tam olması ve birbirine zıt olmamasının açığa çıkarılması ise çoğu zaman kolay bir iş değildir. Bilgi mühendisinin görevlerinden biri bunları yoklayarak uzman sistemin bilgi tabanının doğru biçimlendirilmiş olmasını tespit etmektedir. Bunu yapabilmek için çözüm tablosu dili olarak adlandırılan yöntemden kullanılabilir. Çözüm tablosu dili bilgi tabanındaki kuralların tam ve zıddiyetsiz olup olmadığını otomatik bir biçimde bilgisayar yardımı ile yapmaya ve bu kuralları optimumlaştırmaya imkân veren algoritmik ve prosedürel olmayan, incelenmekte olan süreci görsel bir şekilde sunmaya olanak tanıyan bir dildir. Bu dilin temel bileşeni olan çözüm tablosu kullanıcı ile bilgisayar arasındaki ilişki aracı, soru listesi oluşturmak için, tasarlama ve kontrolün otomatik hale getirilmesi, taklit modellemede, programların analiz, sentez ve optimal hale getirilmesinde vb. uygulamalarda geniş ölçüde kullanılmaktadır.

Çözüm tablosu (ÇT) incelenmekte olan problem alanındaki durumu belirleyen sonlu koşullar kümesinin eleman değerleri ile bu alandaki sonuçları belirleyen sonlu çözümler kümesinin elemanları arasındaki uygunluğu göstermektedir. Başka bir deyimle, ÇT “eğer-o halde” sonlu üretim kurallarının tablo şeklinde yazılmasıdır. Genelde kurallardan oluşan bir bilgi tabanı birbiri ile hiyerarşi şeklinde bağlı olan birkaç ÇT’den oluşabilmektedir. Eğer ÇT’deki koşul yalnız iki değer alırsa, böyle tabloya sınırlandırılmış girişli Çözüm tablosu,

ikiden çok değer alırsa bu tabloya genişletilmiş girişli ÇT denmektedir. Her bir genişletilmiş girişli ÇT ona eşdeğer olan sınırlandırılmış girişli ÇT'ye getirebildiği için burada sunulması daha sıkışık olan sınırlandırılmış girişli ÇT'ye bakılmaktadır. Bazı eklerle birlikte ÇT'nin şekli Tablo 3.1'de gösterilmektedir. Burada ÇT $C=\{C_i\}$, $i=1,...,m$ koşullar (eğer) kümesindeki koşulları sunan veriler; $A=\{A_r\}$, $r=1,...,k$ eylemler (o halde) kümesindeki sonuçları sunan veriler; $R=\{R_j\}$, $j=1,...,n$ kurallar kümesi; $C_{ij}=\{0,1\}$; E aksi durumdaki (Else) durum; $f=(f_i)$, $i=1,...,n$ vektörü çözüm kurallarının (R_j) yerine getirilmesi ihtimali; f_E kuralın “aksi durumda” yerine getirilmesi ihtimali; $t=(t_i)$, $i=1,...,m$, ve $q=(q_r)$, $r=1,...,k$ vektörleri uygun olarak koşulların değerlerinin hesaplanması ve eylemlerin yerine getirilmesi karmaşıklığını belirlemektedirler. Genel olarak t_i ve q_r , uygun fonksiyonların yerine getirilmesinin sürelerini karakterize etmektedirler. f , t ve q vektörlerinin değerleri belli olmayabilir veya verilemeyebilir, bu durumda bütün çözümlerin aynı ihtimalli olması kabul edilir. Matematiksel olarak ÇT aşağıdaki gibi belirlenir:

$$T=<(C,A,\check{C},\check{A}),(f,t,q)>$$

Tablo3.1. Çözüm tablosunun genel yapısı

Koşullar C_i	Çözüm Kuralları $R_1 \ R_2 \ \dots \ R_n$	“Aksi durum” kuralı E (Else)	Koşulların karmaşıklığı T_i
C_1	$c_{11} \ c_{12} \ \dots \ c_{1n}$		t_1
C_2	$c_{21} \ c_{22} \ \dots \ c_{2n}$		t_2
$\cdot$	$\cdot \ \cdot \ \cdot$		$\cdot$
$\cdot$	$\cdot \ \cdot \ \cdot$		$\cdot$
C_m	$c_{m1} \ c_{m2} \ \dots \ c_{mn}$		t_m
Eylemler A_i			Eylemlerin karmaşıklığı
A_1	$a_{11} \ a_{12} \ \dots \ a_{1n}$	a_{1n+1}	q_1
A_2	$a_{21} \ a_{22} \ \dots \ a_{2n}$	a_{2n+1}	q_2
$\cdot$	$\cdot \ \cdot \ \cdot$	$\cdot$	$\cdot$
$\cdot$	$\cdot \ \cdot \ \cdot$	$\cdot$	$\cdot$
A_i	$A_{k1} \ a_{k2} \ \dots \ a_{kn}$	a_{kn+1}	q_k
Kuralların sıklığı	$f_1 \ f_2 \ \dots \ f_n$	F_E	

$R_j=<\check{C}_j,\check{R}_j>$ $j=1,...,n$ çifti çözüm kuralı (R_j kuralı) olarak adlandırılır ve tablodaki j. Sütun-vektöre denk gelir. Tablodaki c_{ij} ve a_{ij} elemanlarının değerleri aşağıdaki gibi belirlenirler.

$$c_{ij} = \begin{cases} C \in \hat{C}_i, \text{ eğer } R_i \text{ kuralı için } C_i \text{ koşulu } c \text{ ise;} \\ \lambda, \text{ eğer } R_j \text{ kuralı için } C_i \text{ koşulu önemsiz ise.} \end{cases}$$

$$a_{ij} = \begin{cases} a \in \{1,2,\dots,k\}, \text{ eğer } R_i \text{ kuralı için } C_i \text{ eylemi yürütülüyor} \\ \quad (E, \text{ eğer } j=n+1) \text{ ve önceliği var ise veya} \\ \quad A, \text{ eğer yürütme sıra avantajı var ise;} \\ 0, \text{ eğer } R_i \text{ kuralı için } A_i \text{ eylemi yürütülmüyor ise.} \end{cases}$$

Eğer ÇT’de “1” ve “0”lardan oluşan bütün C_j değerleri mümkün tüm Boole kombinasyonlarını oluşturmakta ise, o zaman ÇT tamdır, aksi durumda ÇT tam değildir denilir. Sonuncu durum bazı C_i kombinasyonları için herhangi bir eylemin düşünülmediği durumda ortaya çıkar, yani sistemde bazı kombinasyonlar oluştuğunda cevap reaksiyonu olarak sistemin ne yapacağı bilgi mühendisi tarafından belirlenmemiştir. Eğer A_i eylemi için C_j ’nin hem kendisi ve hem de tersi değerleri geçerli ise bu durumda ÇT zıddiyetlidir denmektedir. ÇT tam olmayabilir, ama bu onun zıddiyetli olduğu anlamına gelmez. Tam olmayan ÇTde öngörülmemiş kombinasyonların ortaya çıkma ihtimali sıfır kabul edilebilir ya da bu durumlar hakkında bilgi mühendisi yeniden uzmana danışarak karar almalıdır [25].

BÖLÜM 4

4. GELİŞTİRİLEN HASTALIK TEŞHİS SİSTEMLERİ

Biyokimya, canlı yapı taşlarını ve canlılık olaylarını molekül düzeyinde inceleyen temel bir bilim dalıdır. Klinik Biyokimya ise, sağlıkta ve hastalıktaki biyokimyasal mekanizmaları tanı, ayırıcı tanı ve prognoz tayinindeki testlerin seçimi, tekniği, sonuçların yorumlanması ve klinisyenlere konsültasyonu da içeren, tıba özgün bir laboratuvar bilimidir [26].

Bu bilim dalında yapılabilen analizler aşağıda verilmiştir:

A) Kuruluşlarına göre klinik biyokimyasal analizler

- 1) Kalitatif (nitel) analizler
- 2) Kantitatif (nicel) analizler

B) İncelenecek maddelere göre klinik biyokimyasal analizler

- 1) Metabolitlerin miktar tayini
- 2) Enzimlerin aktivitelerinin tayini
- 3) Vitaminlerin miktar tayini
- 4) Hormonların miktar tayini
- 5) Yükleme testleri

C) Kullanılan materyal miktarına göre klinik biyokimyasal analizler

- 1) Makrometodlar
- 2) Semi-mikrometodlar
- 3) Mikrometodlar
- 4) Ultramikro metodlar (mikrolitreler metodları)
- 5) İleri ultramikro metodlar (nanolitireler metodları)

D) Kullanılan ayıraçlara göre klinik biyokimyasal analizler

- 1) Hazırlanacak ayıraçlarla yapılanlar
- 2) Hazır ayıraçlarla ve kitlerle yapılanlar
- 3) Ayıraçlı şeritler ya da daldırma çubukları ile yapılanlar

E) Elle işleme ve otomatlaşma gereksinimine göre klinik biyokimyasal analizler

- 1) Manuel analizler. Elle işlenen metodlar
- 2) Otomatlaştırılmış analizler. Otoanalizör gerektiren metodlar

F) Kullanılışlarına göre klinik biyokimyasal analizler

- 1) Sık kullanılan analizler
- 2) Acil analizler. Hemoglobin, kan grubu, hematokrit, glukoz, üre, kreatinin, sodyum, potasyum, klorür, kalsiyum, bilirubin, kan gazları, AST, ALT, CK, CK-MB, Amilaz analizi

Biyolojik Sıvılar:

- Kan, BOS, amniyon sıvısı, idrar, seminal sıvı
- Eritrositler
- Lökositler
- Doku biyopsi örnekleri
- Doku hücre kültürleri şeklinde sıralanabilir.

İnsanlarda görülen hastalıkların tanı veya ayırıcı tanısının yapılması ve sağaltımın izlenmesinde enzimatik ölçümlerin uygulanması ile ilgilenen bilim dalına Klinik Enzimoloji denir [26].

- Kan, idrar, beyin-omurilik sıvısı (BOS, serebrospinal sıvı), amniyon sıvısı, mide özsuğu, sperma
- Plevra sıvısı, periton sıvısı, eklem sıvısı (sinovyal sıvı)
- Ovaryum kisti, hidatik kist gibi kist sıvıları ve çeşitli fistüllerden sızan sıvılar

Kan enzimlerinin aktivite tayinlerinde dikkat edilecekler:

- Kan, antikoagülsüz tüpe (düz tüp) alınmalıdır
- Kan genellikle venden alınır
- Kan alırken hemolizden kaçınmalıdır
- Kan, pıhtılaşmasından hemen sonra santrifüj edilerek serum ayrılmalıdır
- Günlük taze kan kullanılması en iyisidir

Klinik tanıda önemli olan serum enzimleri:

- Transaminazlar (AST Ve ALT)
- Laktat Dehidrojenaz (LDH, LD)
- Kreatin Kinaz (CK, CKMB Ve CPK)
- Fosfatazlar (ALP Ve ACP)
- Amilaz (AMS)
- Lipaz (LPS)
- Gama Glutamiltransferaz (GGT, γ -GT)
- Aldolaz (ALS)
- 5'-Nükleotidaz (5'-NT)
- Lösin Aminopeptidaz (LAP)
- Psödokolinesteraz (Che)
- Glukoz-6-Fosfat Dehidrojenaz (G-6-PD) [26]

Kan, kapalı bir kanallar sistemi içinde dolaşan sıvıdır. İçeriğı:

- Plazma → %55-60
- Şekli Elemanlar:

- *Eritrositler*
- *Lökositler*
- *Trombositler*
- Hematokrit

Kanın Görevleri:

- Besin maddelerini taşıma
- Solunum gazlarının alışverişi ve taşınması
- Metabolizmanın atık ürünlerini taşıma
- Enzim, hormon, vitaminleri taşıma
- Organizmayı savunma ve koruma
- Elektrolit, su, asit-baz dengesini düzenleme
- Vücudun ısınıı düzenleme şeklinde sıralanabilir [27].

Kan Plazması ve Serumı: Kan plazması, kanın sıvı kısmıdır. Fibrinojen ve diğer bazı pıhtılaşma faktörlerini içerir. Pıhtılaşmış kanın santrifüj edilmesiyle elde edilen sıvı üst faz serumdur. Fibrinojen ve diğer bazı pıhtılaşma faktörlerini içermez. Kan plazmasının %90'ını su ve %10'unu suda çözünmüş katı maddeler oluşturur [27].

Hastalık teşhis sistemleri, son derece yoğun ve değerli bilgilerin biriktirildiği ortamlarda, toplanan verilerin etkin ve çabuk biçimde süzgeçten geçirilip faydalı ve kolay kullanılabilir bilgiler haline getirilmesini sağlar [28].

Bu tez çalışmasında, kan biyokimya parametreleri ile karar verilebilecek hastalık teşhisinde, hekime yardımcı olacak ve kolaylık sağlayabilecek bir karar destek sistemi oluşturulmuştur. Örüntü tanıma süreci esas alınmış olup, sistemin işleyişi veri madenciliği tekniklerinden olan karar ağaçları yapısı ile sağlanmaktadır. Sisteme giriş olarak, biyokimya parametrelerinden belirlenen hastalıklar için temel belirleyici olan lipidler veya enzimler kullanılarak değerlendirmelerde bulunulmuştur.

Bu enzimler her bir kişinin tanı ve teşhisinde kullanılırken büyük veri yığınları oluşturur. Ayrıca biyoloji ve tıp dünyasındaki bu tip araştırmalar arttıkça gerçek dünya olgularıyla ilgili farklı veri tiplerinin çokluğuna yol açmaktadır [29]. Modern tıpta bu yüzlerce özellik, binlerce görüntüyü insanların kavrayıp anlamlı bilgiler haline getirmesini oldukça zorlaştırmıştır [30, 31]. Halen kullanılan ana veri tiplerinin bazıları DNA [32, 33] ve protein sentezi [34], biyolojik durum ölçümleri [35], grafikler [36] ve enzimlerdir [29]. Klinik tıpta bulguların okunması, sadeleştirilmesi, sınıflandırılması ve karara varılması işlemlerinde tıp uzmanlarının kılavuz hazırlaması işi, bu veriler arttıkça sıkıcı ve zor bir iş haline gelmiştir. Ayrıca bu veriler içerisinde saklı kalan pek çok bulgu, bu şekilde bir veri yığını olarak kalmıştır.

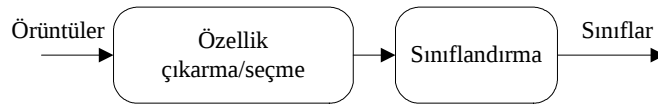
Faydalı bilgiler çıkarmak için verilerin otomatikleştirilmesi gerekmektedir. Bu sorunların çözümü için en çok kullanılan yaklaşım tarzı veri tabanlarından bilgi elde edilmesi veya veri madenciliğidir [22]. Veri madenciliği, hem metine bağlı bilgileri hem de belirli bir hastanın görüntü verilerinin çıkarılıp her bir hasta kaydını birleştirmek için kullanılır [37]. Bu anlamda veri madenciliği tıpta pek çok alanda başarı ile uygulanmıştır [22,38]. Örneğin, kronik hastalıkların teşhis ve tedavisinde kesin kurallar keşfetmek, seçilen kuralları kullanmak, hastalıkla ilgili üretilen, organize edilen, arıtılan ve paylaşılan bilgiyi bir sistem içerisinde çalıştırma işlemlerini genişletme durumlarını tanımlamak için tamamlayıcı veri madenciliği oluşturulması [39], aynı şekilde hipoplastik sol kalp sendromunda görülen bebek ölümleri için bir veri kazanım sistemi tasarlanarak toplanan verilerle anlık tahmin yapılması ve böylece akıllı sistemlerle keşfedilen bilgi ile bebek ölümlerinin kontrol altına alınarak iyileştirilmesine [40] yönelik çalışmalara literatürde rastlanmaktadır. Bu çalışmalarda, amaca ulaşmak için ilk önce veriler, veri tabanında depolanıp, işlenerek anlamlı bilgilere ulaşılmaya çalışılmıştır [29]. Medikal alanda, veri ambarları klinik ambar olarak da bilinmektedir. Sonra, biyolojik, klinik ve yönetsel veriyi içeren klinik ambarlar hasta bilgilerini birleştirir. Böylece hasta ile ilgili sistemlerin kullanım imkânı artırılmış olur [41].

Örüntü Tanıma: Örüntü, ilgilenilen varlıklar ile ilgili gözlenebilir veya ölçülebilir bilgilere verilen addır. Gerçek dünyadaki bu örüntüler, genellikle ilgilenilen verilerin nicel tanımlama şekilleridir. Örüntü tanıma, insanların çeşitli ses, imge ve benzeri tüm örüntülerin biçimsel şekillerinden çıkardıkları dilsel şekillendirmedir. Aslında, örüntü tanıma bilimin, mühendisliğin ve günlük hayatın geniş bir alanındaki etkinlikleri kapsamaktadır. Örüntü tanıma uygulamalarını, insanların yaşantısında da görebiliriz: Hava değişimin algılanması, binlerce çiçek, bitki, hayvan türünü tanımlama, kitap okuma, yüz ve ses tanıma gibi bulanık sınırlara sahip birçok etkinlikte örüntü tanıma kullanılır. İnsan örüntü tanınması, geçmiş tecrübelerle dayalı öğrenme esaslıdır. Böylece, insanlar pratikte karşılaştığı örüntü tanıma olaylarını tecrübeleri ışığında değerlendirebilme yeteneğine sahiptirler. Belirli bir sesi tanımak için kullanılan kuralları tanımlamak mümkün değildir. İnsanlar bu işlemlerin birçoğunu oldukça iyi yapmalarına rağmen, bu işlemleri daha ucuz, iyi, hızlı ve otomatik olarak makinelerin yapmasını arzularlar. Örüntü tanıma, böyle akıllı ve öğrenebilen makineleri gerçekleştirmek için, çok boyutlu bir mühendislik disiplindir.

Örüntü tanıma olayını şu şekilde irdeleyebiliriz: Aralarında ortak özellik bulunan ve aralarında bir ilişki kurulabilen karmaşık işaret örneklerini veya imgeleri, bazı tespit edilmiş özellikler veya karakterler vasıtası ile tanımlama veya sınıflandırmadır. Bu bağlamda, örüntü

tanımanın en önemli amaçları; bilinmeyen örüntü sınıflarına belirli bir şekil vermek ve bilinen bir sınıfa ait olan örüntüyü teşhis etmektir [42].

Örüntü tanıma tekniklerinin uygulamaları birçok mühendislik, tıp, askeri ve bilim alanına açıktır [43]. Bunlardan bazıları; ses tanıma [44], EEG sınıflama, DTMF haberleşme işaretlerini tanıma ve radar hedef sınıflama [45], biyomedikal kontrol verilebilir. Örüntü tanıma olarak bilinen bu uygulamalar, makine öğrenmesi, örüntü sınıflandırma, ayırım analizi ve nitelik tahmini gibi isimlerle de anılmaktadır [46]. Kullanılan bu genel model şekil 4.1’de gösterilmiştir [47].



Şekil 4.1. Örüntü tanıma yapısı.

4.1. Kardiyak Enzim Parametrelerinden Kalp Krizi Teşhisi

Miyokard İnfarktüsü (MI), miyokardın oksijen ihtiyacının koronerler tarafından karşılanamaması sonucu kalp kasında oluşan nekrozun tanımlanmasında kullanılan bir terimdir. MI halen dünyada mortalite ve morbidite nedenleri içinde ilk sıralarda yer alan hastalıklardan biridir. MI tanısı; klinik semptomlar, elektrokardiografi (EKG) değişiklikleri ve kardiyak enzimlerin serum düzeylerindeki artış esaslarına dayanır. Bazı enzimler, serum değerlerinin yükseldiğinin tespit edilmesiyle MI tanısı doğrulanır. Böyle enzimler, akut MI (AMI) ’nün diagnostik laboratuvar belirleyicileri kardiyak enzimleri olarak adlandırılırlar [48]. Kardiyak enzimleri belirlemek için biyokimya laboratuvar test sonuçları kullanılır. Klinik Biyokimya, sağlıkta ve hastalıktaki biyokimyasal mekanizmaları tanı, ayırıcı tanı ve prognoz tayinindeki testlerin seçimi, tekniği, sonuçların yorumlanması ve klinisyenlere konsültasyonunu da içeren, tıbbi özgün bir laboratuvar bilimidir [26]. Kural olarak geniş infarktlar, daha yüksek serum enzim düzeylerine neden olur. Yükselen enzim düzeyi, infarktın zamanlaması açısından da yardımcı olur. MI teşhisinde dikkate alınan enzimler CK, CK-MB, LDH, AST, ALT, Troponin ve Miyoglobindir [49-51].

Bu bölümde, geliştirilen MI teşhis karar destek sisteminin yapısının kavranmasına yönelik bilgiler alt bölümler halinde aşağıda anlatılmaktadır.

4.1.1. Kardiyak Enzimler

MI, kalbi besleyen koroner damarların çok defa bir pıhtı ile tıkanması sonucu oluşur. Böyle bir durumda tıkanmış olan koroner damarların beslediği kalp kası oksijensiz kaldığı için kalbin o bölgesinde bir harabiyet meydana gelir. Halk arasında kalp krizi olarak adlandırılan bu olay sırasında hastalardan üçte biri ani ölümle hayatını kaybetmektedir. Tıkanmış olan koroner damarın 2-4 saat içinde yapılacak girişimlerle yeniden açılması halinde kalp krizleri, enfarktüs oluşturmada atlatılabilir [52]. MI tanısı konulurken; klinik semptomlar, elektrokardiografi (EKG) değişiklikler ve kardiyak enzimlerin serum düzeylerindeki artışları göz önüne alınır. Genellikle EKG sonuçları ve biyokimya testlerinde bazı enzimler serum değerlerinin yükseldiğinin tespit edilmesiyle MI tanısı doğrulanır [53]. Biyokimya verileri ile MI teşhisinde baz alınan CK, CKMB, AST, ALT ve LDH enzimleri kardiyak enzimleri olarak tanımlanmaktadır. Bu parametreler aşağıda kısaca açıklanmıştır:

CK: Kalp kası, iskelet kası ve beyin hasarlarında kanda artar. Başlıca 4 izoenzimden oluşur. Bunlar; CK-MM (İskelet kası kökenli CK), CK-BB (Beyin kökenli CK), CK-MB (Miyokart kökenli CK), CK-MiMi (Mitokondri kökenli CK) dır. Total CK'nın yükselme gösterdiği başlıca durumlar; miyokart infarktüsü, miyokarditler, i.m. enjeksiyonlar, ameliyatlara, travmalar, epileptik ataklar, arteriyel emboli ve oklüzyonlar, reanimasyon, kas hastalıkları, alkolizm ve deliryum tremens, intoksikasyonlar, nekrozitan pankreatitler, akut karaciğer nekrozu, malign hastalıklar, endokrin miyopatiler (hipotirodillilerin ve hipoparatiroidillilerin bir kısmında ve seyrek olarak Addison hastalığındaki gibi), Trichinosis, Coxackie B enfeksiyonları, miyokart kontüzyonu gibi durumlardır [24]. CK, MI tanısının konmasında en değerli enzimdir. Ağrının başlamasından ~ 6-8 saat sonra yükselmeye başlar, 20-24 saatte pik yapar ve 48 saatte normale döner [53].

CK-MB: Total CK'nın %5-20 sini oluşturur. Eğer total CK'nın %5'inden azını oluşturuyorsa yüksek olsa bile bu durumun kalpten değil diğer organlardan ve genellikle de iskelet kasından kaynaklanan bir durum olduğu sonucu çıkarılır. CK-MB düzeyi total CK'nın %20'sinden yüksek olduğunda da bunun izoenzim CK-BB (malignite veya nörolojik hastalık) veya Makro-CK kökenli bir yükselmeden olup olmadığı araştırılmalıdır (Makro CK-1: yaşlı insanların -çoğunlukla kadınların- yaklaşık %1-1.5'inde CK-BB ile IgG arasında oluşan immun kompleksler olup herhangi bir hastalık bulgusu olmayabilir. Makro CK-2: Birden çok CKMiMi molekülünün bir araya gelmesiyle oluşur. Nekrozitan karaciğer hastalıkları ve malign tümörler gibi hastalıklarda artarlar). Bu nedenle CK-MB düzeyi toplam CK'nın %20'sinden yüksekse Troponin T veya I bakılması endikedir [54]. CKMB, CK'ın izoenzimi olan MB fraksiyonu miyokardiyal hücrelere çok daha spesifiktir. CK'ın yükselmesinden, hafif daha önce yükselmeye başlar ve 48-72 saat içinde normale düşer [53].

AST, ALT: Başta karaciğer, kalp ve iskelet kası hastalıklarında yükselebilir. CK/SGOT oranı 10'dan düşükse MI lehine, 10'dan büyükse kas hastalığına veya harabiyeti lehine olarak yorumlanır. Karaciğer hastalıkları veya stazına bağlı SGOT yükselmelerinde ise diğer karaciğer enzimlerinde de yükselme olması beklenir (Örneğin SGPT'de) [24]. AST ve ALT, Plazmada pik yapması oldukça yavaştır (24-48 saatte) ve CK'dan daha uzun süre saptanır. 72 saatte normale inebilir.

LDH: Nonspesifik ama infarktüsün geçirildiği zamanın belirlenmesinde yardımcı olabilen bir parametredir. 5 izoenzimi vardır. LDH-1 (=a-HBDH) kalp kası ve eritrositlerde bulunur [24]. LDH, 3-4 günde pik yapar ve 7-14 gün kadar yüksek seyredebilir. Sıklıkla geçmiş MI tanısında kullanılır. CK negatif ise ve MI'nın en az 2 gün önce meydana geldiği düşünülüyorsa LDH yararlıdır [53].

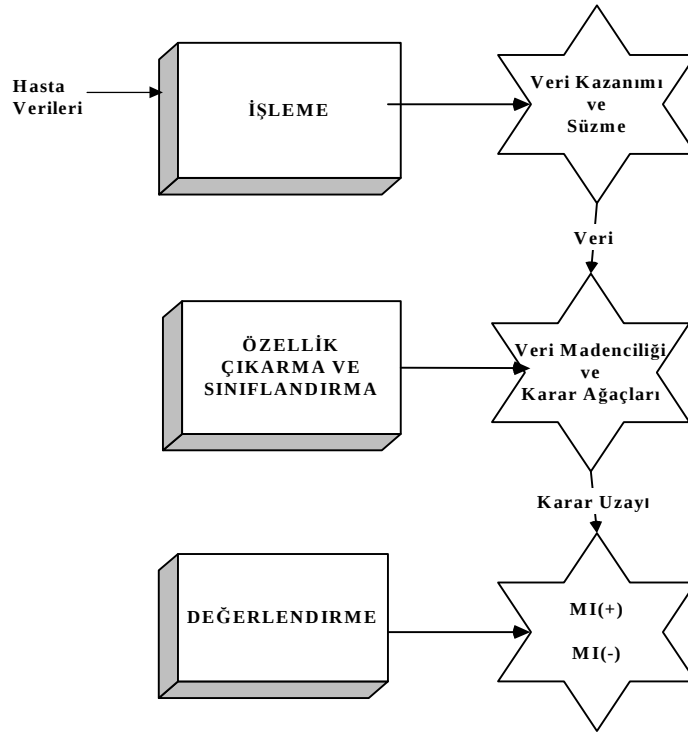
Bu enzimlerin MI geçiren hastalarda yükselmenin zamanlaması tablo 4.1 de verilmiştir [54].

Tablo 4.1. Kardiyak Enzimlerin yükselme zamanlaması

Enzim	Yükselme süresi (saat)	Maksimum düzeye ulaşma (saat)	Normale dönme süresi (gün)
Total CK	4-8	16-36	2-3
CK-MB	4-8	12-18	3-6
AST	4-8	16-48	3-6
ALT	4-8	16-48	3-6
LDH	6-8	24-60	7-15

4.1.2 Yöntem

Hekim açısından elde mevcut olabilecek başlıca tanı koydurucu testler; ECG ve biyokimya testleridir ki; bunlar birçok durumda tanı koydurucudur. Biyokimya test sonuçlarında MI teşhisi için dikkate alınan kas enzim parametreleri veya diğer bir adıyla kardiyak enzimlerinin değerleri kontrol edilerek MI teşhisine varılabilir. Geliştirilen karar destek sisteminin yapısı, yukarıda anlatılan örüntü tanıma tabanlı olarak Şekil 4.2 de verilmiştir [47]. Bu sistemin işleyişi şu adımlardan oluşmaktadır:



Şekil 4.2. Karar destek sistemi algoritması.

Adım – 1: Ön işlemler

Verilerin toplanması, amaca uygun olanların seçildiği aşamadır. Kardiyoloji servisine başvuran hastaların biyokimya test sonuçları doktor gözetiminde toplanarak ve bu verilerden kas enzimleri ile ilgili olanlar ayrıştırılarak gerekli veriler elde edildi. Toplanan verilerden, doktor yönlendirmesi ile MI şüphesi olan hastaların test sonuçları seçilerek veri tabanı oluşturuldu.

Adım – 2: Özellik çıkarma ve sınıflandırma

Derlenen veritabanındaki verileri işlemeye yönelik olarak geliştirilen karar ağacı tabanlı karar destek sisteminin en önemli aşamasıdır. Toplanan veriler bu aşamada işlenmektedir. Bunun için Tablo 4.2. de belirtilen kurallar doktor bilgilerine göre belirlenerek, Şekil 4.3 deki karar ağacı yapısı elde edilmiştir.

Tablo 4.2. Sistemin kural yapısı

Kurallar	Karar Ağacı Sırası
<i>if (ck>= 195) and (ckmb>=25) and (ldh>=450) and (ast>=40) and (alt>=40) then</i>	(a)
<i>else if (ck>= 195) and (ckmb>=25) and (ldh>=450) and (ast>=40) and (alt<40) then</i>	(b)
<i>else if (ck>= 195) and (ckmb>=25) and (ldh>=450) and (ast<40) and (alt>=40) then</i>	(b)
<i>else if (ck>= 195) and (ckmb>=25) and (ldh>=450) and (ast<40) and (alt<40) then</i>	(c)
<i>else if (ck>= 195) and (ckmb>=25) and (ldh<450) and (ast>=40) and (alt>=40) then</i>	(d)
<i>else if (ck>= 195) and (ckmb>=25) and (ldh<450) and (ast<40) and (alt>=40) then</i>	(e)
<i>else if (ck>= 195) and (ckmb>=25) and (ldh<450) and (ast>=40) and (alt<40) then</i>	(e)
<i>else if (ck>= 195) and (ckmb>=25) and (ldh<450) and (ast<40) and (alt<40) then</i>	(f)
<i>else if (ck>= 195) and (ckmb<25) and (ldh>=450) and (ast>=40) and (alt>=40) then</i>	(g)
<i>else if (ck>= 195) and (ckmb<25) and (ldh>=450) and (ast<40) and (alt>=40) then</i>	(h)
<i>else if (ck>= 195) and (ckmb<25) and (ldh>=450) and (ast>=40) and (alt<40) then</i>	(h)
<i>else if (ck>= 195) and (ckmb<25) and (ldh>=450) and (ast<40) and (alt<40) then</i>	(i)
<i>else if (ck>= 195) and (ckmb<25) and (ldh<450) and (ast>=40) and (alt>=40) then</i>	(j)
<i>else if (ck>= 195) and (ckmb<25) and (ldh<450) and (ast<40) and (alt>=40) then</i>	(k)
<i>else if (ck>= 195) and (ckmb<25) and (ldh<450) and (ast>=40) and (alt<40) then</i>	(k)
<i>else if (ck>= 195) and (ckmb<25) and (ldh<450) and (ast<40) and (alt<40) then</i>	(l)
<i>else if (ck< 195) and (ckmb>=25) and (ldh>=450) and (ast>=40) and (alt>=40) then</i>	(m)
<i>else if (ck< 195) and (ckmb>=25) and (ldh>=450) and (ast>=40) and (alt<40) then</i>	(n)
<i>else if (ck< 195) and (ckmb>=25) and (ldh>=450) and (ast<40) and (alt>=40) then</i>	(n)
<i>else if (ck< 195) and (ckmb>=25) and (ldh>=450) and (ast<40) and (alt<40) then</i>	(o)
<i>else if (ck< 195) and (ckmb>=25) and (ldh<450) and (ast>=40) and (alt>=40) then</i>	(p)
<i>else if (ck< 195) and (ckmb>=25) and (ldh<450) and (ast<40) and (alt>=40) then</i>	(q)
<i>else if (ck< 195) and (ckmb>=25) and (ldh<450) and (ast>=40) and (alt<40) then</i>	(q)
<i>else if (ck< 195) and (ckmb>=25) and (ldh<450) and (ast<40) and (alt<40) then</i>	(r)
<i>else if (ck< 195) and (ckmb<25) and (ldh>=450) and (ast>=40) and (alt>=40) then</i>	(s)
<i>else if (ck< 195) and (ckmb<25) and (ldh>=450) and (ast>=40) and (alt<40) then</i>	(t)
<i>else if (ck< 195) and (ckmb<25) and (ldh>=450) and (ast<40) and (alt>=40) then</i>	(t)
<i>else if (ck< 195) and (ckmb<25) and (ldh>=450) and (ast<40) and (alt<40) then</i>	(u)
<i>else if (ck< 195) and (ckmb<25) and (ldh<450) and (ast>=40) and (alt>=40) then</i>	(v)
<i>else if (ck< 195) and (ckmb<25) and (ldh<450) and (ast<40) and (alt>=40) then</i>	(w)
<i>else if (ck< 195) and (ckmb<25) and (ldh<450) and (ast>=40) and (alt<40) then</i>	(w)
<i>else if (ck< 195) and (ckmb<25) and (ldh<450) and (ast<40) and (alt<40) then</i>	(x)

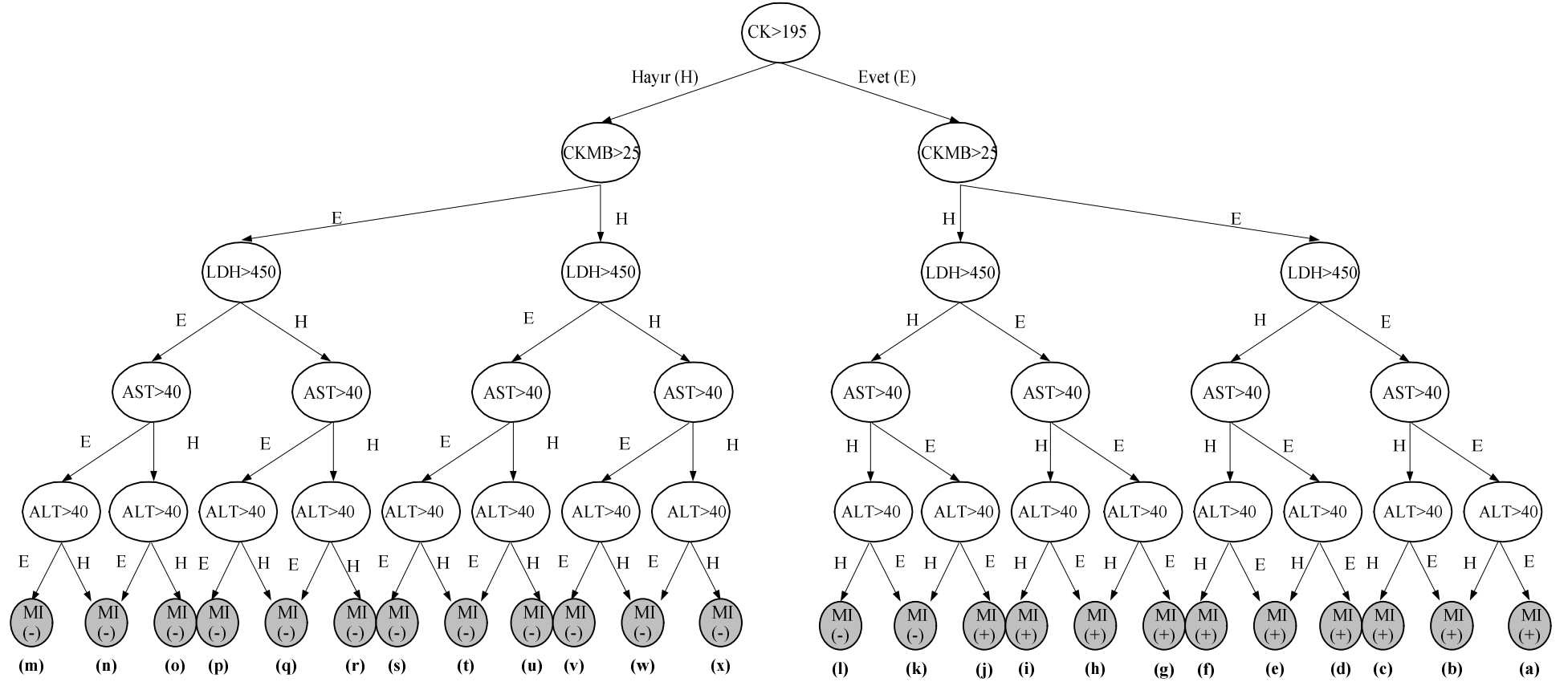
Biyokimya test sonuçlarında sıralanan enzimler arasından kardiyak enzim parametrelerinin normal değerleri Tablo 4.3. de verilmiştir.

Tablo 4.3. Kardiyak enzimlerin normal değerleri

Enzimler	Normal değerleri
CK	24-195
CKMB	0-25
AST	5-40
ALT	5-40
LDH	220-450

Adım – 3: Değerlendirme

Bu aşamada, özellik çıkarma ve sınıflama süreci sonunda oluşan veriler değerlendirilip sunulmaktadır. Bu amaçla tasarlanan karar ağacı yapısının (Şekil 4.3), a,b,c,d,e,f,g,h,i,j yapraklarına ulaşan sonuçlar MI teşhisi ile k,l,m,n,o,p,q,r,s,t,u,v.w.x yapraklarına ulaşan sonuçlar ise MI olarak değerlendirilmemektedir.



Şekil 4.3. Karar Ağacının Yapısı

4.1.3. Uygulama Sonuçları

Geliştirilen karar destek sistemi, 61 hastanın biyokimya test sonuçları üzerinde denenmiştir. Sistemde 61 hastadan 50'sine MI teşhisi, 11 hastaya ise sağlam tanısı konulmuştur ve bu sonuçların uzman bir doktorun verdiği kararlarla örtüştüğü görülmüştür. Bu başarıım yüzdesi, MI teşhisinde doktorun verdiği kararlar, geliştirilen sistemin sonuçlarının bire bir uygunluğu, tasarlanan karar destek sisteminin etkinliğini ve güvenilirliğini açıkça göstermektedir.

Tablo 4.4. Önerilen akıllı karar destek sisteminin başarıımı

	Miyokard İnfarktüsü	
	MI(+)	MI(-)
<i>Verilerin toplam sayısı</i>	50	11
<i>Doğru sınıflananlar</i>	50	11
<i>Hatalı sınıflananlar</i>	---	---
<i>Ortalama tanıma yüzdesi</i>	100	100

4.2. Lipid Parametrelerinden Hiperlipidemi Teşhisi

Hiperlipidemi kardiyovasküler hastalıkların gelişiminde primer ve sekonder korumada değiştirilebilir major risk faktörlerinden biridir [55] ve kanda yağların (kolesterol, trigliseridler ya da her ikisi) normalden daha yüksek düzeylerde olmasıyla meydana gelir [56]. Plazma lipid düzeylerinin ölçülmesi hiperlipidemili kimselerin belirlenmesini sağlar ve böylece erken ateroskleroz tesbitine ve önleyici tedbirlerin alınmasına yardımcı olur. Plazma lipidleri; kolesterol, TG, fosfolipid ve serbest yağ asitlerinden oluşur. Bunlar suda çözünür olmadıklarından kanda proteinle birleşmiş halde yani lipoprotein (Lp) kompleksleri halinde taşınırlar. Kandaki Lp'ler şunlardır: Şilomikronlar, düşük dansiteli Lp'ler (LDL), çok düşük dansiteli Lp'ler (vLDL), yüksek dansiteli Lp'ler (HDL), serbest yağ asitleri ve albumin kompleksleri [57]. Aynı zamanda hiperlipidemisinin medikal tedavisi ile mortalitede düşüşler olduğunu göstermiştir [58]. Protein kaybının olduğu durumlarda plazma lipidlerinde artma olduğu bilinmektedir [59]. Hiperlipidemi saptanan hastalardan hangisinin tedavi edileceği konusunda bugün için ortak bir görüşe varılmıştır. Aterosklerotik kalp hastalığı ve diğer risk faktörleri tedavi kararında göz önünde tutulmalıdır. Statinler, LDL-kolesterol düzeyleri yüksek olan hastalarda ve familial hiperkolesterolemide ilk seçenektir. Pravastatin, simvastatin, fluvastatin, atorvastatin, rosuvastatin ve pitavastatin başlıcalarıdır. Pitavastatin dışındaki tüm statinler intestinal ve/veya hepatik düzeyde ilk geçiş eliminasyonuna uğrar. Statin hücre içine

girdiği zaman biyosentetik kolesterol sentezini düşürür ve böylece hücre içinde kolesterol içeriği azalır. Bu azalmaya yanıt olarak kolesterol homeostazında işlev gören proteinlerin sentezi artar. LDL reseptörlerinin sayısının artması sonucu plazmadan LDL alınımı belirgin şekilde artar. Statinler pleiotropik etkiler de göstererek kardiyovasküler olayların azalmasına katkıda bulunurlar. Endotelial fonksiyonun düzeltilmesi, aterosklerotik plağın stabilitesinin artması, oksidatif stresin azalması ve vasküler inflamasyonun azalmasına katkıda bulunur. Statinler ayrıca koagulatif ve fibrinolitik süreçlerde yararlı rol oynar. Statinler ayrıca endotelial nitrik oksid sentazı (eNOS) stimüle eder. Doku tipi plazminojen aktivatör ekspresyonunu artırarak NO biyoyararlanımını artırır ve endotel disfonksiyonunu düzeltir. Statinlerin en iyi bilinen ve kanıtlanmış olan pleiotropik etkisi hs-CRP seviyelerini azaltmasıdır. CRP ve lipoproteinlerin birlikte değerlendirilmesi aterosklerotik hastalık riskinin belirlenmesi açısından tek başına lipidlerden daha değerlidir. Statinlerin doz-yanıt ilişkisi doğru orantılı değildir. Toksikite ile doz arasındaki ilişki ise doğrusaldır. Statinlerin pek çok yan etkisi olmakla birlikte çoğu hafif ve geçicidir. En sık görülen yan etki myaljidir. Transaminazlarda normalin 3 katı artış olması tedavinin kesilmesini gerektirir. Statinlerin gemfibrozil, fibratlar, nikotinik asid, siklosporin, eritromisin gibi ilaçlarla birlikte kullanılması rabdomyoliz ve hepatotoksikite riskini artırır [55]. Kural olarak hiperlipidemi teşhisinde, serum enzim düzeylerinin değişiklikleri baz alınır. Hiperlipidemi teşhisinde dikkate alınan enzimler Total kolesterol, LDL, Trigliserit, HDL ve vLDL'dir. [60-63].

Bu bölümde, geliştirilen Hiperlipidemi teşhis karar destek sisteminin yapısının kavranmasına yönelik bilgiler alt bölümler halinde aşağıda anlatılmaktadır.

4.2.1.Lipidler

Hiperlipidemi plazma lipoproteinlerinin konsantrasyon artışına bağlıdır. Artmış yapım veya dolaşıma salınım ya da azalmış klerens veya dolaşımdan uzaklaştırılma nedeniyle bir veya daha fazla lipoprotein sınıfı kanda birikebilir. Metabolik olaylardaki bu değişiklikler sıklıkla lipoprotein metabolizması ile ilgili proteinlerdeki apolipoproteinler, reseptörler, enzimler veya kofaktörlerdeki değişikliklere bağlıdır. Genetik değişikliklere bağlı ortaya çıkan böyle değişiklikler lipid metabolizmasının primer bozuklukları olarak sınıflanır. Diabetes mellitus veya hipotiroidizm gibi lipoprotein metabolizmasını değiştiren diğer olaylar plazma lipoprotein konsantrasyonlarını arttırırlar; bu gibi durumlara ise ikincil bozukluklar denir [64]. Hiperlipidemi'de Total Kolesterol, LDL, HDL, Trigliserit ve vLDL enzimleri baz alınır [60-63,65]. Bu parametreler aşağıda kısaca açıklanmıştır:

Total Kolesterol: Kanda total kolesterol düzeyinin 200 mg/dL'nin altında olması istenmektedir. Kolesterol düzeyi 350 mg/dL'nin üstünde olan hastalar, koroner arterlerde erken ateromatöz

tıkanma gelişmesi ve erken yaşta iskemik kalp hastalığı belirtileri ortaya çıkması açısından yüksek risk altındadır. Ancak bu senaryo pek sık görülmemektedir ve 1960'lerden 1980'lerin başlarına kadar ailesel hiperkolesterolemi üzerinde yoğunlaşan dikkatler artık kan kolesterolündeki hafif ve orta derecedeki yükselmelere yönelmiştir. Total kolesterol ve HDL düzeyleri aç olup olmamaktan etkilenmez ve tok karnına da ölçülebilir. Bu strateji hastanın kan incelemesi için tekrar laboratuara başvurmasını önlediği için hem zaman kazandırır hem de maliyeti düşürür [66].

LDL: Bu lipoproteinlerin görevi doku ve organlara kolesterol taşımaktır. Halk arasında kötü kolesterol olarak ta bilinen bu lipoprotein miktarının yetişkinlerde belirlenmiş bazı değerlerin (100 mg/dL) üzerinde olması, istenmeyen bir durumdur. Bu lipoprotein türleri küçük yapıları olduğundan damar duvarını geçerek alt tabakada birikme kabiliyetine sahiptir. Bu birikme bir zaman sonra ateroskleroz (damar sertliği) adı verilen bir rahatsızlığa neden olur. Damar sertliğinde, damar duvarı zamanla damarın içine doğru bir çıkıntı oluşturmaya başlar ve bu, kan akımını azaltarak o damarın beslediği dokularda hasara ve ölüme neden olur. Bu da başta koroner kalp hastalığı (KAH) olmak üzere kalp krizi, inme, böbrek yetmezliği gibi hastalıkların ortaya çıkmasına neden olur. Bu bilgilerin ışığında, devam eden yüksek LDL seviyeleri bazı hastalıklara zemin hazırlayabilen tehlikeli bir durumdur.

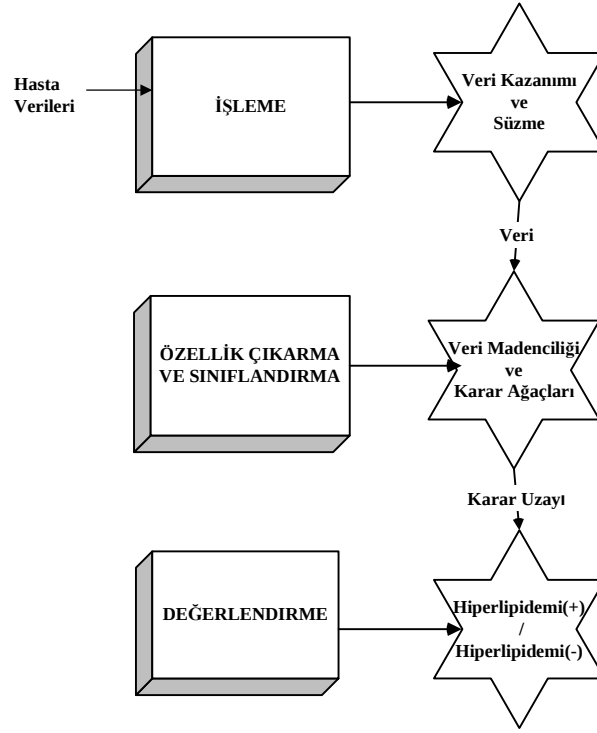
Trigliserit: Yağları oluşturan en küçük yapıtaşlarına yağ asidi adı verilir ve bu asitler vücutta depolanabilmek için gliserin adı verilen bir madde ile birleşerek trigliserid adını alırlar. Trigliseridler vücutta başlıca enerji kaynaklarıdır. Bu sebeple kanda belirli oranlarda bulunmaları gerekir. Kandaki seviyeleri 200mg/dL'nin üzerine çıktığı zaman artmış kabul edilir [67].

HDL: Hiperlipidemi tedavisine ilişkin kararlar verilmesinde HDL kolesterolü düzeyi şarttır. HDL kolesterolü arttıkça hiperlipidemi risk faktörünün yükseldiği gözlenmektedir. 60 mg/dL'nin üzerindeki HDL kolesterolü, koruyucu özellikte negatif bir risk faktörüdür. Total kolesterol gibi HDL'de daha önce yenmiş yemeklerden etkilenmemektedir [66].

vLDL: Lipoproteinler, lipidlerin ve proteinlerin oluşturdukları kompleks bir yapıdır. Protein tarafını apolipoprotein ve enzim adı verilen maddeler oluşturmaktadırlar. Çok düşük dansiteli lipoproteinler (VLDL) trigliserid miktarı bakımından zengin lipoproteinlerdir [67].

4.2.2. Yöntem

Biyokimya test sonuçlarından Hiperlipidemi teşhisi için dikkate alınan lipid parametrelerinin değerleri kontrol edilerek teşhise varılabilir. Geliştirilen karar destek sisteminin yapısı, yukarıda anlatılan örüntü tanıma kuralına göre hazırlanmış olan Şekil 4.4. de gösterilmiştir [47]. Bu sistemin işleyişi şu adımlardan oluşmaktadır:



Şekil 4.4. Karar destek sistemi algoritması.

Adım – 1: Ön işlemler

Dahiliye servisine başvuran hastaların biyokimya test sonuçları doktor gözetiminde toplandı. Toplanan verilerden, doktor yönlendirmesi ile hiperlipidemi şüphesi olan hastaların verileri ayrıştırılarak gerekli veriler elde edildi ve veri tabanı oluşturuldu.

Adım – 2: Özellik çıkarma ve sınıflandırma

Veritabanındaki verileri işlemeye yönelik olarak geliştirilen karar ağacı tabanlı karar destek sistemi sistemin en önemli aşamasıdır. Toplanan veriler bu aşamada işlenmektedir. Bunun için Tablo 4.5. de verilen kurallar doktor bilgilerinden elde edilerek, Şekil 4.5 deki karar ağacı yapısı oluşturulmuştur.

Tablo 4.5. Sistemin kural yapısı

Karar Ağacı Sırası	Kurallar
a	<i>if (tk>= 200) and (ldl>=180) and (tg>=165) and (hdl<=40) and (vldl>=50) then</i>
b	<i>else if (tk>= 200) and (ldl>=180) and (tg>=165) and (hdl<=40) and (vldl<50) then</i>
b	<i>else if (tk>= 200) and (ldl>=180) and (tg>=165) and (hdl>40) and (vldl>=50) then</i>
c	<i>else if (tk >= 200) and (ldl>=180) and (tg>=165) and (hdl>40) and (vldl<50) then</i>
d	<i>else if (tk>= 200) and (ldl>=180) and (tg<165) and (hdl <=40) and (vldl>=50) then</i>
e	<i>else if (tk>= 200) and (ldl>=180) and (tg<165) and (hdl >40) and (vldl>=50) then</i>
e	<i>else if (tk>= 200) and (ldl>=180) and (tg<165) and (hdl <=40) and (vldl<50) then</i>
f	<i>else if (tk>= 200) and (ldl>=180) and (tg<165) and (hdl >40) and (vldl<50) then</i>
g	<i>else if (tk>= 200) and (ldl<180) and (tg>=165) and (hdl <=40) and (vldl>=50) then</i>
h	<i>else if (tk>= 200) and (ldl<180) and (tg>=165) and (hdl >40) and (vldl>=50) then</i>
h	<i>else if (tk>= 200) and (ldl<180) and (tg>=165) and (hdl <=40) and (vldl<50) then</i>
i	<i>else if (tk>= 200) and (ldl<180) and (tg>=165) and (hdl>40) and (vldl<50) then</i>
j	<i>else if (tk>= 200) and (ldl<180) and (tg<165) and (hdl<=40) and (vldl>=50) then</i>
k	<i>else if (tk>= 200) and (ldl<180) and (tg<165) and (hdl>40) and (vldl>=50) then</i>
k	<i>else if (tk>= 200) and (ldl<180) and (tg<165) and (hdl<=40) and (vldl<50) then</i>
l	<i>else if (tk>= 200) and (ldl<180) and (tg<165) and (hdl>40) and (vldl<50) then</i>
m	<i>else if (tk< 200) and (ldl>=180) and (tg>=165) and (hdl<=40) and (vldl>=50) then</i>
n	<i>else if (tk< 200) and (ldl>=180) and (tg>=165) and (hdl<=40) and (vldl<50) then</i>
n	<i>else if (tk< 200) and (ldl>=180) and (tg>=165) and (hdl>40) and (vldl>=50) then</i>
o	<i>else if (tk< 200) and (ldl>=180) and (tg>=165) and (hdl>40) and (vldl<50) then</i>
p	<i>else if (tk< 200) and (ldl>=180) and (tg<165) and (hdl<=40) and (vldl>=50) then</i>
q	<i>else if (tk< 200) and (ldl>=180) and (tg<165) and (hdl>40) and (vldl>=50) then</i>
q	<i>else if (tk< 200) and (ldl>=180) and (tg<165) and (hdl<=40) and (vldl<50) then</i>
r	<i>else if (tk< 200) and (ldl>=180) and (tg<165) and (hdl>40) and (vldl<50) then</i>
s	<i>else if (tk< 200) and (ldl<180) and (tg>=165) and (hdl<=40) and (vldl>=50) then</i>
t	<i>else if (tk< 200) and (ldl<180) and (tg>=165) and (hdl<=40) and (vldl<50) then</i>
t	<i>else if (tk< 200) and (ldl<180) and (tg>=165) and (hdl>40) and (vldl>=50) then</i>
u	<i>else if (tk< 200) and (ldl<180) and (tg>=165) and (hdl>40) and (vldl<50) then</i>
v	<i>else if (tk< 200) and (ldl<180) and (tg<165) and (hdl<=40) and (vldl>=50) then</i>
w	<i>else if (tk< 200) and (ldl<180) and (tg<165) and (hdl>40) and (vldl>=50) then</i>
w	<i>else if (tk< 200) and (ldl<180) and (tg<165) and (hdl<=40) and (vldl<50) then</i>
x	<i>else if (tk< 200) and (ldl<180) and (tg<165) and (hdl>40) and (vldl<50) then</i>

Biyokimya test sonuçlarında sıralanan lipidler arasından lipid parametrelerinin normal değerleri Tablo 4.6.'de verilmiştir.

Tablo 4.6. Lipidlerin normal değerleri		
Lipidler	Normal değerleri	
tT. Kolesterol	0-200	mg/DL
LDL	80-180	mg/DL
Trigliserit	60-165	mg/DL
HDL	40-60	mg/DL
vLDH	20-50	mg/DL

Adım – 3: Değerlendirme

Bu aşamada, özellik çıkarma ve sınıflama süreci sonunda oluşan veriler değerlendirilip sunulmaktadır. Bu amaçla tasarlanan karar ağacı yapısında (Şekil 4.5), *a,b,c,d,e,f,g,h,i,j* yapraklarına ulaşan sonuçlar Hiperlipidemi (+) ve *k,l,m,n,o,p,q,r,s,t,u,v.w.x* yapraklarına ulaşan sonuçlar ise Hiperlipidemi (-) kararına varılmıştır.

4.2.3. Uygulama Sonuçları

Geliştirilen karar destek sistemi, 106 hastanın biyokimya test sonuçları üzerinde denenmiştir. Sistemde 106 hastadan 86'sine hiperlipidemi teşhisi, 20 hastaya ise sağlam tanısı konulmuştur ve bu sonuçların uzman bir doktorun verdiği kararlarla örtüştüğü görülmüştür. Bu başarıım yüzdesi, hiperlipidemi teşhisinde doktorun verdiği kararlar, geliştirilen sistemin sonuçlarının bire bir uygunluğu, tasarlanan karar destek sisteminin etkinliğini ve güvenilirliğini açıkça göstermektedir.

Table 4.7. Önerilen akıllı karar destek sisteminin başarıımı

	Hiperlipidemi	
	+	-
<i>Verilerin toplam sayısı</i>	86	20
<i>Doğru sınıflananlar</i>	86	20
<i>Hatalı sınıflananlar</i>	---	---

4.3. Hematolojik Parametrelerden Demir Eksikliği Anemisi Teşhisi

Çevre kanında hemoglobin miktarının kişinin yaşı ve cinsiyetine göre normal olarak kabul edilen değerlerin altına inmesine anemi denir [68]. Fiziopatoloji açısından, anemi, dolaşımdaki toplam eritrosit kitlesinin azalması olarak tanımlansa da fonksiyonel olarak kanın oksijen taşıma kapasitesinin azalması ve dolayısıyla doku hipoksisi anlamına gelmektedir. Kanın oksijen taşıma kapasitesi, eritrosit sayısının ve/veya hemoglobin miktarının azalması sonucunda azalmaktadır. Günlük uygulamada, eritrosit kitlesinin doğrudan ölçümü gibi yaygın olmayan, oranla pahalı ve zaman alan bir yöntem yerine eritrosit kitlesini değerlendirmede hemoglobin konsantrasyonu (Hb), hematokrit değeri (Hct) ve eritrosit sayısı kullanılmaktadır. Bu parametreler, örneğin, sırasıyla g/dl, % ve mm³'te sayı olarak ifade edilirler ve eritrositleri, içinde bulundukları sıvı ortam olan plazmaya oranla yansıtırlar. Bu nedenle, eritrositlerde değişme olmadan gelişen plazma kompartımanındaki değişikliklerden de etkilenirler. Dolayısıyla, anemi değerlendirmesi yapılırken, plazma hacminde artış olup olmadığı (hemodilüsyon) gözden geçirilmeli ve gerekli olan tedaviler yapıldıktan ya da belli bir süre geçtikten sonra ölçümler tekrarlanmalıdır. Her üç parametre de normal popülasyonda belirli bir dağılım göstermektedir. Bu dağılım yaş, cins, ırk ve yaşanan yükseklik gibi bazı önemli etkenlerden etkilenmekte olup verilen sınırlar toplumun %95'ini (aritmetik ortalama $\pm$ 2 standart sapma) kapsamaktadır [69]. Demir eksikliği [70], hemolitik [71], megaloblastik [72], aplastik [73,74], kronik hastalıklar anemisi [75] gibi çeşitleri vardır. Bizim burada dikkate

aldığımız demir eksikliği anemisi tanım olarak düşük miktarda demire bağlı olarak kanın kırmızı hücrelerindeki azalmadır. Kansızlığın en sık görülen şekli budur. Demir, kanda oksijen taşıyan pigment olan hemoglobinin önemli bir parçasıdır. Vücutta ve kemik iliğindeki demir depolarının harcanması sonucu kansızlık yavaş yavaş gelişir [76]. Özellikle küçük çocuklar ve üreme çağındaki kadınlar arasında sıklıkla görülen demir eksikliği anemisi çocuklarda bağışıklık sistemini zayıflatmakta, zihinsel ve fiziksel gelişimi olumsuz yönde etkilemekte ve enfeksiyona duyarlılığı arttırmaktadır [77]. Anemi kadınlarda dördüncü sırada görülürken, erkeklerde sekizinci sırada ve kadınlarda 38 ve 62 yaşlar civarında görülürken erkeklerde 70 ile 33 yaşlarında kümelenmektedir [78].

Kural olarak demir eksikliği anemisi teşhisinde, serum enzim düzeylerinin değişiklikleri baz alınır. Demir eksikliği anemisi teşhisinde dikkate alınan enzimler Serum demiri, Serum demir bağlama kapasitesi ve Ferritin'dir. [69].

Bu bölümde, geliştirilen Demir eksikliği anemisi teşhisi karar destek sisteminin yapısının kavranmasına yönelik bilgiler alt bölümler halinde aşağıda anlatılmaktadır.

4.3.1. Hemalotik Parametre Değerleri

Demir tedavisi oral veya parenteral uygulanabilir. Tüm hemodiyaliz hastalarında aralıklı olarak vücut depo demiri değerlendirilmelidir. Toplam vücut depo demirinin 800-1200 mg ve hemoglobindeki toplam demirin 2500 mg olduğu akılda tutulmalıdır. Vücut depo demiri 4 yöntemle değerlendirilebilir [79].

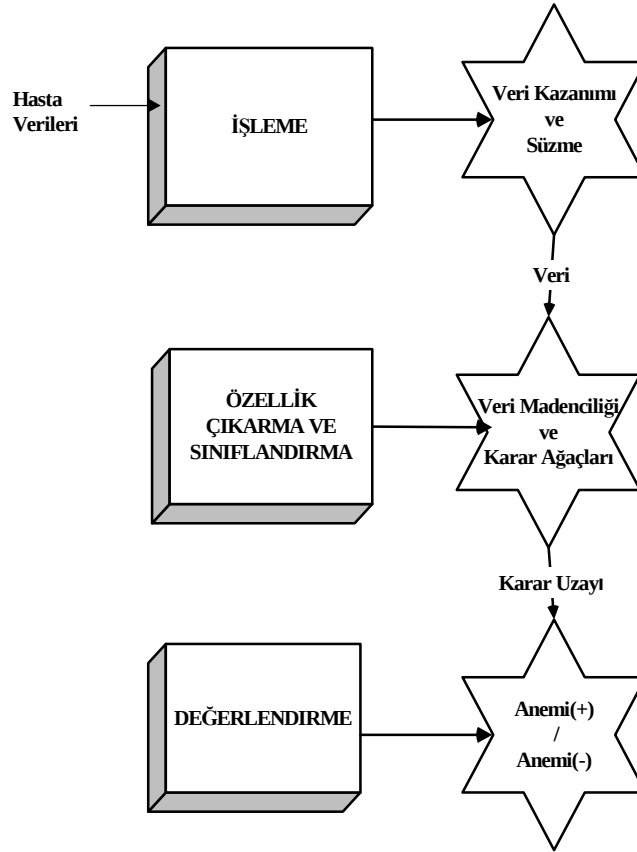
Serum demir ve transferrin düzeyi: Serum demiri, demir bağlama kapasitesi ve transferrin saturasyonu değerleriyle belirlenebilmektedir. Serum demirinin tamamına yakın bölümü transferrine bağlı olarak taşınmakta olup gün içinde değişiklikler göstermektedir. Demir eksikliği anemisinde ise düşük düzeylerde bulunmaktadır. Sabahları daha yüksek akşamları daha düşük değerlerde olması nedeniyle kan örneklerinin sabah ya da öğleden önce alınması önerilmektedir. Total demir bağlama kapasitesi (TDBK) serum demirinden daha az değişiklikler göstermekte demir eksikliğinde artmaktadır. Transferrin saturasyonu ise serum demirinin TDBK'ne oranını göstermekte demir eksikliğinde ve enfeksiyonlarda azalmış düzeylerde bulunmaktadır [80].

Normal serum demiri 60-190 µg/dl, serum demir bağlama kapasitesi (transferrin) 250-450 µg/dl ve normal doyumluk derecesi (serum demir/serum transferrin oranı) yüzde olarak % 20-45'tir. Serum demir ve demir bağlama kapasitesi ölçülerek doyumluk derecesi % olarak hesaplanır. Doyumluk derecesi % 15-20 arasında ise göreceli demir eksikliği, % 15'ten az ise mutlak demir eksikliği kabul edilir.

Serum ferritin düzeyi: Serum ferritin düzeyi, demir deposunun azlığı veya fazlalığını yansıtan bir parametredir [79]. Çocuklarda normal değerleri yaşla değişkenlik göstermekte yenidoğanda yüksek olan değerler bebeklik ve çocukluk çağında azalmaktadır. Düşük ferritin değerleri demir depolarının azalmış olduğunu gösterirken hafif üst solunum yolu enfeksiyonları, gastroenteritler gibi çocuklarda yaygın olarak görülen enfeksiyonlarda ve inflamasyonda normal ya da yüksek düzeyler elde edilebilmektedir. [81].

4.3.2. Yöntem

Biyokimya test sonuçlarında Demir eksikliği anemisi teşhisi için dikkate alınan enzim parametrelerinin değerleri kontrol edilerek teşhise varılabilir. Geliştirilen karar destek sisteminin yapısı, yukarıda anlatılan örüntü tanıma kuralına göre hazırlanmış olan Şekil 4.6'da gösterilmiştir [47]. Bu sistemin işleyişi şu adımlardan oluşmaktadır:



Şekil 4.6. Karar destek sistemi algoritması.

Adım – 1: Ön işlemler

Dahiliye servisine başvuran hastaların biyokimya test sonuçları doktor gözetiminde toplandı. Toplanan verilerden, doktor yönlendirmesi ile Anemi şüphesi olan hastaların verileri ayrıştırılarak gerekli veriler elde edildi ve veri tabanı oluşturuldu.

Adım – 2: Özellik çıkarma ve sınıflandırma

Veritabanındaki verileri işlemeye yönelik olarak geliştirilen karar ağacı tabanlı karar destek sistemi sistemin en önemli aşamasıdır. Toplanan veriler bu aşamada işlenmektedir. Bunun için Tablo 4.8.’de verilen kurallar doktor bilgilerinden elde edilerek, Şekil 4.7.’deki karar ağacı yapısı oluşturulmuştur.

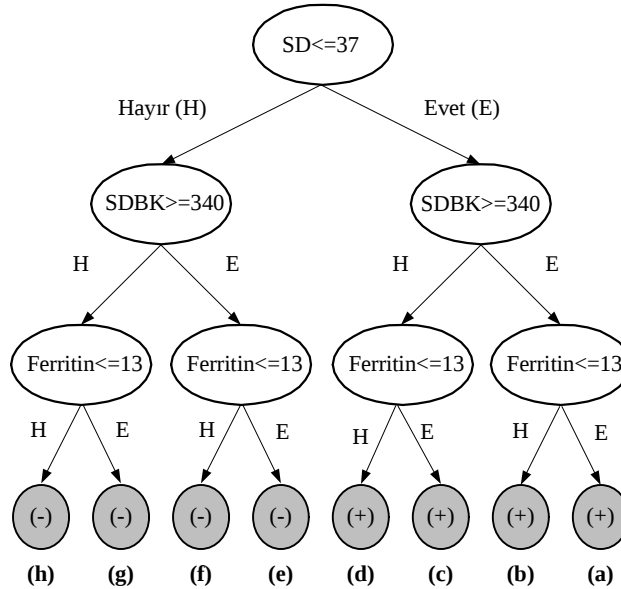
Tablo 4.8. Sistemin kural yapısı

Karar Ağacı Sırası	Kurallar	Karar
a	<i>if (sd <= 37) and (sbdk >= 340) and (ferritin <= 13) then</i>	<i>Anemi(+)</i>
b	<i>else if (sd <= 37) and (sbdk >= 340) and (ferritin > 13) then</i>	<i>Anemi(+)</i>
c	<i>else if (sd <= 37) and (sbdk < 340) and (ferritin <= 13) then</i>	<i>Anemi(+)</i>
d	<i>else if (sd <= 37) and (sbdk < 340) and (ferritin > 13) then</i>	<i>Anemi(+)</i>
e	<i>else if (sd > 37) and (sbdk >= 340) and (ferritin <= 13) then</i>	<i>Anemi(-)</i>
f	<i>else if (sd > 37) and (sbdk >= 340) and (ferritin > 13) then</i>	<i>Anemi(-)</i>
g	<i>else if (sd > 37) and (sbdk < 340) and (ferritin <= 13) then</i>	<i>Anemi(-)</i>
h	<i>else if (sd > 37) and (sbdk < 340) and (ferritin > 13) then</i>	<i>Anemi(-)</i>

Biyokimya test sonuçlarında sıralanan hemalotik parametreler arasından bu parametrelerin normal değerleri tablo 4.9. da verilmiştir.

Tablo 4.9. Hemalotik parametrelerin normal değerleri.

Hemalotik Parametreler	Normal değerleri
Serum Demiri	37-157 µg/dl
SDBK	210-340 µg/dl
Ferritin	13-400 ng/ml



Şekil 4.7. Karar ağacı yapısı

Adım – 3: Değerlendirme

Bu aşamada, özellik çıkarma ve sınıflama süreci sonunda oluşan veriler değerlendirilip sunulmaktadır. Bu amaçla tasarlanan karar ağacı yapısında, a,b,c,d yapraklarına ulaşan sonuçlar Anemi (+) ve e,f,g,h,j yapraklarına ulaşan sonuçlar ise Anemi (-) kararına varılmıştır.

4.3.3 Uygulama Sonuçları

Geliştirilen karar destek sistemi, 96 hastanın biyokimya test sonuçları üzerinde denenmiştir. Sistemde 96 hastadan 76'sine anemi teşhisi, 20 hastaya ise sağlam tanısı konulmuştur ve bu sonuçların uzman bir doktorun verdiği kararlarla örtüştüğü görülmüştür. Bu başarıyı yüzdesi, anemi teşhisinde doktorun verdiği kararlar, geliştirilen sistemin sonuçlarının bire bir uygunluğu, tasarlanan karar destek sisteminin etkinliğini ve güvenilirliğini açıkça göstermektedir.

Table 4.10. Önerilen akıllı karar destek sisteminin başarımı

	Anemi	
	+	-
Verilerin toplam sayısı	76	20
Doğru sınıflananlar	76	20
Hatalı sınıflananlar	---	---

4.4. Tiroid Hormon Parametrelerinden Hipotiroidi-Hipertiroidi Teşhisi

Tiroid hormonları, çeşitli metabolik olayları etkiler. Bu nedenle tiroid hastalıklarında organizmadaki çeşitli organların fonksiyonlarında değişiklik olmaktadır. Tiroid hormonlarının kardiovasküler etkileri ayrıntılı olarak araştırılmıştır. Tiroid hormon eksikliğinin bulguları ise bradikardi, nabız basıncında küçülme, kalp debisinde azalma, sol ventrikül kontraktilesinde azalmadır [82]. Bu durum Hipotiroidi olarak tanımlanır ve sıklıkla otoimmün nedenlere (otoimmün hastalık, vücudun kendi dokularından birine bilinmeyen nedenlerle yabancılaşması ve bu “yabancı” dokuyu bağışıklık sistemiyle vücuttan uzaklaştırmaya yönelik girişimler yapması neticesinde oluşan hastalıktır. Bağışıklık sistemi vücudun aslında kendine ait olan bu dokusunu tahrip ettikçe dokunun işlevleri aksamakta ve buna bağlı sorunlar ortaya çıkmaktadır) bağlı olarak ortaya çıkan bu durumda tiroid bezi tahribat görmüş olması nedeniyle işlevlerini daha az yapmaktadır. Hipotiroidi gelişme riski her yaşta var olmasına karşın risk yaş ilerledikçe artar ve 60 yaşından sonra yüzde 2-4 oranında hipotiroidi görülür [83]. Tiroid bezi yetmezliği TSH, serbest T3 (sT3) ve serbest T4 (sT4) hormonlarının ölçülmesiyle kolayca tanınır ve

tedavisine başlanır [84]. Tiroid hormon fazlalığında taşikardi, nabız basıncında artma, kalp debisinde artma, sol ventrikül kontraktilitesinde artma görülür [82] ve bu durum da Hipertiroidi olarak tanımlanır ve sıklıkla "Toksik Diffüz Guatr" veya Toksik Multinodüler Guatr şeklinde görülen hipertiroidi, hipotiroidi gibi nispeten sık rastlanan bir durumdur. Çeşitli belirtileriyle hipertiroidi düşündüren durumlarda veya tarama amacıyla yapılan TSH incelemesinin düşük bulunması ve sT4 veya sT3 seviyesinin yüksek bulunması tanıyı koymak için yeterlidir. [85]. Günümüzde hem total tiroksin (TT4), total triiyodotironin (TT3) ve hem de serbest tiroksin (ST4) ve serbest triiyodotironin (ST3) düzeyleri direkt olarak serumda ölçülebilmektedir [86].

Bu bölümde, geliştirilen Hipertiroidi-Hipotiroidi teşhisi karar destek sisteminin yapısının kavranmasına yönelik bilgiler alt bölümler halinde aşağıda anlatılmaktadır.

4.4.1.Hormonlar

T3-T4: Tiroidin folliküler hücrelerinden tiroksin (T4) ve triiyodotironin (T3) hormonları salgılanır. Ayrıca parafolliküler hücrelerden de kalsiyumun metabolizmasında etkili olan kalsitonin salgılanmaktadır. T3 ve T4 genel anlamda bazal metabolizmayı düzenleyen hormonlardır. Hücre içinde bulunan nükleus reseptörlerine bağlanarak protein yapımını regüle ederler. Ayrıca mitokondrilerde oksidasyon olaylarını hızlandırırlar, membran yapısında yer alan enzimlerin aktivitesini kontrol etmek gibi diğer fonksiyonları da vardır. Bu bağlamda tiroid hormonları yaşam için mutlak gereklidirler. Tiroidden T3 ve T4 sekresyonu anterior hipofizden salgılanan tiroid stimulan hormonun (TSH) kontrolü altındadır. TSH uyarısı T3 ve T4 salınımını uyarırken, kandaki T3 ve T4 artışı hipofizden TSH salınımını suprese eder (negatif feed-back) ve salınımı ise hipotalamustan salgılanan TRH'nın (tirotropin releasing hormon, tiotropin serbestleştirici hormon) kontrolü altındadır. Tiroksin ve triiyodotironin sekresyonunun artmasıyla metabolizma hızı %60-100 oranında artabilir. Salgının ortadan kalkması ise metabolizma hızını normalin %40 altına düşürür.

Tiroid hormonlarının oluşumu eksojen iyot alımına bağlıdır. Follikül hücresinde tirozine bir iyot bağlanması ile monoiyodotirozin (MIT),iki iyot bağlanması ile diiyodotirozin (DIT) oluşur. İki DIT eşlendiğinde T4,bir MIT ile bir DIT eşlendiğinde T3 meydana gelir. Tiroid hormonları tiroglobuline (Tg) bağlı olarak follikül içindeki kolloidde depolanır. Bu depo vücudun 1-3 aylık ihtiyacını karşılamaya yeterlidir.

T3 ve T4 tiroglobulinden ayrılarak serbest hormon şeklinde kana salgılanırlar ve tamamına yakını plazma proteinlerine bağlanırlar. Bu hormonlara bağlanma eğilimi en yüksek olan taşıyıcı protein bir glikoprotein olan tiroksin bağlayan globulin (TBG)'dir,ki 2/3'ünü bağlar ve T3'e bağlanma eğilimi daha düşüktür.1/4'ü tiroksin bağlayan prealbumine

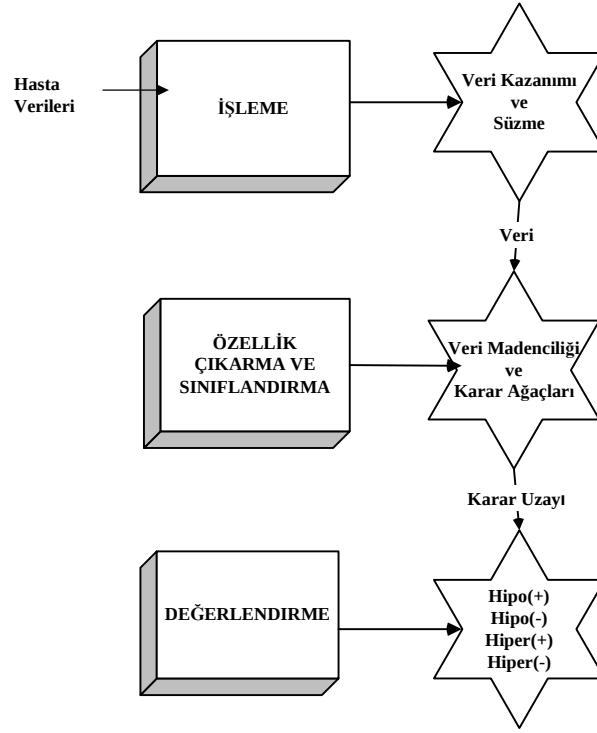
(TBPA),1/10 kadarı da albumine bağlanır.Plazmadaki tiroid hormonlarının %0,02'si serbest haldedir ve bunlar fizyolojik olarak aktif fraksiyonu oluşturur.

Tiroid bezinden salgılanan hormonun %90'ı T4,%10'u ise T3'tür.Bununla birlikte tiroksinin önemli bir bölümü (%75-85) kanda triiyodotironine çevrilir (T4'ün T3'e deiyodinasyonu).Bu çevrilme çok önemlidir çünkü T3 plazmada 10-20 kat daha az miktarda bulunsada T4'ten dört kat daha aktiftir.T3'ün yarılanma ömrü bir gün iken T4'ün yedi gündür.

TSH: TSH, primer hipotiroidizmin ve subklinik hipertiroidizm tanısında,tiroid hormon replasman (hipotiroidizm) ve süpresyon (benign guatr ve tiroidektomi yapılmış diferansiye tiroid karsinomu) tedavilerini değerlendirmede ideal bir testtir. Ancak serum TSH düzeyi her zaman tiroidin fonksiyonel durumunu doğru olarak yansıtmayabilir. Örneğin hiper yada hipotiroidizmin tedavisi sırasında tiroid hormon düzeyleri hızla değiştiğinden TSH'nın FT4 ile sabit bir dengeye ulaşması uzun sürer (bazen 6 aya kadar). Tiroid hormonlarına direnç varsa ve klinik ötiroid veya hipertiroidili hastalarda, FT3 konsantrasyonunun yükselmiş olmasına karşın TSH düzeyi normal ya da yüksektir [87].

4.4.2. Yöntem

Biyokimya test sonuçlarında Hipertiroidi-Hipotiroidi teşhisi için dikkate alınan enzim parametrelerinin değerleri kontrol edilerek teşhise varılabilir. Geliştirilen karar destek sisteminin yapısı, yukarıda anlatılan örüntü tanıma kuralına göre hazırlanmış olan Şekil 4.8.'de gösterilmiştir [47]. Bu sistemin işleyişi şu adımlardan oluşmaktadır:



Şekil 4.8. Karar destek sistemi algoritması.

Adım – 1: Ön işlemler

Verilerin toplanarak, amaca uygun olanların seçildiği aşamadır. Dahiliye servisine başvuran hastaların biyokimya test sonuçları doktor gözetiminde toplandı. Toplanan verilerden, doktor yönlendirmesi ile Anemi şüphesi olan hastaların verileri ayrıştırılarak gerekli veriler elde edildi ve veri tabanı oluşturuldu.

Adım – 2: Özellik çıkarma ve sınıflandırma

Veritabanındaki verileri işlemeye yönelik olarak geliştirilen karar ağacı tabanlı karar destek sistemi sistemin en önemli aşamasıdır. Toplanan veriler bu aşamada işlenmektedir. Bunun için Tablo 4.11. de verilen kurallar doktor bilgilerinden elde edilerek, Şekil 4.9.daki karar ağacı yapısı oluşturulmuştur.

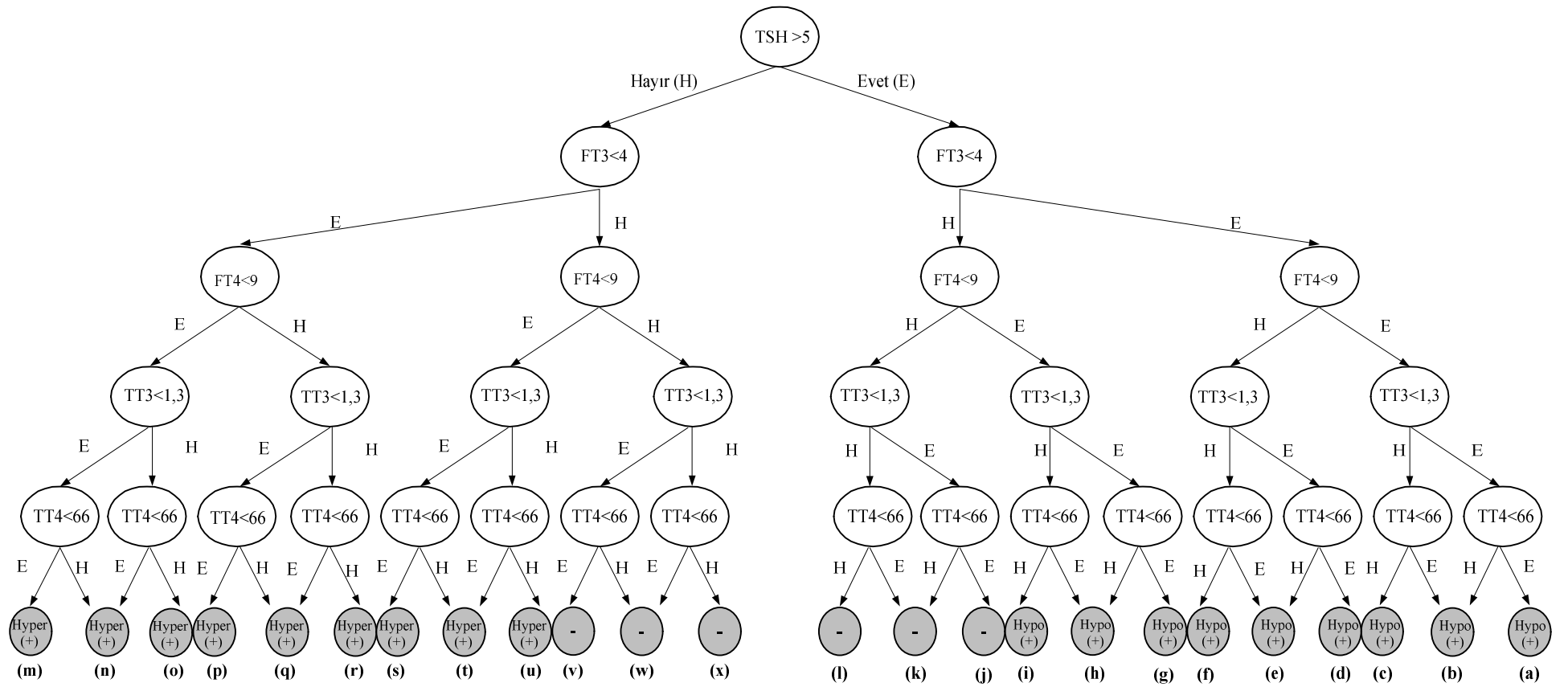
Tablo 4.11. Sistemin kural yapısı

Karar Ağacı Sırası	Kurallar
a	<i>if (tsh >= 5) and (ft3 < 4) and (ft4 < 9) and (tt3 < 1,3) and (tt4 < 66) then</i>
b	<i>else if (tsh >= 5) and (ft3 < 4) and (ft4 < 9) and (tt3 < 1,3) and (tt4 >= 66) then</i>
b	<i>else if (tsh >= 5) and (ft3 < 4) and (ft4 < 9) and (tt3 >= 1,3) and (tt4 < 66) then</i>
c	<i>else if (tsh >= 5) and (ft3 < 4) and (ft4 < 9) and (tt3 >= 1,3) and (tt4 >= 66) then</i>
d	<i>else if (tsh >= 5) and (ft3 < 4) and (ft4 >= 9) and (tt3 < 1,3) and (tt4 < 66) then</i>
e	<i>else if (tsh >= 5) and (ft3 < 4) and (ft4 >= 9) and (tt3 >= 1,3) and (tt4 < 66) then</i>
e	<i>else if (tsh >= 5) and (ft3 < 4) and (ft4 >= 9) and (tt3 < 1,3) and (tt4 >= 66) then</i>
f	<i>else if (tsh >= 5) and (ft3 < 4) and (ft4 >= 9) and (tt3 >= 1,3) and (tt4 >= 66) then</i>
g	<i>else if (tsh >= 5) and (ft3 >= 4) and (ft4 < 9) and (tt3 < 1,3) and (tt4 < 66) then</i>
h	<i>else if (tsh >= 5) and (ft3 >= 4) and (ft4 < 9) and (tt3 >= 1,3) and (tt4 < 66) then</i>
h	<i>else if (tsh >= 5) and (ft3 >= 4) and (ft4 < 9) and (tt3 < 1,3) and (tt4 >= 66) then</i>
i	<i>else if (tsh >= 5) and (ft3 >= 4) and (ft4 < 9) and (tt3 >= 1,3) and (tt4 >= 66) then</i>
j	<i>else if (tsh >= 5) and (ft3 >= 4) and (ft4 >= 9) and (tt3 < 1,3) and (tt4 < 66) then</i>
k	<i>else if (tsh >= 5) and (ft3 >= 4) and (ft4 >= 9) and (tt3 >= 1,3) and (tt4 < 66) then</i>
k	<i>else if (tsh >= 5) and (ft3 >= 4) and (ft4 >= 9) and (tt3 < 1,3) and (tt4 >= 66) then</i>
l	<i>else if (tsh >= 5) and (ft3 >= 4) and (ft4 >= 9) and (tt3 >= 1,3) and (tt4 >= 66) then</i>
m	<i>else if (tsh < 5) and (ft3 < 4) and (ft4 < 9) and (tt3 < 1,3) and (tt4 < 66) then</i>
n	<i>else if (tsh < 5) and (ft3 < 4) and (ft4 < 9) and (tt3 < 1,3) and (tt4 >= 66) then</i>
n	<i>else if (tsh < 5) and (ft3 < 4) and (ft4 < 9) and (tt3 >= 1,3) and (tt4 < 66) then</i>
o	<i>else if (tsh < 5) and (ft3 < 4) and (ft4 < 9) and (tt3 >= 1,3) and (tt4 >= 66) then</i>
p	<i>else if (tsh < 5) and (ft3 < 4) and (ft4 >= 9) and (tt3 < 1,3) and (tt4 < 66) then</i>
q	<i>else if (tsh < 5) and (ft3 < 4) and (ft4 >= 9) and (tt3 >= 1,3) and (tt4 < 66) then</i>
q	<i>else if (tsh < 5) and (ft3 < 4) and (ft4 >= 9) and (tt3 < 1,3) and (tt4 >= 66) then</i>
r	<i>else if (tsh < 5) and (ft3 < 4) and (ft4 >= 9) and (tt3 >= 1,3) and (tt4 >= 66) then</i>
s	<i>else if (tsh < 5) and (ft3 >= 4) and (ft4 < 9) and (tt3 < 1,3) and (tt4 < 66) then</i>
t	<i>else if (tsh < 5) and (ft3 >= 4) and (ft4 < 9) and (tt3 < 1,3) and (tt4 >= 66) then</i>
t	<i>else if (tsh < 5) and (ft3 >= 4) and (ft4 < 9) and (tt3 >= 1,3) and (tt4 < 66) then</i>
u	<i>else if (tsh < 5) and (ft3 >= 4) and (ft4 < 9) and (tt3 >= 1,3) and (tt4 >= 66) then</i>
v	<i>else if (tsh < 5) and (ft3 >= 4) and (ft4 >= 9) and (tt3 < 1,3) and (tt4 < 66) then</i>
w	<i>else if (tsh < 5) and (ft3 >= 4) and (ft4 >= 9) and (tt3 >= 1,3) and (tt4 < 66) then</i>
w	<i>else if (tsh < 5) and (ft3 >= 4) and (ft4 >= 9) and (tt3 < 1,3) and (tt4 >= 66) then</i>
x	<i>else if (tsh < 5) and (ft3 >= 4) and (ft4 >= 9) and (tt3 >= 1,3) and (tt4 >= 66) then</i>

Biyokimya test sonuçlarında sıralanan hormonlar arasından hormon parametrelerinin normal değerleri Tablo 4.12.'de verilmiştir.

Tablo 4.12. Hormonların normal değerleri

Tiroid Hormanları	Normal Değerler
TSH	0,25-5 mg/dl
FT3	4-8,3 mg/dl
FT4	9-20 mg/dl
TT3	1,3-3,1 nmol/L
TT4	66-181 nmol/L



Şekil 4.9. Karar Ağacının Yapısı

Adım – 3: Değerlendirme

Bu aşamada, özellik çıkarma ve sınıflama süreci sonunda oluşan veriler değerlendirilip sunulmaktadır. Bu amaçla tasarlanan karar ağacı yapısında (Şekil.4.9.), $a, b, c, d, e, f, g, h, i$ yaprağı için Hipotiroidi(+) $m, n, o, p, q, r, s, t, u$ yaprağı için Hipertiroidi(-), j, k, l, v, w ve x yaprakları için (-) kararına varılmıştır.

4.4.3. Uygulama Sonuçları

Geliştirilen karar destek sistemi, 120 hastanın biyokimya test sonuçları üzerinde denenmiştir. Sistemde 120 hastadan 35'ine Hipertiroidi, 35'ine Hipotiroidi, 50 hastaya ise sağlam tanısı konulmuştur ve bu sonuçlar uzman bir doktorun verdiği kararlarla örtüştüğü görülmüştür. Bu başarımlar yüzdesi, Hipertiroidi-Hipotiroidi teşhisinde doktorun verdiği kararlar, geliştirilen sistemin sonuçlarının bire bir uygunluğu, tasarlanan karar destek sisteminin etkinliğini ve güvenilirliğini açıkça göstermektedir.

Tablo 4.13. Karar destek sisteminin başarımları

	Hipertiroidi		Hipotiroidi	
	+	-	+	-
<i>Verilerin toplam sayısı</i>	35	25	35	25
<i>Doğru sınıflananlar</i>	35	25	35	25
<i>Hatalı sınıflananlar</i>	---	---	---	---

BÖLÜM 5

5. SONUÇ VE DEĞERLENDİRME

Veri madenciliği yaklaşımlarıyla tıp dünyasında hastalara hizmet sunulması günümüzdeki popüler konulardan olmuştur. Bu nedenle tezde bu amaca yönelik olarak biyokimya verileri ile hastalık teşhisi, uzman doktor gözetiminde hazırlanan kurallara göre hastalık teşhisinde bulunulması için karar destek sistemi tasarımı ve gerçekleştirimi yapılmıştır. Tasarımlarda veri madenciliği tekniklerinden karar ağacı yöntemi kullanılmıştır. Böylece kalp krizi, hiperlipidemi, hipertiroidi-hipotiroidi ve demir eksikliği anemisi gibi hastalıklarının teşhislerinde uzman veya pratisyen bir doktora yardımcı olabilecek karar destek sistemleri tasarlanmıştır.

Geliştirilen sistemde:

- Biyokimya verileri ile teşhis edilebilecek hastalıkların belirlenmesi,
 - Belirlenen hastalıklara göre hastanenin gereken bölümlerine başvurarak hasta verilerinin toplanarak bir veri tabanı oluşturulması,
 - Belirlenen hastalıklar için uzman doktorların belirlediği kurallara göre karar ağaçlarının oluşturulması
 - Oluşturulan karar ağaçlarına göre hasta verilerinin sınıflandırılması ve değerlendirilmesi
 - Hazırlanan hastalık teşhis sisteminin güvenilirliğinin ve etkinliğinin kontrolünün yapılması,
- başarıyla gerçekleştirilmiştir.

5.1. Sonuçların İrdelenmesi

Karar verme işlemi doğrudan bilgisayar tarafından yapılmıştır. Bu değerlendirme doktorlar için doğrudan bir karar sonucu değil, karar sürecine etki edecek bir bulgudur. Geliştirilen karar destek sisteminde kalp krizi için 61 hasta verisi kullanılmıştır ve 50 hasta verisi için kalp krizi, 11 hasta verisi için sağlam tanısı konulmuştur. Hiperlipidemi için 106 hasta verisi kullanılıp 86 hasta verisi için hiperlipidemi, 20 hasta verisi için sağlam kararına varılmıştır. Demir eksikliği anemisinde 96 hasta verisi kullanılmış ve 76 hasta için anemi 20 hasta için sağlam kararına varılmıştır. Hipotiroidi ve Hipertiroidi için 120 hasta verisi kullanılmıştır ve 35 hasta verisi için hipertiroidi, 35 hasta için hipotiroidi ve 50 hasta için sağlam tanısı konulmuştur. Bu hastalarda sistemin ulaştığı sonuç doktorun verdiği kararlar % 100 doğrulukla örtüşmektedir.

Geliştirilen sistem esnek bir yapıya sahip olduğundan, sisteme yeni hastalıklar eklenebilir ve sistem geliştirilebilir.

Ayrıca sistem uzman bir hekimin kontrollünde internet tabanlı olarak geliştirilebilirse, teknik donanımın eksikliği ile olumsuz koşullar altında çalışan diğer hekimlere yardımcı bir karar destek organı olabileceği gibi hekimlerin bulunamadığı ortamlarda internet üzerinden kontrolle hastalara geniş kapsamlı bir hizmet sunulabilir.

Sistemin ticari açılardan kullanılabilmesi için istisnai durumların ele alınması, güvenlik açısından geliştirilmesi gerekmektedir.

KAYNAKLAR

- [1] Tuğ, E., Şakiroğlu, A., Bulun , M., Tıbbi Veri Tabanlarında Gizli Bilgilerin Keşfedilmesi, Türkiye Bilişim Derneği 20. Bilişim Kurultayı , 2003.
- [2] Internet: Veri Madencilik Uygulamaları ve Eğitimi, <http://courses.cs.deu.edu.tr/cse572>, Ocak 2006.
- [3] Internet: İşletme Fakültesi Dergisi, <http://www.isletme.istanbul.edu.tr/dergi/nisan2000/1.htm>, Ocak 2006.
- [4] Bilim ve Teknik, İnsan Genomu, Mart 2001.
- [5] Çömlekçioglu, U., Biyoinformatik Üzerine, 2003.
- [6] Cebeci, Z., Bek, Y., 1999. Tarımsal Enformatik Eğitimi Mümkün mü? 3. Tarımda Bilgisayar Uygulamaları Sempozyumu (Çağrılı Bildiri), Çukurova Üniversitesi, Adana, 3-6 Ekim 1999.
- [7] Zaki, M., Scalable Data Mining for Rules, University of Rochester, New York U.S.A., 1998.
- [8] Alonso, F., Caraça-Valente, J. P., Combining Expert Knowledge and Data Mining in a Medical Diagnosis Domain, Expert Systems with Applications 23, 367-375, 2002.
- [9] Türkoğlu, İ., Arslan, A. ve İlkay, E., “An Intelligent System for Diagnosis of Heart Valve Diseases with Wavelet Packet Neural Networks”, Computer in Biology and Medicine, 33(4), 319-331, (2003).
- [10] Öger, G. S., Osteoporoz Hastalığının Tanısı İçin Veri Madenciliği Kullanımı, Fırat Üniversitesi, Yüksek Lisans Tezi, 2003.
- [11] Sever, H., Oğuz, B., Veri Tabanlarında Bilgi Keşfine Formal Bir Yaklaşım, Ankara, 2003.
- [12] Bojarczuk C. C., Lopes, H. S., A Constrained-Syntax Genetic Programming System for Discovering Classification Rules: Application to Medical Data Sets, Artificial Intelligence in Medicine 30, 27–48, 2004.
- [13] Agrawal R., Mehta M., The Quest Data Mining System, IBM Almaden Research Center San Jose 6s California, U.S.A., 1996.
- [14] Information discovery in databases and data mining, I. U. The school of business administration magazine, C:29, S: 1, 2000.
- [15] İlhan, S., Veri Madenciliği, Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, 2005.
- [16] Üzgünyıldız, T., Veri Ambarı ve Veri Madenciliği, 2002.
- [17] Alpaydin, E., Zeki Veri Madenciliği, Bilişim 2000 Eğitim Semineri, 2000.
- [18] Internet: Information Management, Data mining and information discovery, http://www.bilgiyonetimi.org/cm/pages/mkl_gos.php?nt=538, 2000.
- [19] SPSS Inc. AnswerTree 2.0 User’s Guide, 1998, ISBN 1-56827-254-5, 1998.

- [20] Bishop, C. M., Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1996.
- [21] Hogl, O., Müller, M., On Supporting Medical Quality with Intelligent Data Mining, Proceedings of the 34th Hawaii International Conference on System Sciences, 2001.
- [22] Tan, K. C., Yu, Q., Evolutionary Computing for Knowledge Discovery in Medical Diagnosis, Artificial Intelligence in Medicine 27, 129–154, 2002.
- [23] Kusiak, A., Dixon, B., Predicting Survival time for Kidney Dialysis Patients: a Data Mining Approach, Computers in Biology and Medicine 35, 311–327, 2005.
- [24] Aydoğan, F., E-Ticarete Veri Madenciliği Yaklaşımlarıyla Müşteriye Hizmet Sunan Akıllı Modüllerin Tasarımı ve Gerçekleştirimi, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 2003.
- [25] Allahverdi, N., Uzman Sistemler ile Bir Yapay Zeka Uygulaması, Atlas Yayın Dağıtım, İstanbul, 2002.
- [26] Altınışık M., Klinik Biyokimyaya Ve Enzimolojiye Giriş, Adü Tıp Fakültesi, 2000.
- [27] Altınışık M., Kan ve İdrar Biyokimyası, Adü Tıp Fakültesi, 2003.
- [28] Özata, M., Aslan, Ş., Klinik Karar Destek Sistemleri ve Örnek Uygulamalar, Afyon Kocatepe Üniversitesi, Kocatepe Tıp Dergisi, 5 (1), 11-19, 2004.
- [29] Barrera, J., Cesar-Jr, R. M., An Environment For Knowledge Discovery in Biology, Computers in Biology and Medicine 34, 427–447, 2003.
- [30] Podgorelec, V., Kokol, P., Knowledge Discovery with Classification Rules in a Cardiovascular Dataset, Computer Methods and Programs in Biomedicine 80 Suppl. 1, S39-S49, 2005.
- [31] Bojarczuk, C. C., Lopes, H. S., A Constrained-Syntax Genetic Programming System for Discovering Classification Rules: Application to Medical Data Sets, Artificial Intelligence in Medicine 30, 27–48, 2004.
- [32] Wood, R. D., Mitchell, M., Human DNA Repair Genes, Analysis of Genomic Information, Vol 291 Science, 2001.
- [33] Yin, M. M., Wang, J. T. L., GeneScout: a Data Mining System For Predicting Vertebrate Genes in Genomic DNA Sequences, Information Sciences 163, 201–218, 2003.
- [34] Ayub, J. M., Smulski, C. R., Protein–Protein Interaction Map of the Trypanosoma Cruzi Ribosomal P Protein Complex, Gene 357, 129 – 136, 2005.
- [35] Shebuski, R. J., Utility of Point-of-Care Diagnostic Testing in Patients with Chest Pain and Suspected Acute Myocardial Infarction, Current Opinion in Pharmacology, 2:160–164, 2002.

- [36] Turkoglu, İ., Arslan, A., “An Intelligent System for Diagnosis of Heart Valve Diseases with Wavelet Packet Neural Networks”, *Computer in Biology and Medicine*, 33(4), 319-331, 2003.
- [37] Cooke, CD., Santana, CA., Validating Expert System Rule Confidences Using Data Mining of Myocardial Perfusion SPECT Databases, *Computers in Cardiology*, 27: 785–788, 2000.
- [38] Unger, T., Korade, Z., True and False Discovery in DNA Microarray Experiments: Transcriptome Changes in the Hippocampus of Presenilin 1 Mutant Mice, *Methods* 37, 261–273, 2005.
- [39] Huang, M. J., Chen, M. Y., Integrating Data Mining with Case-Based Reasoning for Chronic Diseases Prognosis and Diagnosis, *Expert Systems with Applications*, 2006.
- [40] Kusiak, A., Caldarone, C. A., Hypoplastic Left Heart Syndrome: Knowledge Discovery with a Data Mining Approach, *Computers in Biology and Medicine* 36, 21–40, 2006.
- [41] Mullins, I. M., Siadat, M. S., Data Mining and Clinical Data Repositories: Insights from a 667,000 Patient Data Set, *Computers in Biology and Medicine* 2005.
- [42] Türkoğlu, İ., Durağan olmayan işaretler için zaman-frekans entropilerine dayalı akıllı örüntü tanıma, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, 2002.
- [43] Toraman, S., Histopatolojik İmgelerin Değerlendirilmesinde Örüntü Tanıma Temelli Karar Destek Sistemleri, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2006.
- [44] Phan, F., Tzanakou, E.M. and Sideman, S., 2000, Speaker identification using neural networks and wavelets, *IEEE Engineering in Medicine and Biology*, 19(1), 92-101, 2000.
- [45] Türkoğlu, İ. ve Hanbay, D., Yapay sinir ağı ve HFD kullanarak DTMF sinyal örüntülerini tanıma sistemi, *Elektrik-Elektronik-Bilgisayar Mühendisliği 9.Ulusal Kongresi*, 431-434, Kocaeli, 19-23 Eylül 2001.
- [46] Duda, R.O. and Hart, P.E., *Pattern Classification and Scene Analysis*. Stanford Research Institute, 1989.
- [47] Turkoglu İ., Arslan, A., An Expert System for Diagnose of The Heart Valve Diseases, *Expert Systems with Applications*, 23(3), 229-236, 2002.
- [48] Altınışık M, Akut Miyokard İnfarktüsünün Diagnostik Laboratuvar Belirleyicileri, 2006.
- [49] Özçakır A, Birinci Basamak Sağlık Hizmetinde Göğüs Ağrısı Olan Hastalara Yaklaşım,
- [50] Güngör Ö., Sunar B., Özçelik F., Aktaş Z., Gökmen S., Serum Sialic Acid Levels in Acute Myocardial Infarction and Its Relationship to Ceruloplasmin, 2004
- [51] Heper C, Myocardial Infarction, 1998.
- [52] Internet: www.hastarehberi.com/article_read.asp?id=1539 - 40k, Mart 2006.

- [53] Braunwald, E., Zipes, D. P., Heart disease: a textbook of cardiovascular medicine, 6 th edition, 2001.
- [54] Harff, G. A., Jeurissen, R. W. M., Differentiation between transmural perioperative myocardial infarction and subendocardial injury after coronary artery bypass grafting using biochemical tests, elaborated by cluster and discriminant analysis, Clinica Chimica Acta 274, 29–40, 1998.
- [55] Azezli, A., Kubat, A., Hiperlipidemi Tedavisinde Statinlerin Yeri, Türkiye Klinikleri J Int Med Sci, 1(20):42-48, 2005.
- [56] Internet: MSD- Hiperlipidemi, <http://www.msd-turkey.com/content/patients/therapy/t5.html>, Mart 2005.
- [57] Saka, B., Koroner Kalp Hastalığı, Haydarpaşa Numune Eğitim ve Araştırma Hastanesi, Biyokimya ve Klinik Biyokimya Bölümü, 2004.
- [58] Güven, A., Sasmaz, S., Serum Lipid Parameters Evaluation in Patients with Psoriasis, T Klin Dermatoloji, 13:91-94, 2003.
- [59] Stinson J, O'Toole E., Cholesterol and lipoprotein (a) levels in psoriasis; 88:128-9, Ir Med J 1995.
- [60] Koba, S., Pakala, R., Hyperlipemic-very low density lipoprotein, intermediate density lipoprotein and low density lipoprotein act synergistically with serotonin on vascular smooth muscle cell proliferation, Atherosclerosis 149, 61–67, 2000.
- [61] Lupattelli, G., Marchesi, S., Mechanisms of High-Density Lipoprotein Cholesterol Effects on the Endothelial Function in Hyperlipemia, Metabolism, Vol 52, No 9: pp 1191-1195, (September), 2003.
- [62] Wang, T. D., Chen, W. J., Efficacy of fenofibrate and simvastatin on endothelial function and inflammatory markers in patients with combined hyperlipidemia: relations with baseline lipid profiles, Atherosclerosis 170, 315–323, 2003.
- [63] Lee, H. S., Ahn, H. C., Hypolipemic effect of water extracts of Picrorrhiza rhizoma in PX-407 induced hyperlipemic ICR mouse model with hepatoprotective effects: A prevention study, Journal of Ethnopharmacology 105, 380–386, 2006.
- [64] Sarandöl, E., <http://www.msd-turkey.com/content/patients/therapy/t5.html>, Mart 2006.
- [65] Karpuz, H., Hiperlipidemiye Güncel Yaklaşım, Sık Gorülen Kardiyolojik Sorunlarda Guncelleme Sempozyum Dizisi No: 40, s. 69-74, 2004.
- [66] Khan, M. G., Kalp Hastalıkları Tanı ve Tedavisi, Turgut yayıncılık ve ticaret, İstanbul, 1998.
- [67] Internet: Izedebiyat- Hiperlipidemi, <http://www.izedebiyat.com/yazi.asp?id=30679>, Mart 2006.

- [68] Akman, N., Erişkinde anemilere genel yaklaşım, Anemiler Sempozyumu, İstanbul, s. 9-16, Nisan 2001.
- [69] Başlar, Z., Anemili Hastada Ayırıcı Tanı Rehberi, Anemiler Sempozyumu, İstanbul, s. 107-116, Nisan 2001.
- [70] Ülkü, B., Demir Eksikliği Anemisi: Klinik Hematolojinin ABC'si, Anemiler Sempozyumu, İstanbul, s. 23-32, Nisan 2001.
- [71] Öngören, Ş., Hemolitik Anemiler (Tanım, Sınıflama, Genel Tanı Prensipleri), Anemiler Sempozyumu, İstanbul, s. 49-60, Nisan 2001.
- [72] Soysal, T., Megaloblastik Anemiler, Anemiler Sempozyumu, İstanbul, s. 33-47, Nisan 2001.
- [73] Aktuğlu, G., Aplastik Anemi ve Saf Eritroid Dizi Aplazisi, Anemiler Sempozyumu, İstanbul, s. 87-98, Nisan 2001.
- [74] Apak, H., Aplastik Anemi, Anemiler Sempozyumu, İstanbul, s. 163-171, Nisan 2001.
- [75] Öngören, Ş., Kronik Hastalıklar Anemisi ve Diğer Sekonder Anemiler, Anemiler Sempozyumu, İstanbul, s. 99-105, Nisan 2001.
- [76] Internet: <http://www.bilkent.edu.tr/~bilheal>, Mayıs 2006.
- [77] Toksöz, P., Ceylan, A., Diyarbakır Bölgesinde Okul Çağı Çocuklarında Hematokrit Ölçümü ile Anemi Sıklığının Belirlenmesi, Dicle Tıp Dergisi (Journal of Medical School) C:27 S:1 2000.
- [78] Ertürk N, Süt N, Sipahioğlu F. A three years profile of patients referring to family medicine outpatient clinics. Cerrahpaşa J Med 2004; 35: 115-121.
- [79] Akpolat, T., Utaş, C., Anemi, 2004.
- [80] Internet: Demir Eksikliği, <http://www.karaman.saglik.gov.tr/egitim/demir.htm>, Ocak 2006.
- [81] Internet: Sürekli Tıp Eğitim Dergisi, www.ttb.org.tr/STED/sted1199/st11992.html - 32k, Ocak 2006.
- [82] Küçükylmaz, H., Tütüncü, Y., The relation between pulmonary hypertension, hypothyroidism and hyperthyroidism, Journal of Ankara University Faculty of Medicine, 58:3033, 2005.
- [83] Internet: <http://www.hormonlar.com/tiroid1.html>, Mayıs 2006.
- [84] Özata, M., Tiroit Hakkında Bilmeniz Gereken Her Şey, Epsilon yayınevi, 2005
- [85] Internet: <http://www.hormonlar.com/tiroid2.html>, Mayıs 2006.
- [86] Dayan C. M., Interpretation of thyroid function tests. Lancet Vol 357; pp 619-624, 2001.
- [87] Abanuz, Ü., Hipertiroidide Tiroid Kanseri İnsidansı, Taksim Eğitim ve Araştırma Hastanesi 2. Genel Cerrahi Kliniği, Uzmanlık Tezi, 2005.