

T.C.
MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ELEKTRONİK VE BİLGİSAYAR EĞİTİMİ
ANABİLİM DALI

KALP KRİZİ RİSKİNİN BİR VERİ MADENCİLİĞİ
UYGULAMASI İLE ANALİZİ

YÜKSEK LİSANS TEZİ

FATİH ELMAS

EYLÜL 2014
MUĞLA

MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü

TEZ ONAYI

FATİH ELMAS tarafından hazırlanan **KALP KRİZİ RİSKİNİN BİR VERİ MADENCİLİĞİ UYGULAMASI İLE ANALİZİ** başlıklı tezinin, 22/07/2014 tarihinde aşağıdaki jüri tarafından Elektronik ve Bilgisayar Eğitimi Anabilim Dalı'nda yüksek lisans derecesi için gerekli şartları sağladığı oybirliği/oyçokluğu ile kabul edilmiştir.

TEZ SINAV JURİSİ

Doç. Dr. İlhan TARIMER (Danışman)

Bilişim Sistemleri Mühendisliği Anabilim Dalı,
Muğla Sıtkı Koçman Üniversitesi, Muğla

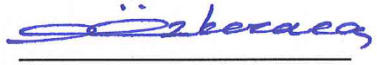
İmza:



Yrd. Doç. Dr. Osman ÖZKARACA (Üye)

Bilişim Sistemleri Mühendisliği Anabilim Dalı,
Muğla Sıtkı Koçman Üniversitesi, Muğla

İmza:



Yrd. Doç. Dr. Gürcan ÇETİN (Üye)

Bilişim Sistemleri Mühendisliği Anabilim Dalı,
Muğla Sıtkı Koçman Üniversitesi, Muğla

İmza:

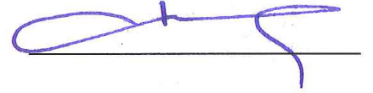


ANA BİLİM DALI BAŞKANLIĞI ONAYI

Doç. Dr. İlhan TARIMER

Elektronik ve Bilgisayar Eğitimi Ana Bilim Dalı Başkan Vekili,
Muğla Sıtkı Koçman Üniversitesi, Muğla

İmza:



Savunma Tarihi: 22/07/2014

Tez çalışmalarım sırasında elde ettiğim ve sunduğum tüm sonuç, doküman, bilgi ve belgelerin tarafımdan bizzat ve bu tez çalışması kapsamında elde edildiğini; akademik ve bilimsel etik kurallarına uygun olduğunu beyan ederim. Ayrıca, akademik ve bilimsel etik kuralları gereği bu tez çalışması sırasında elde edilmemiş başkalarına ait tüm orijinal bilgi ve sonuçlara atıf yapıldığını da beyan ederim.

Fatih ELMAS

12/09/2014

ÖZET
KALP KRİZİ RİSKİNİN BİR VERİ MADENCİLİĞİ
UYGULAMASI İLE ANALİZİ

Fatih ELMAS

Yüksek Lisans Tezi

Fen Bilimleri Enstitüsü

Elektronik ve Bilgisayar Eğitimi Anabilim Dalı

Danışman: Doç. Dr. İlhan TARIMER

Eylül 2014, 49 sayfa

Bu tez çalışmasında her geçen gün büyüyen veri tabanları içerisinde, karar verme süreçlerine etki edebilecek anlamlı bilgileri çıkarmak için kullanılan Veri Madenciliği konusu süreç ve yöntemleriyle incelenmiştir. Bu çalışmada veri madenciliğinin sağlık alanında uygulanabilirliği araştırılmış ve buna örnek teşkil edecek, Kalp krizi geçiren hastaların kan değerlerinde meydana gelen ortak değişimler tespit edilerek kalp krizini etkileyen faktörlerin tespitine çalışılmış, elde edilen sonuçlar hasta grubu üzerinde test edilerek başarımları ölçülmüştür. Bu çalışmada birliktelik kuralı algoritmalarından Apriori ve Gri algoritmaları kullanılmış, oluşturdukları kural kümelerinin başarımları üzerinden karşılaştırılarak incelenmiştir. Sonuç olarak Gri algoritmasının Apriori'ye göre daha az kural ürettiği halde yakın başarımlar gösterdiği tespit edilmiştir.

Anahtar sözcükler: Veri madenciliği, Apriori ve Gri algoritması, Kalp krizi.

ABSTRACT
ANALYSIS OF HEART ATTACK RISK WITH AN APPLICATION OF
DATA MINING

Fatih ELMAS

Master of Science (M.Sc.)

Graduate School of Natural and Applied Sciences

Department of Electronics and Computer Training

Supervisor: Assoc. Prof. İlhan TARIMER

September 2014, 49 pages

In this thesis, among databases growing everyday Data Mining topic, which is used to define meaningful information that can affect deciding process, is studied including processes and techniques. In this study, the applicability of data mining in the health field has been researched and it has been attempted to identify factors that affect heart attack by determining undergoing common changes occurring in the blood values of heart attack patients. The results obtained were measured by testing with group of patients. In this study, association rule algorithms which are Apriori and Gri algorithms have been used, and these have been examined over the success of rule sets which they create. As a result, it has been found that Gri Algorithm shows the close performance although it produces fewer rules than Apriori algorithm.

Keywords: Data mining, Apriori and Gri algorithm, Heart attack.

ÖNSÖZ

Bu tez çalışmasının planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteğini esirgemeyen, engin bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren Sayın Doç. Dr. İlhan TARIMER'e sonsuz teşekkürlerimi sunarım. Yine bu tez çalışması sürecinde, bilgi ve tecrübelerini benimle paylaşarak beni motive eden Nedim YILDIZ ve Selim YENEN'e teşekkürü bir borç bilirim.

İÇİNDEKİLER

ÖNSÖZ.....	vii
İÇİNDEKİLER.....	viii
ÇİZELGELER DİZİNİ	x
ŞEKİLLER DİZİNİ	xi
1. GİRİŞ	1
1.1. Amaç ve Kapsam	3
1.2. Sağlık Alanında Veri Madenciliği Konusunda Dünyada ve Ülkemizde Yapılmış Çalışmalar	5
2. MALZEME VE YÖNTEM.....	7
2.1. Kalp Krizi Ve Sonuçları	7
2.2. Veri Madenciliği	8
2.2.1. Veri madenciliği sürecinde problemin tanımlanması.....	10
2.2.2. Veri madenciliği sürecinde verilerin hazırlanması.....	10
2.2.2.1. Verinin toplanması	10
2.2.2.2. Birleştirme ve temizleme	11
2.2.2.3. Dönüştürme	11
2.2.3. Modelin kurulması ve değerlendirilmesi	11
2.2.4. Modelin kullanılması	13
2.2.5. Modelin izlenmesi	13
2.4. Veri Madenciliği Yöntemleri.....	13
2.4.1. Kümeleme yöntemi	13
2.4.2. Sınıflama ve regresyon yöntemi.....	14
2.4.2.1. Karar ağaçları.....	15
2.4.2.2. Yapay sinir ağları	16
2.4.2.3. Genetik algoritmalar	16
2.4.2.4. Tahminleme modelleri	17
2.4.3. Birliktelik kuralları çıkarma yöntemi	17
2.4.3.1. Apriori algoritması.....	18
2.4.3.2. Gri (Generalized rule induction) algoritması	20
3. KALP KRİZİ RİSKİNİN BİR VERİ MADENCİLİĞİ UYGULAMASI İLE ANALİZİ.....	21
3.1. Verilerin İncelenmesi	21

3.2. Verilerin Düzenlenmesi.....	22
3.3. Referans Değer Aralıklarına Göre Dönüşüm İşleminin Yapılması	23
3.4. Veri Setinin İşlenmesi	26
3.5. Kural Kümeleri İçerisinde Anomali Oluşturan Durumların Tespiti	29
3.6. Bulgular	31
3.7. Elde Edilen Bulguların Test Edilmesi	32
4. SONUÇ VE ÖNERİLER	39
KAYNAKLAR	41
EKLER.....	45
Ek A. Apriori Algoritması Kullanılarak Elde Edilen Kurallar	45
Ek B. Gri Algoritması Kullanılarak Elde Edilen Kurallar	48
ÖZGEÇMİŞ	49

ÇİZELGELER DİZİNİ

Çizelge 3.1. 32 Kategorilik Veri seti bilgileri	21
Çizelge 3.2. Hasta laboratuvar verilerinden bir kesit	23
Çizelge 3.3. Kullanılan laboratuvar referans aralıkları.....	23
Çizelge 3.4. Hasta laboratuvar verilerinin referans değerlere göre dönüştürülmüş halinden bir kesit	25
Çizelge 3.5. Apriori Algoritmasına göre Hemogram ve Biyokimya testleri verilerinin değerlendirilmesi ile elde edilen bulgular	31
Çizelge 3.6 Gri Algoritması ile Hemogram ve Biyokimya testleri verilerinin birlikte değerlendirilmesi sonucu elde edilen bulgular	31
Çizelge 3.7. Apriori algoritması kullanarak elde edilen test kurallarından bir kesit	34
Çizelge 3.8. Tespit edilen hastalar ve bu hastalarda geçerli olan kurallar kesiti.....	34
Çizelge 3.9. Her bir kuralın geçerlilik görünümünden bir kesit.....	35
Çizelge 3.10. Tespit edilen hastalarda geçerli olan kural istatistiği	36
Çizelge 3.11. Gri algoritması kullanılarak elde edilen test kurallarından bir kesit	36
Çizelge 3.12. Her bir kuralın geçerlilik görünümünden bir kesit.....	37
Çizelge 3.13. Tespit edilen hastalarda geçerli olan kurallar istatistiği	37

ŞEKİLLER DİZİNİ

Şekil 2.1 Veri madenciliği süreci.....	9
Şekil 2.2. Verilerin kümelenmesi; Küme içi ve kümeler arası uzaklıklar	14
Şekil 3.1. Verileri referans değerlerine göre dönüştüren program kodundan bir kesit.....	25
Şekil 3.2. SPSS Clementine programında Hasta verilerine Aprori ve Gri Algoritmalarının uygulanması.....	27
Şekil 3.3. Apriori ve Gri algoritmalarının Güven ve Destek değerleri.....	28
Şekil 3.4. Apriori algoritması ile elde edilen kurallardan bir kesit.....	29
Şekil 3.5. Anomali olan kuralları tespiti akış şeması.....	30
Şekil 3.6. Elde edilen bulguların test edilmesi	33

1. GİRİŞ

İçinde bulunduğumuz çağ, bilgi çağıdır. Her geçen gün bilginin miktarı artmaktadır ve bu hızlı gelişme içerisinde bilgisayar ve internet teknolojilerinin yeri çok büyüktür. Geçmişin kütüphane raflarını dolduran kitaplar dolusu bilgi, bilgisayarlarda küçük alanlarda saklanabilmekte ve paylaşılabilmektedir. Bilişim teknolojilerinin ilerlemesinin getirdiği veri saklama kolaylığından yararlanan ticari kuruluşlar her türlü faaliyetlerine dair verilerini, eğitim, adalet, sağlık ve benzeri resmi kurumlar bütün işlemlerine dair kayıtlarını elektronik ortam da tutmaya başlamışlardır. Bu kayıtları bilgisayar hafızasında salt olarak tutmak büyük bir kolaylık olarak görülmüş ise de her geçen gün miktarı artan bu kayıtları arama ve düzenlemede yaşanan erişim sorunları Veri Tabanı ve Yönetim Sistemleri ile çözülmeye çalışılmaktadır.

Her geçen gün büyüyen Veri Tabanları içerisinde Veri Tabanı yöneticileri için, karar verme süreçlerine etki edecek anlamlı bilgileri çıkarmak hiç şüphesiz değerli olduğu kadar zor bir iştir. Bilgisayar teknolojilerinden, istatistik biliminden ve programcı zekâsından yararlanarak yapılan çalışmalar sonucu Veri Tabanlarında tutulan ham veri den bilgi keşfetme süreci geliştirilmiş; bunun adına da “Veri Madenciliği” denilmiştir.

Veri madenciliği, veriyi farklı yönlerden analiz etme ve onu kullanışlı bilgi elde etmek üzere özetleme süreci olarak tanımlanabilir (Şentürk, 2011). Veri madenciliği ile eldeki verilerden çok net olmayan ve önceden bilinmeyen ancak potansiyel olarak kullanışlı bilgiler çıkarılır. Veriler arasındaki ilişkiyi, kuralları ve özellikleri belirlemekten bir bilgisayar yazılımı sorumludur. Amaç, daha önceden fark edilmemiş veri desenlerini tespit edebilmektir (Vahaplar ve İncoğlu, 2000).

Veri madenciliği, eldeki yapısız veriden, anlamlı ve kullanışlı bilgiyi çıkarmaya yarayacak tümevarım işlemlerini formüle etmek, analiz etmek ve uygulamaya

yönelik çalışmaların tümünü içerir. Geniş veri kümelerinden desenleri, değişiklikleri, düzensizlikleri ve ilişkileri çıkarmakta kullanılır (Baykasoğlu, 2004).

Veri madenciliği uygulamalarında mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modeller kolaylıkla çıkarılır. Bu bağlamda, veri madenciliği insan merkezlidir ve bazen insan – bilgisayar ara yüzü birleştirilir (Tiryaki, 2006).

Günümüzde Veri Madenciliği; bilim ve mühendislik, sağlık, ticaret, alışveriş, bankacılık ve finans, eğitim ve internet gibi pek çok farklı alanda kullanılmaktadır (Azimli, 2011). İnsanların ortalama yaşam sürelerinin uzaması, beraberinde bazı sorunları da getirmiştir. Kronik hastalıkların görülme sıklığı ve buna paralel olarak getirdiği mali yük giderek artan bir seyir izlemektedir. Bu noktadan hareketle, kronik hastalıkların ortaya çıkmasını engelleyecek önleyici çözümler geliştirilmesi gerekmektedir (Obenshain, 2004). Bu hastalıkların hem tıbbi açıdan hem de hastane kaynak ve maliyetleri açısından ele alınarak doğru yönetilmesi gerekmektedir. Bu noktada bilgi sistemleri üzerinde çalıştırılabilecek klasik sorgulama yöntemleri yeterli kalmamaktadır. Hastane bilgi sistemleri hastalara ait demografik bilgiler, hastalık ve tedavi durumları, yapılan tetkikler, faturalama ve idari işlere ait bilgileri içerirler. Sağlık ve tıp, çağımızın en önemli bilimsel araştırma alanları olduğu için bu alandaki bilgi sistemleri de araştırmalar için en büyük veri kaynaklarıdır (Yıldırım vd., 2008).

Veri Madenciliğinin; sağlık hizmetlerinin sunumu, her düzeydeki sağlık kurumlarının yönetimi ve sağlık politikalarının oluşturulmasında bir karar destek aracı olarak kullanılması sağlık profesyonellerinin en uygun kararları almasına yardımcı olacaktır (Koyuncugil ve Özgülbaş, 2009). Özellikle tarama testlerinden elde edilen veriler kullanılarak çeşitli kanserlerin ön tanısı, kalp krizi riskinin tespiti, acil servislerde hasta semptomlarına göre risk ve önceliklerin tespiti gibi çok geniş bir uygulama sahası söz konusudur (Baykasoğlu, 2004).

Bu yüksek lisans tez çalışmasında veri madenciliği yöntemleri incelenmiş, birliktelik kuralları yöntemi ile kalp krizi geçirmiş hastalara ait veriler birliktelik kuralları yöntemi ile incelenerek; kalp krizini tetikleyen risk faktörleri tespit edilmeye çalışılmış, Veri Madenciliğinin sağlık alanında uygulanmasının faydaları ortaya konmuştur.

Bu çalışmanın 2. Bölümünde çalışmaya temel teşkil eden konular; Kalp krizi ve sonuçları ile Veri madenciliği konusu incelenmiştir; 3. Bölümünde kalp krizi geçiren hastalara ait veriler kullanılarak kalp krizi riskinin bir veri madenciliği uygulaması ile analizi gerçekleştirilmiş; 4. Bölümünde ise elde edilen sonuçlar ortaya konulmuş ve ileride yapılacak çalışmalar için önerilerde bulunulmuştur.

1.1. Amaç ve Kapsam

Sağlık sektörü bilginin içerik ve yapısal anlamda en hızlı değiştiği alanlardandır. Sağlık hizmetlerinin en hızlı, en doğru, en yüksek kalitede ve ihtiyaca cevap verecek şekilde sunulabilmesi için, sağlık çalışanları en doğru ve güncel bilgilere ulaşmalıdır. Karar vermeye yardımcı olacak bu bilgiye, karar verme sürecinde kullanıcının sistemle karşılıklı olarak etkileşimde bulunduğu, bilgisayar tabanlı bir bilişim sistemi olan karar destek sistemleri içerisinde kullanılan veri madenciliğinden faydalanılarak ulaşılabilir. (Delen vd., 2011)

Sağlık alanında veri madenciliğinin uygulanabileceği başlıca alanlar şunlardır;

- Kronik hastalıklar için erken uyarı sinyallerinin veri madenciliği ile tespiti
- Klinik karar destek sistemlerinin geliştirilmesi
- Laboratuvar testleri için hata tespiti
- Suistimallerin ve yolsuzluklarının tespiti
- Sağlık personelinin performansının izlenmesi
- Hasta akış planlarının yapılması
- Tıbbi tedavi süreçlerinin optimizasyonu (klinik rehber)
- İlaç kullanım hata ve yan etkileri için erken uyarı sinyallerinin belirlenmesi
- Veri Madenciliğine dayalı olarak hasta ve ilaç kullanımının profilendirilmesi ve Türkiye ilaç kullanım haritasının hazırlanması
- Kronik hastalıklarda veri madenciliğine dayalı olarak ilaç kullanım alışkanlıkları ve risk tespiti (Koyuncugil ve Özgülbaş, 2009).

Veri Madenciliği büyük miktarda veri içinden, gizli kalmış, değerli, kullanılabilir bilgileri açığa çıkarmak ve stratejik karar desteği sağlamak amacıyla kullanılan; verilerin analizini temel alarak karar verme modelleri oluşturan bir yöntemdir. Bu nedenle sağlık hizmetlerinin sunumu, her düzeydeki sağlık kurumlarının yönetimi ve sağlık politikalarının oluşturulmasında bir karar destek aracı olarak veri madenciliğinin kullanılması sağlık çalışanlarının en uygun kararları almasına yardımcı olacaktır.

Kronik hastalıklar için erken uyarı sinyallerinin tespit edilmesi gereklidir. Ortalama yaşam süresinin artışıyla beraber, kronik hastalıkların görülme sıklığı ve buna paralel olarak getirdiği mali yük giderek artan bir seyir izlemektedir. Bu noktadan hareketle, kronik hastalıkların ortaya çıkmasını engelleyecek öngörüye dayalı çözümler geliştirilmesi gerekmektedir (Obenshain, 2004). Bu amaçla her bir kronik hastalığa yönelik; sosyal, ekonomik, demografik, coğrafi vb. tüm değişkenler dikkate alınarak, hastalığın ortaya çıkışında etkisi olan faktörler tespit edilebilir. Bunun için Sınıflama ve Regresyon veya Birliktelik Kuralları yöntemleri kullanılarak hastaların ortak özellikleri veya hastalığa etki eden faktörler ve bu faktörlerin sınır değerleri tespit edilerek; hastalığın ortaya çıkışına işaret edebilecek risk sinyalleri geliştirilebilir.

Klinik karar destek sistemlerinin geliştirilmesinde hastalığın teşhis ve tedavisi için hekime yardımcı olacak; hekime veri, çözüm, risk ve önerileri sunabilecek sistemler geliştirilmesi için veri madenciliği uygulamalarına ihtiyaç duyulmaktadır. Bunun için; sayısallaştırılmış hasta kayıtları üzerinde çalışan, hekimin isteğine uygun yöntemi teorik karmaşaya girmeden çalışıp sunabilen sistemler tasarlanmalıdır. Bir başka deyişle, amaca uygun Veri Madenciliği yöntemlerinin kullanıcı grafik ara yüzünün arkasına gizlendiği karar destek sistemleri geliştirilmelidir. Böyle bir akıllı sistem, birçok veriye anında ulaşabilme, değişkenleri anında ilişkilendirip analiz edebilme, kolay kullanım ve erken uyarı özelliklerine sahip olacaktır.

Ülkemizde Sağlık Bakanlığı tarafından yapılan değerlendirmelere göre; sağlık alanında politika üretmek için hayati öneme sahip verilerin toplanmasında, saklanmasında ve analiz edilmesinde, ulusal veya uluslararası standartların olmadığı, özellikle veri toplama konusunda ciddi bir karmaşanın mevcut olduğu tespit edilmiştir. Çözüm olarak; Sağlık bakanlığı 2008 yılında başlattığı “Sağlıkta Dönüşüm Programı” kapsamında “Karar Sürecinde Etkili Bilgiye Erişim: Sağlık

Bilgi Sistemi” başlığı ile birtakım çalışmalara ivme kazandırılmıştır. Ulusal Sağlık Veri Sözlüğü, Minimum Veri Setleri, Sağlık Kodlama Referans Sunucusu ve sağlık verilerinin toplandığı Elektronik Sağlık Kaydı (ESK) veri tabanı ve Karar Destek Sistemi bileşenleri bu çalışmaların kapsamını oluşturmaktadır (Akdağ vd., 2008). Toplanan bu verileri, analiz amaçlı göstergelere dönüştürmek ve karar vermede yol gösterici modeller geliştirmek için Veri Madenciliği çözümlerine başvurulması kaçınılmazdır (Koyuncugil ve Özgülbaş, 2009).

1.2. Sağlık Alanında Veri Madenciliği Konusunda Dünyada ve Ülkemizde Yapılmış Çalışmalar

Sağlık alanında hastaların elektronik tıbbi kayıtları, laboratuvar test sonuçları ve hastanın demografik bilgileri gibi verileri kullanarak birçok veri madenciliği araştırması yapılmıştır. Bu çalışmalardan yurtdışında yapılan bazıları şunlardır;

Vinnakota ve Lam (2006) “Amerika Birleşik Devletleri’nde sosyoekonomik farklılıkların kanser ölümlerine etkisine veri madenciliği ile yaklaşım” adındaki çalışmalarında birliktelik kuralı algoritmalarını kullanarak kanserden ölen kimselerin karakteristiklerini belirlemeye çalışmışlardır. İnceledikleri 426 vakadan 3002 birliktelik kuralı elde etmişler. Bu kuralları yorumlayarak en fazla; düşük eğitim oranı, yüksek işsizlik ve düşük ücretli işlerde çalışma ilişkilerini bulmuşlardır.

Kaur vd. (2006) Göğüs kanseri olmuş 103 hastaya ait yaş, cinsiyet, hipertansiyon ve şeker gibi biyomedikal verileri kullanarak, sınıflama yöntemlerinden olan karar ağacı yöntemleri ile analiz etmişler ve göğüs kanseri olup olmadıklarını tahmin etmişlerdir.

Palaniappan ve Awang (2008) yaptıkları veri madenciliği çalışmasında, Sınıflama yöntemlerinden olan karar ağaçları, yapay sinir ağları ve Naive Bayes’i kullanarak kalp hastalarına ait 13 faktörü değerlendirmiş ve kalp hastalarını % 86 oranında doğru tahmin etmişler. En yüksek başarılı sonuçları Naive Bayes sonra yapay sinir ağları sonra da karar ağaçları yöntemini kullanarak elde etmişler.

Krishniah vd. (2012) biyomedikal 17 veriyi kullanarak kalp krizi belirtilerinin neler olduğunu araştırmışlardır. K-mean, Nearest Neighbor ve Entropy algoritmalarını kullanarak, Kalp krizi geçiren hastaları kümeleme yöntemi ile tespit etmeye

alışmışlar. Entropy tabanlı algoritma ile hastaları % 82,9 doğruluk oranında tespit etmişler.

Saęlık alanında veri madencilięi konusunda lkemiz araştırmacılarının da eşitli alışmaları vardır. Bu alışmalara örnek verilecek olursa;

Aksoy (2009) dalgıçların vurgun vakalarının analizi iin bir veri madencilięi uygulaması gerekleştirmiş, bu alışmada kümeleme yöntemi k-ortalama, COBWEB, EM algoritmaları ile uygulanmıştır. Dalgıçların acil durum bilgi aęının dalış yaralanmaları bildirim formlarından elde edilen belirti listeleri kullanılarak vurgun vakaları sınıflandırılmış ve sonuçlar klasik sınıflandırma yöntemleri, yeni yapılan istatistiksel sınıflandırma yöntemleri ve tedavi sonuçları ile karşılaştırılmıştır. Ayrıca teşhiste yardımcı olabilecek birliktelik kuralları elde edilmiştir. Sonuç olarak, kümeleme yöntemleriyle elde edilen sınıfların yeni yapılan istatistiksel sınıflandırmalarla ve klasik sınıflandırmalarla uyumlu olduęu ve hafiften şiddetli vakalara giden hiyerarşik yapıda olduęu gözlemlenmiştir.

Doęan ve Türkoęlu (2008) tarafından yapılan bir alışmada, demir eksikliği anemisi hastalığı iin temel belirleyicileri olan Serum demiri, Serum demir bağlama kapasitesi ve Ferritin enzimleri faktörleri kullanılarak demir eksikliği anemisi teşhisinde, hekime yardımcı bir karar destek sistemi geliştirilmiştir. Sınıflama yöntemlerinden olan karar aęaçları kullanılarak Anemi(+) ve Anemi(-) deęerlendirmesinde bulunmuştur. Tasarlanan karar destek sisteminin sonuçları, doktorun verdięi kararlarla karşılaştırılmış tamamen örtüştüęü görölmüştür.

Kökver vd. (2014) tarafından yapılan bir alışmada hipertansiyona etki eden faktörler, sınıflandırma yöntemlerinden Naive Bayes ve Karar aęaçları algoritmalarını kullanarak araştırılmıştır. Hasta ve saęlam kimseleri ieren bir veri tabanında Yaş, Cinsiyet, Vücut Kütle İndeksi, HDL, LDL, Trigliserid, Ürik Asit ve Sigara Kullanımı olmak üzere 8 faktör incelenerek hipertansiyon hastaları ve saęlam kimseler ayırt edilmiştir.

2. MALZEME VE YÖNTEM

2.1. Kalp Krizi Ve Sonuçları

Kalp krizi, kalbin koroner arterlerinde gerçekleşen bir bozukluk sonrası meydana gelen yetersizlik sonucu, şiddetli göğüs ağrısıyla ortaya çıkan ve ölümlle sonuçlanması olası fizyolojik duruma denir. Kalp krizi; kalbi besleyen damarlarda kan akımının çeşitli nedenlerle ani azalmasına veya kesilmesine bağlı olarak oluşan oksijen kesintisi sonucu ortaya çıkar. Tıkanan damarın beslediği kalp kasında çeşitli derecede hasara veya dokunun ölümüne neden olabilmektedir (Ateş, 2010).

Kalp Krizini netice veren kalp-damar sistemi arızaları değişik türlerde olabilir. Bu türler arasında en fazla görülen ve en çok ölüme neden olan koroner arter hastalığıdır.

Kalp içerisinde her dakikada 5 litre kadar kan geçer ve kalbimiz bu kanı vücudun bütün organlarına gönderir. Bu anlamda kalbimiz bir tür pompa görevi görür. Kalbimiz pompa görevini yaparken her dakikada ortalama 70-80 kez kasılır ve gevşer. Dakikada 75 kez kasılıp gevşeme hareketi yapan kalp bir saatte 4500 kez, bir günde 100 bin kez, bir yılda ise 40 milyon kez çalışır. Bu kadar yoğun çalışan bir organın da beslenmesi gerekir. Kalbin içinden her dakika litrelerce kan geçmesine karşın kalp bu kan ile beslenmez. Kalbin kendi besleyici damarları vardır ve kalp bu damarlardan gelen kan ile beslenir. Kalbi besleyen damarlara koroner damar adı verilir. Bu damarların daralması veya tıkanması sonucu ortaya çıkan hastalığa koroner arter hastalığı denir (Bilir, 2008).

Koroner arter hastalığının değişik klinik şekilleri olabilir ve bu hastalığın başlıca belirtisi göğüste hissedilen ağrıdır. Göğüs ağrısı çoğu kez efor durumunda ortaya çıkar. Bazen fazla yemeği takiben veya soğuk havalarda da göğüs ağrısı olabilir. Koroner arter hastalığı Kalp spazmı ve Kalp krizinin başlıca sebebidir. Kalbi besleyen koroner damarların dışında başka damarlarda oluşabilecek plaklardaki

kırılmalar sonucu kana karışan parçalar da pıhtı oluşturarak kalp damarlarını tıkayıp kalp krizine neden olabilir.

Kalp krizi geçirenlerin %50-60'ı ilk saatlerde vefat etmekte, sağ kalanların ise %2-55'i hastaneden çıktıktan sonraki bir yıl içinde yaşamlarını yitirmektedir. Bu yönüyle kalp krizi insan hayatı için ciddi bir tehlikedir. Bir hastalığın toplumsal önemi ise hastalığın görülme sıklığı (insidans) ile ilgilidir. 20-64 yaş arası erkeklerde bir yıllık sürede Kalp krizi geçirme olasılığı binde 1,7 ile 7,3 arasında değişmektedir. Aynı olasılık kadınlar için binde 0,2-1,6 arasındadır. Yani ortalama olarak her yıl 20-64 yaş arası 1000 erkeğin 5 tanesi, 1000 kadının da bir tanesi Kalp krizi geçirmektedir. Gelişmiş ülkelerde ölümlerin yaklaşık yarısı, gelişmekte olan ülkelerde ise 1/5'i kalp hastalıkları nedeni ile olmaktadır (Bilir, 1988).

Kalp krizi sonucu ölüm, dünyada ve ülkemizde en başta gelen ölüm sebeplerindendir. Dünya Sağlık Örgütü'nün 2012 yılındaki kalp ve damar hastalıkları araştırması sonucuna göre; 2030 yılında dünyada ölümlerin önde gelen sebebinin kalp ve damar hastalıkları olacağını öngörülmüştür. Bu araştırmaya göre 2030 yılında ölümlerin 23 milyonu geçeceğini ve bunun büyük bir bölümünün kalp krizi ve inme kaynaklı olacağını öngörmektedir.

Türk Kardiyoloji Derneği ve Sağlık Bakanlığı'nın 1990 yılında başlattığı TEKHARF (Türk Erişkinlerinde Kalp Hastalığı ve Risk Faktörleri) çalışmasının izlem verilerine göre ise Türkiye'de 2 milyona yakın koroner kalp hastasının olduğu tahmin edilmektedir. Bu rakamın önümüzdeki 10 yılda iki kat artış göstererek 5-6 milyona ulaşması beklenmektedir. Ülke genelinde yılda 260 bin civarında kalp krizi vakası yaşanmakta ve bunların 85 bini ölümle neticelenmektedir.

Yapılan bu araştırmalar neticesi görüldüğü üzere kalp hastalıkları ve kalp krizine karşı önleyici tedbirler alınması toplum için büyük önem taşımaktadır.

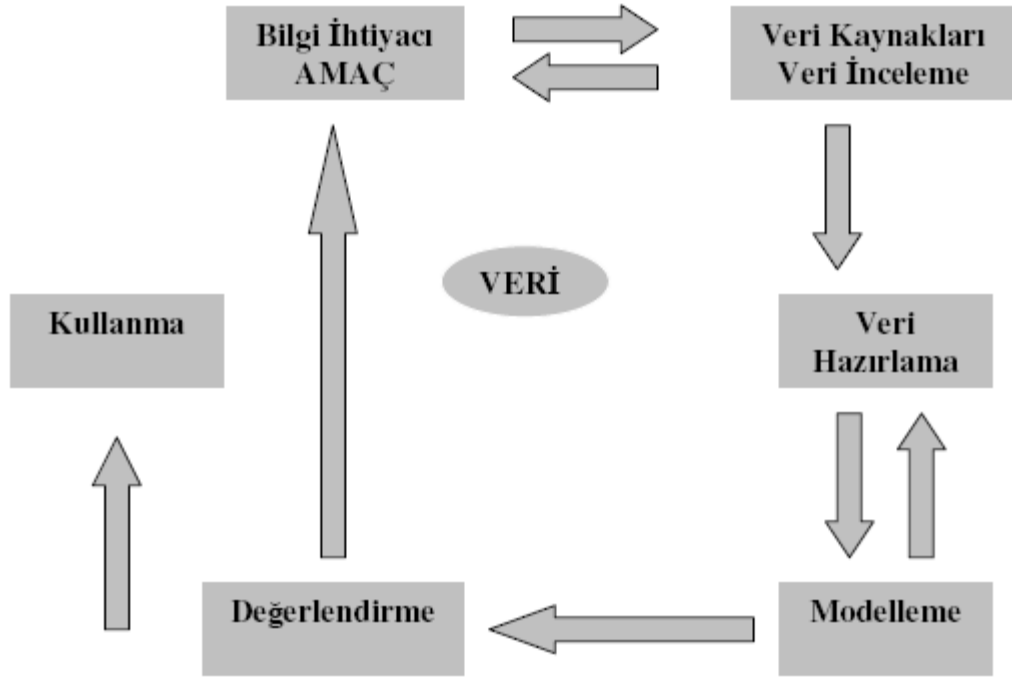
2.2. Veri Madenciliği

Veri madenciliği, büyük hacimli veri yığınları içerisinde karar alabilmek için potansiyel olarak faydalı olabilecek, uygulanabilir ve anlamlı bilgilerin çıkarılmasına verilen addır. Veri madenciliği geniş anlamda veri analiz teknikleri bütünüdür ve tek

başına bir çözüm değildir. Mevcut problemleri çözmek, kritik kararları almak veya geleceğe yönelik tahminleri yapmak için gerekli olan bilgileri elde etmeye yarayan bir araçtır. Ortaya çıkarılması hedeflenen bilgiler; üstü kapalı, çok net olmayan, önceden bilinmeyen, daha önce keşfedilmemiş ancak potansiyel olarak kullanışlı, anlamlı ve kritik bilgilerdir (Argüden ve Erşahin, 2008).

Veri madenciliği çalışmalarında başarılı olmanın ilk şartı, uygulamanın hangi amaç için yapılacağını açık bir şekilde tanımlanmasıdır. İlgili işletme amacı, işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalıdır. 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 (Ayık, 2007).

Veri madenciliği projelerinde önce problem tanımlanır; ardından veriler hazırlanır; model kurulur, değerlendirilir ve kullanılır, sonra da bu model izlenir (Argüden ve Erşahin, 2008). Şekil 2.1’de veri madenciliği işleyiş süreci gösterilmiştir.



Şekil 2.1 Veri madenciliği süreci (Topaloğlu, 2007)

2.2.1. Veri madenciliği sürecinde problemin tanımlanması

Veri madenciliği çalışmalarında başarılı olmak için, geliştirilecek uygulamanın hangi işletme amacı için nasıl yapılacağı açık bir şekilde tanımlanmalıdır. İlgili işletme amacı işletme problemi üzerine odaklanmış ve açık bir dille ifade edilmiş olmalı, elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceği tanımlanmalıdır (Döşlü, 2008). Bu aşamada üretilen sonucun fayda - maliyet analizinin başka bir deyişle üretilen bilginin işletme için değerinin doğru analiz edilmesi gerekmektedir.

2.2.2. Veri madenciliği sürecinde verilerin hazırlanması

Veri madenciliğinin en önemli aşamalarından biri olan verinin hazırlanması aşaması analistin toplam zaman ve enerjisinin çoğunu harcamasına neden olmaktadır. Bu aşamada veri hazırlama işini yapanların, mevcut bilgi sistemleri üzerinde ürettiği sayısal bilgiyi iyi analiz etmesi, veriler ile mevcut iş problemi arasında ilişki kurması gerekir. Geliştirilecek uygulama kapsamında kullanılacak sayısal verilerin seçimi yapılırken geleneksel kalıplardan çıkılmalıdır. Çünkü veri madenciliğinde insanın görevi problemin çözümünü bulmak değil, aksine verinin kendi kendine bu sonuçlara yaklaşmasına yardım etmektir. Bu yönüyle veri madenciliği süreci, uzman kişinin eğilimlerinden korunmalıdır. Aksi halde başarılı sonuçlara ulaşmak zorlaşacaktır (Telcioğlu, 2007).

Veri madenciliği uygulaması için kullanılacak verilerin hangi süreçler ile oluşturulduğu da bu veriler kullanılmadan analiz edilmeli ve analist veri kalitesi hakkında fikir sahibi olmalıdır. Verilerin hazırlanması aşaması kendi içerisinde toplama, birleştirme ve temizleme, dönüştürme adımlarından meydana gelmektedir.

2.2.2.1. Verinin toplanması

Verinin toplanması tanımlanan problem için gerekli verilerin ve bu verilerin toplanacağı veri kaynaklarının belirlenmesi adıımıdır. Verilerin toplanmasında kuruluşun kendi veri kaynaklarının dışında, nüfus sayımı, hava durumu, merkez bankası kara listesi gibi veri tabanlarından veya veri pazarlayan kuruluşların veri tabanlarından da faydalanılabilir (Tiryaki, 2006).

2.2.2.2. Birleřtirme ve temizleme

Bu adımda toplanan verilerde bulunan farklılıklar giderilmeye alıřılır. Hatalı veya analizin yanlış ynlenmesine sebep olabilecek verilerin temizlenmesine alıřılır. Genellikle yanlış veri giriřinden veya bir kereye zg bir olayın gerekleřmesinden kaynaklanan verilerin, nemli bir uyarıcı enformasyon ierip iermedięi kontrol edildikten sonra veri kmesinden atılması tercih edilir. Toplanan verilerdeki sorunların giderilmesi basit ve bařtan savma řekilde yapılmamalıdır; aksi halde ileride yanlış kurallar veya hatalı sınıflamalar gibi sorunlar ortaya ıkabilir (Topaloęlu, 2007).

2.2.2.3. Dnřtrme

Kullanılacak model ve algoritma erevesinde verilerin tanımlama veya gsterim řeklinin de deęiřtirilmesi gerekebilir. rneęin ticari bir alandaki veri madencilięi uygulamasında mřterilerin alıřtıęı sektr, gelir aralıęı ve yař aralıęı gibi deęiřkenlerin gruplanması faydalı olacaktır.

2.2.3. Modelin kurulması ve deęerlendirilmesi

Bu safhada eřitli modelleme teknikleri seilip, uygulanır ve parametreler en uygun deęerlere ayarlanır. Tanımlanan problem iin en uygun modelin bulunabilmesi, olabildięince ok sayıda modelin kurularak denenmesi ile mmkndr. Bazı teknikler belli veri formatlarına ihtiya duymaktadır. Bu nedenle model kurma ařaması, en iyi olduęu dřnlen modele varılıncaya kadar yinelenir ve genellikle veri hazırlama safhasına geri dnřler gerekir.

İyi kurulmuř bir model, analiz sonucunda elde edilecek sonuların kalitesini de etkileyecektir. İyi bir veri madencilięi uygulayıcısı, analiz sonucunda hangi rntlerin bulunabileceęini tahmin edebilmelidir. Model kurma sreci, analiz iin hangi verilerin elde hazır bulunduęunu kullanıcıya sunar. Eęer model doęru kurulmazsa, veri seti ierisinde bulunabilecek kritik iliřkiler doęru bir řekilde sunulamaz ve nemli rntler tespit edilemez. Dolayısıyla modelden bařarılı sonu elde etme olasılıęı da azalır (Berry ve Linoff, 2000).

Modeli işletip sonuçları elde etmeye başlamadan önce, modelin kalitesini ve geçerliliğini test etmek gereklidir. Çünkü model kuruluđu çalışmaları sonucuna bağılı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiğı değışik modeller kurulabilir. Model kuruluş çalışmalarına başlamadan önce, hangi tekniğın en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere denemeler yapılmasında yarar bulunmaktadır.

Modeli değıerlendirirken önemli bir kriter modelin anlaşılabilirliğidir. Genelde verilen kararların niçin verildiğinin yorumlanabilmesi büyük önem taşır. Sözelimi; kaldırıp oranı ve grafiğı, bir modelin sağladığı faydanın değıerlendirilmesinde önemli bir yardımcı gösterge olabilir. Mesela kredi kartını muhtemelen iade edecek müşterilerin belirlenmesi amacını taşıyan bir uygulamada, kullanılan modelin belirlediğı 100 kişinin 35'i gerçekten bir süre sonra kredi kartını iade ediyorsa ve tesadüfi olarak seçilen 100 müşterinin aynı zaman diliminde sadece 5'i kredi kartını iade ediyorsa kaldırıp oranı 7 olarak hesaplanmaktadır. Ancak kurulan modelin doğruluk derecesi ne denli yüksek olursa olsun, gerçek dünyayı tam anlamı ile modellediğini garanti edebilmek mümkün değıildir.

Bir diğeri değıerlendirme kriteri ise öngörü fonksiyonlarından sınıflandırma fonksiyonunda geçerli olan, hata oranlarının kalite göstergesi olarak kullanımıdır. Bu durumda elimizdeki verinin bir kısmı modelin öğrenilmesi, diğeri kısmı ise modelin geçerliliğinin test edilmesi için kullanılır. Modelin öğrenilmesi, öğrenim kümesi kullanılarak gerçekleştirildikten sonra, test kümesi ile modelin doğruluk derecesi belirlenir. Modelde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır. Eğer sınırlı miktarda veri varsa; bu durumunda kullanılabilir bir yöntem, çapraz geçerlilik testidir. Bu yöntemde veri kümesi rasgele iki eşit parçaya ayrılır. İlk aşamada bir parça üzerinde model eğitimi ve diğeri parça üzerinde test işlemi; ikinci aşamada ise ikinci parça üzerinde model öğrenimi ve birinci parça üzerinde test işlemi yapılarak elde edilen hata oranlarının ortalaması kullanılır (Göral, 2007).

2.2.4. Modelin kullanılması

Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir. Kurulan modeller risk analizi, dolandırıcılık tespiti, müşteri alışkanlıklarının tespiti veya hastalık risk faktörlerinin tespiti gibi uygulamalarda doğrudan kullanılabilir. İhtiyaç duyulan stok düzeyleri yeniden sipariş noktasının altına düştüğünde, otomatik olarak sipariş verilmesini sağlayacak bir uygulamanın içine de gömülebilir.

2.2.5. Modelin izlenmesi

Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde değişiklikler ortaya çıkabilir. Bu durumda kurulan modellerin genel olarak izlenmesi ve muhtemelen yeniden düzenlenmesi gerekebilir. Bu noktada tahmin edilen ve gözlenen değişkenler arasındaki farklılığı gösteren grafikler, model sonuçlarının izlenmesinde kullanılabilecek yararlı göstergelerdir.

2.4. Veri Madenciliği Yöntemleri

Veri madenciliği ile bilgi keşfi çalışmaları başladığı günden beri pek çok yöntem geliştirilmiştir. Bu yöntemler üç başlık altında alt kısımlarda açıklanmaktadır.

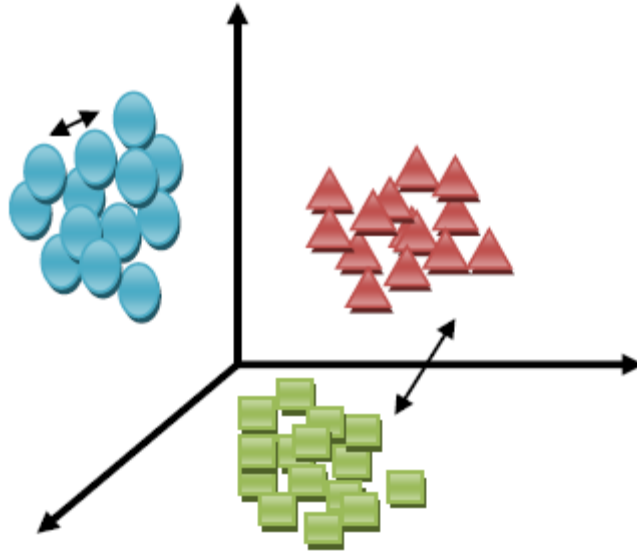
2.4.1. Kümeleme yöntemi

Kümeleme analizi, gruplanmamış bir veri topluluğunu birbirlerine olan benzerliklerine göre gruplara ayırıp sınıflamaktır. Kümeleme analizin de, veriler incelenerek birbirlerine benzeyenler bir kümeye, benzemeyenler ise bir başka kümeye toplanırlar. Bu sınıflama sayesinde veri araştırmacıya uygun, işe yarar özetleyici bilgiler biçimine dönüşür (Özdamar, 1999).

Verilerin benzerliğinden kasıt ise aralarındaki mesafenin ölçülmesi ve değerlendirilmesidir. Bu değerlendirme, veri tabanındaki diğer verilere kıyasla iki

verinin ne kadar yakın ya da benzer oldukları açısından yapılabileceği gibi önceden belirlenmiş kısıtlar eşik değerleri çerçevesinde de yapılabilir.

Kümeleme analizinde nesneler (birimler), sahip oldukları karakteristik özellikler temel alınarak gruplanır. Kümeleme analizi, nesneleri küme içerisinde çok benzer biçimde, kümeler arasında ise farklı olacak biçimde kümeler. Kümeleme işlemi başarılı olursa, bir geometrik çizim yapıldığında nesneler küme içerisinde birbirine çok yakın, kümeler ise birbirinden uzak olurlar (Hair, 1998). Şekil 2.2.'de başarılı bir kümeleme gösterilmektedir.



Şekil 2.2. Verilerin kümelenmesi; Küme içi ve kümeler arası uzaklıklar

Kümeleme analizi birçok disiplin tarafından kullanılmaktadır. İstatistik, bilgisayar ve matematik bilimi bunların en önde gelenleridir. Bilgisayar biliminde ses, karakter ve resim tanınması ve makine öğrenmesi kümeleme analizine dayanan çalışmalarla sağlanır. Ayrıca web sitelerinin benzerliklerinden yola çıkarak, onları gruplama ve benzer grupları da birbirleriyle ilişkilendirerek arama sonuçlarında beraber getirebilme imkânlarını ortaya çıkarmıştır.

2.4.2. Sınıflama ve regresyon yöntemi

Mevcut verilerden hareket ederek geleceğin tahmin edilmesinde faydalanan ve veri madenciliği teknikleri içerisinde en yaygın kullanıma sahip olan bir yöntemdir.

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 (Alpaydın, 2000).

Amaç bir şeyin özellikleri ile başka bir şeyin özellikleri arasında ilişki kurmaktır. Sınıflama ve Regresyon yöntemlerini kullanarak veriler üzerinde bir model kurmak mümkündür. Bu kurulan model ile daha önceden yapılan ayrıştırımların sebeplerini açıklamak veya yeni verileri sınıflandırmak imkânı doğar. Böylelikle geçmişe ait verilerden yola çıkarak ileriye dönük tahminler yapılabilir. Bir sigorta şirketi geçmişte sigortalarını devam ettirmeyen müşterilerinin özelliklerini analiz ederek bir risk grubu profili oluşturabilir. Bu profil model olarak kullanılarak yeni müşterilerin risk grubuna dahil olup olmadığı test edilebilir (Topaloğlu, 2007).

2.4.2.1. Karar ağaçları

Karar ağaçları sınıflandırma ve tahmin etmede kullanılan bir tekniktir. Kural çıkarma algoritmalarına sahiptir ve çıkarılan if-else biçimindeki kurallardan bir ağaç türetirler. İstatistikte, makine öğrenmesi başlığı altında birçok karar ağacı algoritması geliştirilmiştir. Karar ağaçları akış şemalarına benzeyen yapılardır. Her bir nitelik bir düğüm tarafından temsil edilir. Dallar ve yapraklar ağaç yapısının elemanlarıdır.

Karar ağaçlar, bir dizi soru sorup bunların cevapları doğrultusunda hareket ederek en kısa sürede sonuca gider. Karar ağaçları, sorduğu bir soruya gelen cevap ile soracağı diğer soruları belirler. Eğer sorular iyi seçilmiş olursa, yeni gelen bir kaydı sınıflandırılması işlemi, en az sayıda soru sorarak gerçekleştirilebilir. Sorulacak sualler ve bunlara gelebilecek cevapların yönlendirdiği başka suallerin bulunduğu bir ağaç yapısı olarak adlandırılan karar ağaçları ile değerlendirme yaparken, yeni gelen bir kayıt ağacın kökünden giriş yapar. Kökte test edilen bu yeni kayıt, yapılan testin sonucuna göre bir alt düğüme gönderilir. Bu süreç, yeni kayıt herhangi bir yaprak düğüme gelene kadar devam eder. Ağacın belirli bir yaprağına gelen bütün yeni kayıtlar aynı şekilde sınıflandırılırlar. Kökten her bir yaprağa giden sadece tek bir yol

vardır. Bu yol, kayıtları sınıflandırmak için kullanılan bir kuralı tanımlamaktadır. Bazı yapraklar aynı sınıflandırmayı yapabilirler fakat her bir yaprak bu sınıflandırmayı farklı nedenlere dayanarak yapar (Göral, 2007).

2.4.2.2. Yapay sinir ağları

Yapay sinir ağları (YSA), insan beyninin bilgi işleme teknolojisinden esinlenerek geliştirilmiş bir bilgiişlem teknolojisidir. YSA ile basit biyolojik sinir sisteminin çalışma şekli simüle edilir (benzetilir). Simüle edilen sinir hücreleri nöronlar içerirler ve bu nöronlar çeşitli şekillerde birbirlerine bağlanarak ağı oluştururlar. Bu ağlar öğrenme, hafızaya alma ve veriler arasındaki ilişkiyi ortaya çıkarma kapasitesine sahiptirler. Diğer bir ifadeyle, YSA'lar, normalde bir insanın düşünme ve gözlemlemeye yönelik doğal yeteneklerini gerektiren problemlere çözüm üretmektedir. Bir insanın, düşünme ve gözlemleme yeteneklerini gerektiren problemlere yönelik çözümler üretebilmesinin temel sebebi ise insan beyninin ve dolayısıyla insanın sahip olduğu yaşayarak veya deneyerek öğrenme yeteneğidir.

YSA'lar, ağırlıklandırılmış şekilde birbirlerine bağlanmış birçok işlem biriminden (nöronlar) oluşan matematiksel sistemlerdir. Bir işlem birimi, aslında sık sık transfer fonksiyonu olarak anılan bir denklemdir. Bu işlem birimi, diğer nöronlardan sinyalleri alır; bunları birleştirir, dönüştürür ve sayısal bir sonuç ortaya çıkartır. Genelde, işlem birimleri kabaca gerçek nöronlara karşılık gelirler ve bir ağ içinde birbirlerine bağlanırlar; bu yapı da sinir ağlarını oluşturmaktadır (Elmas, 2011).

2.4.2.3. Genetik algoritmalar

Genetik algoritmalar, doğal seçim ilkelerine dayanan bir arama ve optimizasyon yöntemidir. Temel ilkeleri John Holland tarafından ortaya atılmıştır (Holland, 1992). Temel ilkelerinin ortaya atılmasından sonra, genetik algoritmalar hakkında birçok bilimsel çalışma yayınlanmıştır. Ayrıca, genetik algoritmaların teorik kısmı ve uygulamaları hakkında birçok uluslararası konferans da düzenlenmektedir. Genetik algoritmaların, fonksiyon optimizasyonu, çizelgeleme, mekanik öğrenme, tasarım, hücresel üretim gibi alanlarda başarılı uygulamaları bulunmaktadır. Geleneksel optimizasyon yöntemlerine göre farklılıkları olan genetik algoritmalar, parametre

kümesini değil kodlanmış biçimlerini kullanırlar. Olasılık kurallarına göre çalışan genetik algoritmalar, yalnızca amaç fonksiyonuna gereksinim duyar. Çözüm uzayının tamamını değil belirli bir kısmını tararlar. Böylece, etkin arama yaparak çok daha kısa bir sürede çözüme ulaşırlar. Diğer bir önemli üstünlükleri ise çözümlerden oluşan popülasyonu eş zamanlı incelemeleri ve böylelikle yerel en iyi çözümlere takılmamalarıdır (Emel ve Taşkın, 2002).

2.4.2.4. Tahminleme modelleri

Bu veri madencilik fonksiyonu bir nesneler kümesinin belirli özelliklerinin değer dağılımını veya bazı eksik verilerin muhtemel değerlerini tahmin eder. İlgili özelliğine bağlı bazı istatistiksel analizler gibi özellikler kümesini bulmayı ve seçilen nesneler benzer veri kümesine dayalı veri dağılımını tahmin etmeyi içerir (Küçüksille, 2009).

2.4.3. Birliktelik kuralları çıkarma yöntemi

Birliktelik kuralları çıkarmanın amacı, büyük veri kümeleri arasında birliktelik ilişkilerini bulmaktır. Depolanan verilerin sürekli artması nedeniyle kurumlar, veri tabanlarındaki birliktelik kurallarını ortaya çıkarmak isterler. Büyük miktarda depolanan verilerden değişik birliktelik ilişkileri bulmak, kurumların karar alma süreçlerini olumlu etkilemektedir (Küçüksille, 2009).

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 eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır (Alpaydın, 2000).

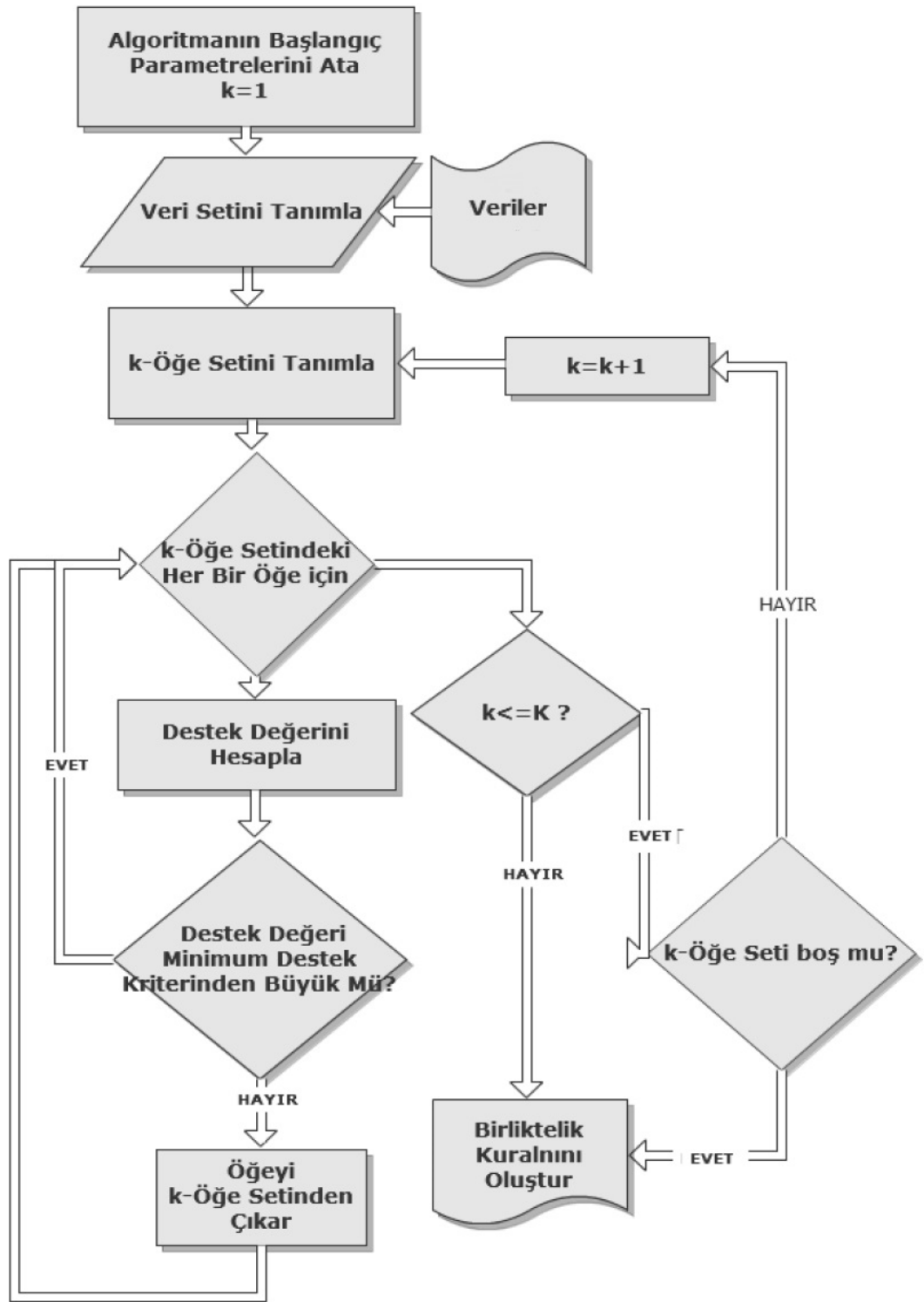
Bu kapsamda bazı birliktelik kuralları çıkarma algoritmalarından bahsedilmesi yerinde olacaktır:

2.4.3.1. Apriori algoritması

Literatürde birliktelik kuralı çıkaran değişik algoritmalar bulunmaktadır. Apriori Algoritması, birliktelik kuralı çıkarım algoritmaları içerisinde en fazla bilinen algoritmadır (Agrawal vd., 1993). Bu algoritmada sık geçen öge kümelerini bulmak için birçok kez veri tabanını taramak gerekir. İlk taramada bir elemanlı minimum destek metriğini sağlayan sık geçen öge kümeleri bulunur. İzleyen taramalarda bir önceki taramada bulunan sık geçen öge kümeleri aday kümeler adı verilen yeni potansiyel sık geçen öge kümelerini üretmek için kullanılır. Aday kümelerin destek değerleri tarama sırasında hesaplanır ve aday kümelerinden minimum destek değerini sağlayan kümeler o geçişte üretilen sık geçen öge kümeleri olur. Sık geçen öge kümeleri bir sonraki geçiş için aday küme olurlar. Bu süreç yeni bir sık geçen öge kümesi bulunmayana kadar devam eder (Alataş, 2003).

Bu algoritmada temel yaklaşım “eğer k -öge kümesi minimum destek değerini sağlıyorsa bu kümenin alt kümeleri de minimum destek değerini sağlar” şeklindedir. Öge küme, 1 veya daha fazla elemandan oluşan kümedir. k -öge küme (k -itemset) ise içinde k adet öge bulunan kümedir (Ay ve Çil, 2008).

Apriori algoritması, $(k+1)$ sık geçen öge kümesini bulmak için k sık geçen öge kümesine ihtiyaç duyar. Sık geçen öge kümeleri bulmak için ilk olarak minimum destek kriterini sağlayan 1-sık geçen öge kümeleri bulunur, 2- sık geçen öge kümeleri bulmak için 1- sık geçen öge kümeleri, 3- sık geçen öge kümeleri bulmak için 2- sık geçen öge kümeleri kullanılır. Bu süreç, algoritma sık geçen öge kümeleri bulamayınca kadar devam eder (Özseven ve Düğenci, 2011). Şekil 2.3.’de Apriori algoritmasının akış şeması görülmektedir.



Şekil 2.3. Apriori algoritması akış diyagramı

Apriori algoritmasının çalışması sonucunda elde edilen her bir kural (öğeler arasındaki birliktelik), destek (support) ve güven (confidence) kriterleri ile ifade edilir. Destek kriteri, öğeler arasındaki birlikteliğinin sıklığını, güven kriteri ise bu birlikteliklerin doğruluğunu ifade etmektedir.

2.4.3.2. Gri (Generalized rule induction) algoritması

Gri algoritması 1992 yılında Smyth ve Goodman tarafından geliştirilmiştir. Aslında bu algoritma ITRule algoritmasına dayanır ve onun bazı fonksiyonlar eklenerek geliştirilmiş halidir (Smyth ve Goodman, 1992) .

Gri bilgiye ulaşmak için uyguladığı yaklaşım aday birliktelik kuralları içerisinde farklılıkları ölçmek için nicel bir ölçü (J) kullanır. Gri bu ölçüyü (J) her bir kuralın mümkün olan sınır değerlerden farklılığını ölçerek hesaplar. Böylelikle anlamlı kuralları arama alanı daralmış olur. Gri en yüksek bilgi içeren kuralları bulmak için oluşturduğu (J) endeksine göre aday birliktelik kurallarını özetler ve güven ve destek değerlerini dikkate alarak sıralar. Bunun sonucu olarak da daha az sayıda ve daha anlamlı kurallar seçilmiş olur.

Gri algoritmasının diğer birliktelik kuralı algoritmalarına nazaran üstünlüğü; anlamlı olmayan kuralları anlamlı olanlar içerisinden elemesidir. Bu yönüyle birliktelik kuralı algoritmalarının en önemli sorunu olan, çok sayıda kural içeren kural kümesi içerisinden anlamlı kuralların bulunmasını kolaylaştırır. Gri daha az sayıda ve daha kaliteli kurallar bulur.

Zaten birliktelik kuralı algoritmalarının sonucu olarak beklenen ve istenilen; önemli ve anlamlı kurallar içeren ve anlamsız kurallar içermeyen bir kural kümesidir. Bu özelliklerde bir kural kümesi değerlendirme ve sonuç çıkarma sürecinde büyük değer ifade eder (Aggelis ve Christodoulakis, 2004).

3. KALP KRİZİ RİSKİNİN BİR VERİ MADENCİLİĞİ UYGULAMASI İLE ANALİZİ

3.1. Verilerin İncelenmesi

Bu uygulamada Muğla’da özel bir hastanenin anjiyo birimine kalp krizi tanısıyla sevk edilen hastaların teşhis aşamasında elde edilen laboratuvar verileri kullanılmıştır. Veriler Microsoft Excel veri tabanında saklanmış ve Microsoft Visual Basic kullanılarak düzenlenmiştir. Veri tabanında toplam 153 kalp krizi vakası bulunmaktadır. Veri madenciliği araştırması içerisinde; bu vakalardan 103’ü öğrenme grubu olarak, kural çıkarmak için, 50’si ise elde edilen kuralları test etmek için kullanılmıştır. Bu veriler hemogram, biyokimya, Ck-Mb ve sedimentasyon testi sonuçları ile yaş ve cinsiyet verilerinden oluşmaktadır. Veri setimizdeki bilgilerin toplamı otuz iki kategoriden oluşmaktadır (Çizelge 3.1.).

Çizelge 3.1. 32 Kategorilik Veri seti bilgileri

HEMOGRAM DEĞERLERİ
WBC (White Blood Cells)
LYMP %
MONO %
NEUT %
NEUT#
GRAN %
GRAN#
RBC (Red Blood Cells)
HGB (Hemoglobin)
HCT (Hematokrit) %
MCV (Mean Corpuscular Volume)
MCH (Mean Corpuscular Hemoglobin)
MCHC (Mean Corpuscular Hemoglobin Concentration)

Çizelge 3.1.'in devamı

PLT (Platelets)
MPV (Mean Platelet Volume)
PCT (Prokalsitonin)
PDW (Platelet Distribution Width)
BİYOKİMYA DEĞERLERİ
AÇLIK KAN SEKERE
KOLESTEROL
HDL KOLESTEROL
LDL KOLESTEROL
TRİGLİSERİD
ÜRE (BUN)
KREATİNİN
SODYUM
ÜRİK ASİT
POTASYUM
ALT (SGPT)
CK-MB
SEDİMENTASYON

Çizelge 3.1.'de Hemogram başlığı altında gösterilen değerler; kan hücrelerinin nitelik ve nicelik açısından incelenmesi sonucu ortaya çıkan değerleri, Biyokimya başlığı altındaki değerler; kanda bulunan kimyasalların oranlarını, Ck-Mb verisi Kalp krizi geçiren hastalarda kalp kaslarındaki yıpranmaya işaret eden CK izoenziminin oranını, Sedimentasyon verisi ise kandaki çökecek oranını ifade etmektedir.

3.2. Verilerin Düzenlenmesi

Kalp hastalıkları anjiyo biriminden toplanan tüm hasta verileri incelendiğinde eksik bilgi içeren kategoriler, veri madenciliği açısından bir değer ifade etmemektedir. Hatta bu eksik veri içeren kategoriler yanlış bulgulara sebep olabileceğinden bu sorunu aşmak için %40 ve üzeri eksik veri içeren kategoriler göz ardı edilmiştir. Eksik veri oranı %40'ın altında olan kategorilerde ise eksik verinin yerine o kategorinin ortalama değeri verilmiştir. Çizelge 3.2.'de veri tabanından bir kesit gösterilmektedir.

Çizelge 3.2. Hasta laboratuvar verilerinden bir kesit

	Sediment	Wbc	Lymp	Mono	Neut	Neut#	Lymp	Mono	Rbc	Hgb
1	24,000	10,40	9,000	6,200	80,244	10,998	0,900	0,600	4,80	12,70
2	19,000	20,20	10,600	2,200	86,900	17,500	2,100	0,400	4,72	15,60
3	2,000	12,80	5,800	1,600	92,300	11,800	0,700	0,200	4,96	15,20
4	3,000	9,300	28,900	8,900	80,244	10,998	2,700	0,800	4,66	13,30
5	5,000	7,800	17,400	4,900	74,300	5,800	1,400	0,400	5,36	15,50
6	78,000	15,40	18,900	12,000	67,000	10,300	2,900	1,900	4,43	13,80
7	3,000	9,600	13,100	5,400	79,700	7,700	1,300	0,500	4,78	14,90
8	5,000	12,10	17,500	5,300	76,100	9,200	2,100	0,600	4,67	15,40
9	48,000	9,600	13,600	2,400	83,300	8,000	1,300	0,200	4,45	13,60
1	6,000	15,60	11,100	4,000	84,500	13,200	1,700	0,600	5,91	16,70

3.3. Referans Değer Aralıklarına Göre Dönüşüm İşleminin Yapılması

Veri tabanına aktarılan veriler, eksik bilgilerin tamamlanması ve hatalı sonuçlara sebep olacak bilgiler elimine edilerek elde kalan bu bilgilerin sağlıklı bir bireyin sahip olması gereken kan değerlerinde olup olmadığı test edilmiştir.

Kan değerlerini analiz eden her laboratuvarın kullandığı donanıma ve tekniğe göre referans aralıkları değişebilmektedir. Bu nedenle yapılan bu çalışmada hasta verilerinin alındığı hastane laboratuvar referans aralıkları kullanılmıştır.(Çizelge 3.3.)

Çizelge 3.3. Kullanılan laboratuvar referans aralıkları

HEMOGRAM DEĞERLERİ	REFERANS ARALIĞI
WBC (White Blood Cells)	4 < >11
LYMP %	20 < > 50
MONO %	2 < > 10
NEUT %	50 < >80
NEUT#	2< >8
GRAN %	35< >80
GRAN#	1.4< >7
RBC (Red Blood Cells)	4< >6.2
HGB (Hemoglobin)	11< >18.8
HCT (Hematokrit) %	35< >55
MCV (Mean Corpuscular Volume)	80< >100
MCH (Mean Corpuscular Hemoglobin)	26< >34

Çizelge 3.3.'ün devamı

MCHC (Mean Corpuscular Hemoglobin Concentration)	31< >37
PLT (Platelets)	130< >400
MPV (Mean Platelet Volume)	6< >10
PCT (Prokalsitonin)	0.108 < >0.282
PDW (Platelet Distrubition Width)	15< >17
BİYOİMYA DEĞERLERİ	
AÇLIK KAN SEKERİ	70< >110
KOLESTEROL	220<
HDL KOLESTEROL	35< >65
LDL KOLESTEROL	150<
TRİGLİSERİD	40< >160
ÜRE (BUN)	10< >50
KREATİNİN	0.4< >1.4
SODYUM	135< >145
URİK ASİT	2.4< >7
POTASYUM	3.6< >5
ALT (SGPT)	0< >50
CK-MB	25<
SEDİMENTASYON	20<

Eğer bu değerler normal düzeyde ise N, olması gereken referans değerlerinin altındaysa A ve referans değerlerden yüksek ise Y harf şekline dönüştürülür. Bu dönüşüm, uygun program kodları aracılığıyla yapılmıştır. Şekil 3.1.'de program kodlarından bir kesit görülmektedir.

```
Private Sub CommandReferans_Click()
    Sheets("Sayfa3").Select
    Range("b3").Select
    While (ActiveCell.Value <> "")
        'WBC REFERANS DEĞERLERİ 4 < >11 ARALIĞI
        i = ActiveCell.Offset(0, 1).Value
        If (i > 11) Then i = "Y" Else If (i < 4) Then i = "A" Else i = "N"
        ActiveCell.Offset(0, 1).Value = i
    End While
End Sub
```

```

'LYPMP% REFERANS DEĞERLERİ 20 <> 50 ARALIĞI
i = ActiveCell.Offset(0, 2).Value
If (i > 50) Then i = "Y" Else If (i < 20) Then i = "A" Else i = "N"
ActiveCell.Offset(0, 2).Value = i

'MONO% REFERANS DEĞERLERİ 2 <> 10 ARALIĞI
i = ActiveCell.Offset(0, 3).Value
If (i > 10) Then i = "Y" Else If (i < 2) Then i = "A" Else i = "N"
ActiveCell.Offset(0, 3).Value = i

'NEUT% REFERANS DEĞERLERİ 50 <>80 ARALIĞI
i = ActiveCell.Offset(0, 4).Value
If (i > 80) Then i = "Y" Else If (i < 50) Then i = "A" Else i = "N"
ActiveCell.Offset(0, 4).Value = i

.
.
.

'ARALIKLARA GÖRE YAŞ DÖNÜŞÜM
i = ActiveCell.Offset(0, 30).Value
If i < 50 Then ActiveCell.Offset(0, 30).Value = "ilk"
If i >= 50 Then ActiveCell.Offset(0, 30).Value = "orta"
If i >= 70 Then ActiveCell.Offset(0, 30).Value = "ileri"

ActiveCell.Offset(1, 0).Select ' Bir sonraki kayda git

Wend
End Sub

```

Şekil 3.1. Verileri referans değerlerine göre dönüştüren program kodundan bir kesit

Dönüşüm işlemlerinin ardından veri tabanının yeni hali, çizelge 3.4.'de gösterilmiştir.

Çizelge 3.4. Hasta laboratuvar verilerinin referans değerlere göre dönüştürülmüş halinden bir kesit

	Sediment.	Wbc	Lymp %	Mono %	Neut %	Neut#	Lymp #	Rbc	Hgb	Hct %	Mcv
1	Y	N	A	N	Y	Y	N	N	N	N	N
2	N	Y	A	N	Y	Y	N	N	N	N	N
3	N	Y	A	A	Y	Y	N	N	N	N	N
4	N	N	N	N	Y	Y	N	N	N	N	N
5	N	N	A	N	N	N	N	N	N	N	N
6	Y	Y	A	Y	N	Y	N	N	N	N	N
7	N	N	A	N	N	N	N	N	N	N	N
8	N	Y	A	N	N	Y	N	N	N	N	N
9	Y	N	A	N	Y	N	N	N	N	N	N
10	N	Y	A	N	Y	Y	N	N	N	N	N

Çizelge 3.4. 'de gösterilen hasta verilerinin cinsiyet ve yaş bilgileri dönüşümü yapılmıştır. Cinsiyet için E/K kodlaması kullanılırken, yaş bilgisi 30–50 yaş aralığı “ilk”, 50–70 yaş aralığı “orta” ve 70 üzeri “ileri” olarak aralıklara göre gruplanmıştır.

3.4. Veri Setinin İşlenmesi

Veri madenciliği araştırmaları yapmak amacıyla geliştirilmiş çok sayıda paket program vardır. Bunların başlıcaları; SPSS Clementine, Matlab, Rapid Miner, Weka, Orange ve diğerleridir.

Bunlar arasında SPSS tarafından geliştirilmiş SPSS CLEMENTINE, veri madenciliği uygulamaları için görsel bir modelleme aracı olarak kullanılmaktadır. Clementine pek çok farklı veri kaynağı ile çalışmaya olanak sağlayan, modelleri hızlı bir şekilde oluşturma ve karşılaştırma olanaklarıyla ön plana çıkan bir yazılım olduğundan dünya üzerinde Veri Madenciliği uzmanlarının en fazla tercih ettiği yazılımlardan birisidir. Bu nedenle bu tez çalışmasında veri madenciliği yapmada SPSS CLEMENTINE kullanımı tercih edilmiştir.

SPSS CLEMENTINE programına, daha önceden temizleme, birleştirme ve dönüşüm işlemlerine tabi tutulan “hastalar veri tabanını” dâhil edebilmek için öncelikle veriler virgülle ayrılmış değerler dosyası formatına (CSV) çevrilmiştir. Ardından Type nesnesi ile veri setindeki değişken türleri atandı. Ardından Filter nesnesi ile hangi verilerin analizinin yapılacağı hangilerinin göz ardı edileceği tercihleri yapıldı. Gri ve Apriori nesneleri ise Gri ve Apriori algoritmalarının veri setimize uygulanması için kullanıldı. Web nesnesi ise veriler arasındaki ilişkiler ağının yapısını gözlemlemek için kullanıldı. Şekil 3.5’de Clementine kullanıcı ara yüzündeki uygulama gösterilmiştir.

CİNSİYET

Model name: ☒ Auto ☐ Custom

☒ Use partitioned data

Minimum antecedent support (%):

Minimum rule confidence (%):

Maximum number of antecedents:

Maximum number of rules:

☐ Only true values for flags

Fields **Model** Annotations

Şekil 3.3. Apriori ve Gri algoritmalarının Güven ve Destek değerleri

Güven (Confidence) değeri kuralın geçerliliğini ifade etmektedir; yani A durumu ortaya çıktığında B durumunun da ortaya çıkma sıklığı en az %80'dir. Destek (Support) değeri ise A ve B durumunu sağlayan hastaların sayısının, bütün hasta kayıtları içerisindeki en az değerini oransal olarak gösterir ki bu oran, çalışmada %20'dir. Bu şartları sağlayan en fazla beş öncülle kurallar üretilmesi istenilmiş ve şekil 3.4.'de gösterilen sonuçlar elde edilmiştir.

CINSİYET			
File Generate			
Sort by: Confidence %			
5328 of 5328			
Consequent	Antecedent	Support %	Confidence %
CINSİYET = E	NEUTD = Y RBC = N MCH = N	62,14	100,0
CINSİYET = E	NEUTD = Y RBC = N MCV = N	62,14	100,0
CINSİYET = E	NEUTD = Y RBC = N Hgb = N MCH = N	62,14	100,0
CINSİYET = E	NEUTD = Y RBC = N Hgb = N MCV = N	62,14	100,0
CINSİYET = E	NEUTD = Y RBC = N HctY = N MCH = N	60,19	100,0
CINSİYET = F	NEUTD = Y		
Model	Summary	Annotations	OK

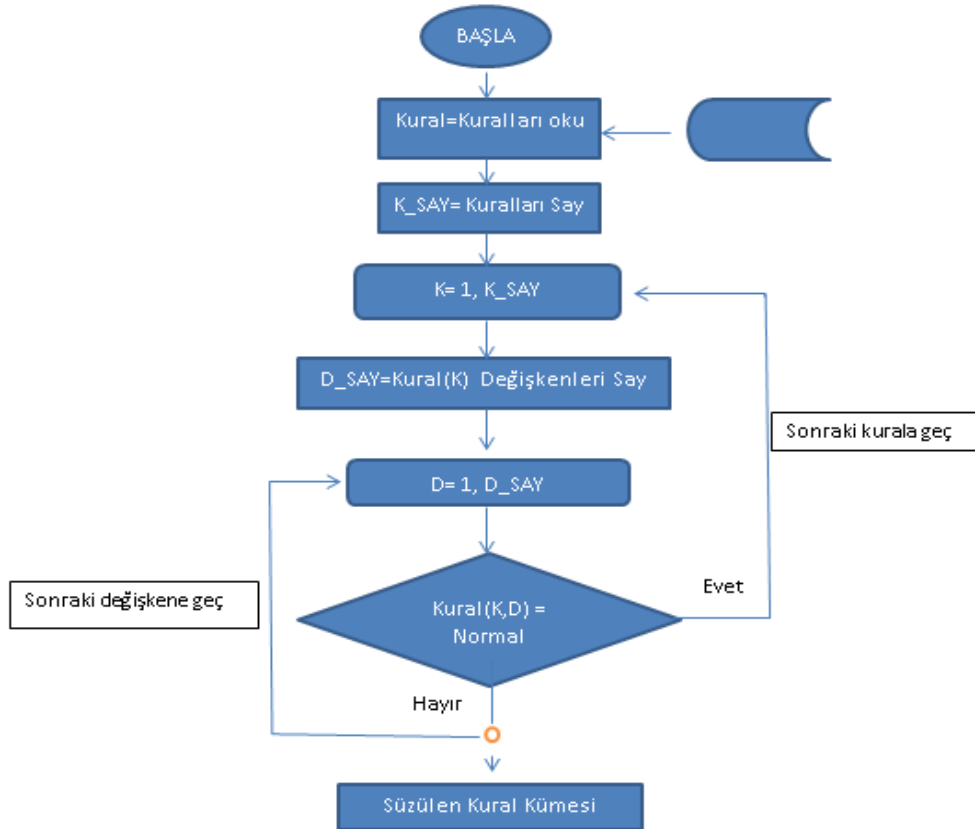
Şekil 3.4. Apriori algoritması ile elde edilen kurallardan bir kesit

Kalp krizi geçiren hastaların kan değerlerinde ortaya çıkan ortak anormalliklerin tespitine çalışan bu uygulamada, hemogram ve biyokimya testlerine ait veriler ayrı ayrı süzgeçlenerek ve birlikte olarak farklı güven ve destek değerleriyle hem Apriori hem Gri algoritmalarıyla pek çok defa işlenmiş ve en sağlıklı kural kümelerine ulaşabilmek için çalışılmıştır. Burada aranılan kurallar; birliktelik oranı ve görülme sıklığı yüksek olmanın yanında, mümkün olduğunca çok ihtimali test ederek tespit edilen kurallardır.

3.5. Kural Kümeleri İçerisinde Anomali Oluşturan Durumların Tespiti

Kalp krizi riskinin tespiti için yapılan veri madenciliği uygulamasında SPSS Clementine gerek Apriori algoritmasında gerekse Gri algoritmasında oldukça çok

sayıda örüntü tespit etmiş ve şekil 3.4’de görüldüğü gibi çok sayıda birliktelik kuralı üretmiştir. Şüphesiz bunun en önemli sebebi uygulamada her bir hastaya ait çok sayıda veri kullanılmış olmasıdır. Elde edilen kurallar incelendiğinde; hastanın kan değerlerinde görülen anormalliklerin birliktelikleri ile beraber çok sayıda normal değer birlikteliğinden oluşan kurallar göze çarpmaktadır. Bu kurallar içinde bizim için anlamlı olan kurallar anormalliklerdir. Yani kalp krizini tetikleyenler veya başka değerlerde anormalliklerin ortaya çıkmasına sebep olan değişimlerdir. Bu noktada çok sayıda kural içerisinde anlamlı olanların tespit edilmesi için elde edilen örüntüler, şekil 3.5.’de akış şeması görülen program kodları ile süzölmüş ve sadece anomali durumları tespit edilmiştir.



Şekil 3.5. Anomali kurallarını tespit eden akış şeması

3.6. Bulgular

Yapılan bu veri madenciliği uygulaması sonucu olarak Gri algoritmasıyla otuz sekiz, Apriori algoritması ile ise yüz on dokuz adet kural çıkarılmış; yani kalp krizini etkileyen veya kalp krizine gösterge teşkil eden durumlar tespit edilmiştir (Ek A ve Ek B). Elde edilen bu kural kümeleri incelendiğinde kalp krizi riskini en fazla etkileyen faktörler sırasıyla aşağıdaki çizelgelerde verilmiştir.

Çizelge 3.5. Apriori Algoritmasına göre Hemogram ve Biyokimya testleri verilerinin değerlendirilmesi ile elde edilen bulgular

NO	DEĞİŞKEN	DURUM (Referans Değerlerden)
1	HDL (KOLESTEROL)	ALÇAK
2	NEUT#	YÜKSEK
3	YAŞ	50-70 ARALIĞI
4	KOLESTEROL	YÜKSEK
5	TRİGLİSERİD	YÜKSEK
6	WBC	YÜKSEK
7	AÇLIK KAN ŞEKERİ	YÜKSEK
8	LYMP%	ALÇAK
9	CKMB	YÜKSEK
10	NEUT%	YÜKSEK

Bu uygulama içerisinde ayrı iki algoritma kullanılarak sonuçlar birbiriyle kıyaslanmak istenmiş bunun için Apriori algoritması ile yapılan analizler aynı şekilde Gri algoritması ile de tekrarlanmıştır. Elde edilen sonuçlar çizelge 3.6.'deki gibidir.

Çizelge 3.6 Gri Algoritması ile Hemogram ve Biyokimya testleri verilerinin birlikte değerlendirilmesi sonucu elde edilen bulgular

NO	DEĞİŞKEN	DURUM (Referans Değerlerden)
1	LYMP %	ALÇAK
2	TRİGLİSERİD	YÜKSEK
3	CKMB	YÜKSEK
4	HDL(KOLESTEROL)	ALÇAK

Çizelge 3.6.^(devam)

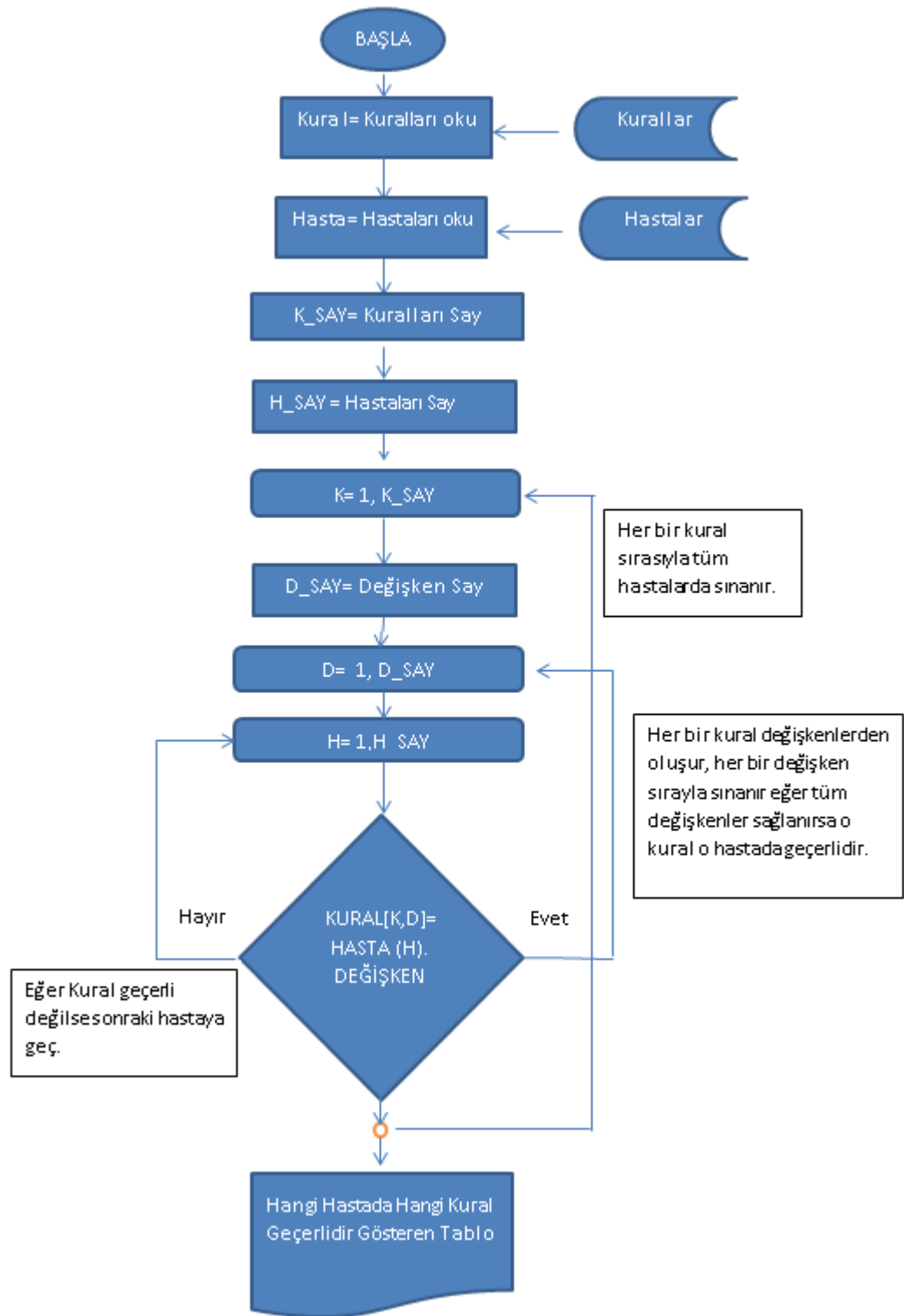
5	SODYUM	ALÇAK
6	WBC	YÜKSEK
7	NEUT #	YÜKSEK

Çizelge 3.5. ve 3.6. incelendiğinde Apriori algoritması kullanılarak elde edilen sonuçların Gri algoritması ile elde edilen sonuçları kapsadığı görülmüştür. Bu da iki algoritmanın başarımının birbirine yakınlığını göstermektedir.

3.7. Elde Edilen Bulguların Test Edilmesi

Yukarıdaki çizelgelerde gösterilen bulguların sağlıklı değerler olup olmadığı ve hastalarda yeni teşhisler için kullanılıp kullanılamayacağı test edilmesi gerekir. Bu amaçla alınan bulguların test edilmesi için bir program yazılmıştır. Programın akış şeması şekil 3.6.'da gösterilmiştir.

Bu program da öncelikle Apriori ve Gri algoritmaları ile elde edilen kurallar ayrı ayrı olarak hastalar veri tabanının da yer alan her bir hasta için sorgulanmış; kaç hastanın tespit edilebildiği, her bir kuralın kaç hastada geçerli olduğu ve her bir hastada kaç kuralın geçerli olduğu tespit edilmiştir. Elde edilen istatistiki veriler aşağıdaki çizelgelerde gösterilmektedir.



Şekil 3.6. Elde edilen bulguların test edilmesi

Çizelge 3.7. Apriori algoritması kullanarak elde edilen test kurallarından bir kesit

Kural No										
1	CİNSİYET	E	YAŞ	ORTA	CKMB	Y	LYMP%	A		
2	CİNSİYET	E	KOLESTEROL	Y	WBC	Y	NEUT#	Y		
3	CİNSİYET	E	KOLESTEROL	Y	CKMB	Y	LYMP%	A		
4	CİNSİYET	K	YAŞ	ORTA	WBC	Y	CKMB	Y	LYMP%	A
5	CİNSİYET	K	MCV	A	AKANSEKERİ	Y	SODYUM	A		
6	CİNSİYET	E	TRIGLİSERİD	Y	YAŞ	ORTA	LYMP%	A		
7	CİNSİYET	E	TRIGLİSERİD	Y	NEUT#	Y	CKMB	Y	LYMP%	A
8	CİNSİYET	E	KOLESTEROL	Y	WBC	Y				
9	CİNSİYET	E	TRIGLİSERİD	Y	CKMB	Y	LYMP%	A	SODYUM	A
10	CİNSİYET	E	TRIGLİSERİD	Y	WBC	Y	CKMB	Y	SODYUM	A

Yapılan veri madenciliği uygulaması sonucu elde edilen kuralları test eden programın tespit ettiği hastalara ait her bir hastada hangi kuralın geçerli olduğuna dair sonuçlar çizelge 3.8.'deki gibidir.

Çizelge 3.8. Tespit edilen hastalar ve bu hastalarda geçerli olan kurallar kesiti

HASTA	KURAL
Hasta 16	6
hasta 16	13
*****	Toplam 2 kural geçerli..
hasta 17	6
hasta 17	13
*****	Toplam 2 kural geçerli..
hasta 18	3
hasta 18	6
hasta 18	7
hasta 18	8
hasta 18	9
hasta 18	10

Çizelge 3.8'in devamı

hasta 18	11	
hasta 18	12	
hasta 18	13	
*****	Toplam 9	kural geçerli..

Çizelge 3.9. tespit edilen kuralların hastaların yüzde kaçında geçerli olduğunu göstermektedir. Apriori algoritmasına ait bu sonuçlarda en geçerli kuralların % 26 oranında hastada geçerli olduğu görülmektedir.

Çizelge 3.9. Her bir kuralın geçerlilik görünümünden bir kesit

1. Kural	%21
2. Kural	%26
3. Kural	%26
4. Kural	%22
5. Kural	%21
6. Kural	%24
7. Kural	%26
8. Kural	%24
9. Kural	%22
10. Kural	%24

Kalp krizi riskinin tespiti için yapılan bu veri madenciliği uygulamasından Apriori algoritması kullanılarak elde edilen örüntüler, riskli hastaların tespiti için birer kural olarak geçerliliği sınanmış; kalp krizi geçiren hasta sayısı tespitinin % 72 oranında çıktığı gözlemlenmiştir. Apriori algoritmasına göre her bir hastada kaç kuralın geçerli olduğu bilgileri, çizelge 3.10.'da verilmiştir.

Çizelge 3.10. Tespit edilen hastalarda geçerli olan kural istatistiği

Kural Sayısı	Oranı
1	%12
2	%7
3	%3
4	%22
5	%2
6	%4
7	%4
8	%4
9	%4
10	%0
Toplam Tespit Edilen Hasta	%72

Bu uygulama içerisinde ayrı iki algoritma kullanılarak sonuçlar birbiriyle kıyaslanmak istenmiş bunun için Apriori algoritması ile yapılan test işlemi aynı şekilde Gri algoritması için de tekrarlanmıştır. Elde edilen sonuçlar çizelge 3.11-12-13'deki gibidir.

Çizelge 3.11. Gri algoritması kullanılarak elde edilen test kurallarından bir kesit

Kural No										
1	CİNSİYET	E	AKANSEKERİ	Y	HDL	A	SODYUM	A		
2	CİNSİYET	E	LYMP%	A	PDW	Y				
3	CİNSİYET	E	NEUT%	Y	TRIGLİSERİD	Y	CKMB	Y		
4	CİNSİYET	E	LYMP%	A	TRIGLİSERİD	Y	SODYUM	A	CKMB	Y
5	CİNSİYET	E	WBC	Y	KOLESTEROL	Y				
6	CİNSİYET	K	MCV	A	AKANSEKERİ	Y	SODYUM	A		
7	CİNSİYET	K	MCV	A	AKANSEKERİ	Y				
8	CİNSİYET	E	LYMP%	A	TRIGLİSERİD	Y	CKMB	Y		
9	CİNSİYET	E	WBC	Y	LYMP%	A	TRIGLİSERİD	Y	CKMB	Y
10	CİNSİYET	E	WBC	Y	NEUT#	Y	KOLESTEROL	Y		

Çizelge 3.12. tespit edilen kuralların hastaların yüzde kaçında geçerli olduğunu göstermektedir. Gri algoritmasına ait bu sonuçlarda en geçerli kuralların % 30 oranında hastada geçerli olduğu görülmektedir. Bu geçerlilik oranları Apriori' ye nispeten daha yüksektir.

Çizelge 3.12. Her bir kuralın geçerlilik görünümünden bir kesit

1. Kural	%30
2. Kural	%28
3. Kural	%28
4. Kural	%28
5. Kural	%26
6. Kural	%26
7. Kural	%26
8. Kural	%26
9. Kural	%30
10. Kural	%24

Kalp Krizi Riskinin Tespiti için yapılan bu Veri madenciliği uygulamasından Gri algoritması kullanılarak elde edilen örüntüler; hastaların tespiti için birer kural olarak geçerliliği sınanmış; mevcut hasta verilerinden % 68 oranındaki bir tespit yapılabilmektedir. Her bir hastada kaç kuralın geçerli olduğunu gösteren bilgiler, çizelge 3.13.'de verilmiştir.

Çizelge 3.13. Tespit edilen hastalarda geçerli olan kurallar istatistiği

Kural Sayısı	Oranı
1	%25
2	%4
3	%10
4	%0
5	%0
6	%2
7	%4

Çizelge 3.13^(devam)

8	%8
9	%0
10	%5
11	%0
12	%10
Toplam Tespit Edilen Hasta	%68

Her iki algoritmanın hasta kümesi üzerinde test edilmesi sonucu; Gri algoritması ile elde edilen kuralların Apriori algoritması ile elde edilen kurallara göre geçerliliklerinin daha yüksek olduğu ve aynı zamanda Gri algoritmasındaki kuralların bir birine daha yakın geçerlilik değerlerine sahip olduğu çizelge 3.9. ve 3.12.'nin yorumlanmasıyla söylenebilmektedir.

4. SONUÇ VE ÖNERİLER

Bu tez çalışmasında kalp krizi geçiren hastalardan alınan veriler üzerine veri madenciliği süreç ve yöntemleri uygulanmıştır. Gerçekleştirilen çalışmada birliktelik kuralları yönteminde kullanılan Apriori ve Gri algoritmaları seçilmiş, kalp krizi geçirmiş hastaların kan değerleri incelenerek meydana gelen ortak değişimler tespit edilmiştir.

Apriori ve Gri birliktelik kuralı algoritmaları ile elde edilen kurallar incelendiğinde; Apriori algoritması ile beş binden fazla kural oluşturulurken, Gri algoritmasında bu rakam iki bin olmuştur. Bunun sebebinin, Apriori'nin çalışma mantığına uygun olarak bir, iki, üç, dört eleman içeren ve birbirinin alt kümesi durumunda olan ve pek çok defa tekrar eden, daha küçük kurallar içermesinden kaynaklanmasıdır. Çünkü Apriori'nin çalışma mantığında önce en fazla görülen tek değişkenli değişimler tespit edilir, ikinci basamakta bu tek değişken ile birlikte değişen ikinci değişkenler ve sonraki basamaklarda hep bir önceki basamakta çıkarılan kurallar ile birlikte yeni kurallar üretilmektedir. Gri algoritması ise çıkarılan kuralları genelleştirerek daha az ve daha anlamlı kurallar üretmektedir.

Yine bu veri madenciliği uygulaması sonucu elde edilen kurallardan en geçerli ve en anlamlı kurallar, hasta kümesi üzerinde sınındığı zaman Apriori algoritması ile %72, Gri algoritması ile %68 oranında kalp krizi vakasının tespit edilebildiği görülmüştür.

Tez çalışması sürecinde tüm bu sonuçlar yanında hasta bilgilerinin saklandığı veri tabanlarında özellikle hasta öyküsü bilgilerinin eksik tutulduğu gözlenmiş ve bu nedenle kalp krizi riskine etki eden veya tetikleyen hastalıklar tespit edilememiştir. Yapılan veri madenciliği uygulamasına bu verilerin de eklenmesi ile daha sağlıklı sonuçlar alınması beklenen bir durumdur.

Son yıllardaki Sağlık Bakanlığı Projeleri ile hasta bilgileri veri tabanının oluşturulması sayesinde, bu merkezi veri tabanında yapılacak veri madenciliği uygulamalarıyla kalp krizi riski gibi pek çok hastalığın erken tespiti mümkündür. Bu

genel veri tabanlarında yer alan hasta öyküsü, laboratuvar, demografik, yaşadığı coğrafi bölge, şehir merkezi veya kırsal alanda yaşam gibi veriler kullanılarak yapılacak veri madenciliği uygulamaları ile pek çok hastalığın ön sinyalleri anlamına gelen gizli örüntülerin tespit edilmesi mümkündür. Elde edilen bu ön sinyallerin, online olarak bir bilgi işlem sistemi içerisine entegre edilmesi durumunda hasta ilerde yaşayacağı hastalık için erkenden uyarılabilir. Yine bu erken uyarı sistemi hastalığın teşhisi için uzman hekimlere fayda sağlayabilir.

Bu tez çalışması ile Kalp Krizi Riskinin Tespitine fayda sağlayacak ve ön sinyaller anlamına gelen bazı örüntüler bir veri madenciliği uygulamasıyla keşfedilmeye çalışılmıştır. Yapılan bu çalışmanın genelde veri madenciliği konusuna, özelde ise veri madenciliğinin sağlık alanında uygulanabilirliğine hizmet etmesi beklenmektedir.

KAYNAKLAR

- Akdağ, R., Aydın, S., Buzgan, T., Demirel, H. ve Gündüz, F. (2008) *Türkiye sağııkta dönüşüm programı ve temel sağıık hizmetleri*, T.C. Sağıık Bakanlığı Yayınları, Ankara, 175s
- Aksoy, B. (2009) Cluster analysis of decompression illness, Yüksek Lisans Tezi, Galatasaray Üniversitesi, İstanbul, 44s
- Alataş, B. (2003) Nicel birliktelik kurallarının keşfinde bulanık mantık ve genetik algoritma yaklaşımı, Yüksek Lisans Tezi, Fırat Üniversitesi, Elazığ, 113s
- Alpaydın, E. (2000) Zeki veri madenciliğı: Ham veriden altın bilgiye ulaşma yöntemleri, Bilişim 2000 Eğitim Semineri, 6-9 Eylül 2000, İstanbul
- Aggelis, V. ve Christodoulakis, D. (2004) E-trans association rules for e-banking transactions, IV. International Conference on Decision Support for Telecommunications and Information, Eylül 2004, Warsaw Poland
- Agrawal, R., Imielinski, T. ve Swami, A. (1993) Mining association rules between sets of items in large databases, Proceedings Of The 1993 Acın Sıgmod International Conference On Management Of Data, Washington Dc (Usa)
- Arguden, Y. ve Erşahin, B. (2008) Veri madenciliğı veriden bilgiye, masraftan değere, 1. Basım, İstanbul, 73s
- Ateş, S. (2010) Coğrafi bilgi sistemleri ile kalp krizi vakalarına yönelik en uygun ambulans yerlerinin belirlenmesi, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, İstanbul, 111s
- Ay, D. ve Çil, İ. (2008) Migros Türk A.Ş. de Birliktelik Kurallarının Yerleşim Düzeni Planlamada Kullanılması, Endüstri Mühendisliğı Dergisi, 21:14-29
- Ayık, Y. Z., Özdemir, A. ve Yavuz, U. (2007) Lise türü ve lise mezuniyet başarısının, kazanılan fakülte ile ilişkisinin veri madenciliğı tekniğı ile analizi, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 10:441-454.

- Azimli, M. (2011) Tıpta veri madenciliği uygulamaları, Yüksek Lisans Tezi, Gazi Üniversitesi, Ankara, 63s
- Baykasoğlu, A. Veri madenciliği ve çimento sektöründe bir uygulama, Akademik Bilişim 2005 Konferansı, 2-4 Şubat 2005, Gaziantep
- Berry M. J. ve Linoff G. (2000) The Art and Science of Customer Relationship Management, 2. Baskı, Wiley Computer Publishing, New York, 512s
- Bilir, N.(1988) Halk Sağlığı Yönünden Koroner Kalp Hastalığı, H.Ü.T.F.Halk Sağlığı A.B.D.Yayını, Ankara
- Bilir, N. (2008) Sigara ve kalp-damar hastalıkları, T.C. Sağlık Bakanlığı Yayınları,1. Basım, Ankara,16s
- Delen, D., Sharda, R. ve Turban, E. (2011) Decision Support and Business Intelligence Systems, 9. Baskı, Prentice Hall, New Jersey (Usa), 720s
- Doğan, Ş. ve Türkoğlu, İ. (2008) Iron-Deficiency Anemia Detection From Hematology Parameters By Using Decision Trees, International Journal of Science & Technology, 3: 85-92.
- Döşlü, A. (2008) Veri madenciliğin de market sepet analizi ve birliktelik kurallarının belirlenmesi, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, İstanbul, 100s
- Erdoğan, Ş.Z. ve Timor, M. (2005) A Data Mining Application In A Student Database, Journal Of Aeronautics and Space Technologies, 2: 53-57
- Elmas, Ç. (2011) Yapay Zeka Uygulamaları, Seçkin Yayınevi, Ankara, 424s
- Emel, G.G. ve Taşkın, Ç. (2002) Genetik algoritmalar ve uygulama alanları, Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 1:129-152.
- Kour, H. ve Wasan, S.K. (2006) Empirical study on applications of data mining techniques in healthcare, Journal of Computer Science, 2:194-200

- Koyuncuđil, A.S. ve Özgölbař, N. (2009) Veri madenciligi: tıp ve sađlık hizmetlerinde kullanımı ve uygulamaları, Biliřim Teknolojileri Dergisi, 2: 21-32.
- Kökver, Y., Barıřçı, N., Çiftçi, A. ve Ekmekçi, Y. (2014) Hipertansiyona etki eden faktörlerin veri madenciligi yöntemleriyle incelenmesi, New World Sciences Academy, 9: 15-25.
- Küçüksille, E. (2009) Veri madenciligi süreci kullanılarak portföy performansının deđerlendirilmesi ve imkb hisse senetleri piyasasında bir uygulama, Doktora Tezi, Süleyman Demirel Üniversitesi, Isparta, 118s
- Krishniah, J.R., Sekar, C. ve Rao, R.H. (2012) Predicting the heart attack symptoms using biomedical data mining techniques, The International Journal of Computer Science & Applications, 1:10-18.
- Göral, M.A. (2007) Kredi kartı başvuru aşamasında sahtecilik tespiti için bir veri madenciligi modeli, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, İstanbul, 146s
- Holland, J.H. (1992) Genetic algorithms, Scientific American, 267: 66-72.
- Obenshain, M.K.(2004) Application of data mining techniques to healthcare data, Data Infect Control Hosp Epidemiol, 25: 690-695.
- Özdamar, K. (2013) Paket Programlarla İstatistiksel Veri Analizi 2, Kaan Kitabevi, Eskiřehir, 474s
- Özseven, T. ve Düğenci M. (2011) LOG Analiz: erişim kayıt dosyaları analiz yazılımı ve GOP üniversitesi uygulaması, Biliřim Teknolojileri Dergisi, 4:55-66
- Palaniappan, S. ve Awang, R. (2008) Intelligent heart disease prediction system using data mining techniques, International Journal of Computer Science and Network Security, 8: 343-350.
- Smyth, P. ve Goodman, R. M. (1992) An information theoretic approach to rule induction from databases, IEEE Transactions on Knowledge and Data Engineering, 4:301-312.

Şentürk, Z.K. (2011) Veri madenciliği ile kanser tanısı, Yüksek Lisans Tezi, Düzce Üniversitesi, Düzce, 58s

Telcioglu, M.B. (2007) Veri madenciliğinde genetik programlama temelli yeni bir sınıflandırma yaklaşımı ve uygulaması, Yüksek Lisans Tezi, Erciyes Üniversitesi, Kayseri, 135s

Tiryaki, S. (2006) Lojistik alanında bir veri madenciliği uygulaması, Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, İstanbul, 55s

Topaloğlu, F. (2007) Veri madenciliği ile meteorolojik parametrelerin analizi ve ziraî meteoroloji haritasının çıkarılması, Yüksek Lisans Tezi, Fırat Üniversitesi, Elazığ, 54s

Vahaplar, A. ve İncoğlu, M.M. (2001) Veri madenciliği ve elektronik ticaret, İstatistik araştırma sempozyumu, 12-14 Kasım 2001, Ankara

Vinnakota, S. ve Lam, N. (2006) Socioeconomic inequality of cancer mortality in the United States: a spatial data mining approach, International Journal of Health Geographics, 5: 9-21.

Yıldırım, P. Uludağ, M. ve Görür, A. Hastane Bilgi Sistemlerinde Veri Madenciliği, Akademik Bilişim 2008 Konferansı, 30 Ocak - 01 Şubat 2008, Çanakkale

EKLER

Ek A. Apriori Algoritması Kullanılarak Elde Edilen Kurallar

Kural										
1	Cinsiyet	E	Yaş	Orta	Ckmb	Y	Lymp%	A		
2	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut#	Y		
3	Cinsiyet	E	Kolesterol	Y	Ckmb	Y	Lymp%	A		
4	Cinsiyet	K	Yaş	Orta	Wbc	Y	Ckmb	Y	Lymp%	A
5	Cinsiyet	K	Mcv	A	Akansekeri	Y	Sodyum	A		
6	Cinsiyet	E	Trigliserid	Y	Wbc	Y	Ckmb	Y	Lymp%	A
7	Cinsiyet	E	Trigliserid	Y	Neut#	Y	Ckmb	Y	Lymp%	A
8	Cinsiyet	E	Kolesterol	Y	Wbc	Y				
9	Cinsiyet	E	Trigliserid	Y	Ckmb	Y	Lymp%	A	Sodyum	A
10	Cinsiyet	E	Trigliserid	Y	Wbc	Y	Ckmb	Y	Sodyum	A
11	Cinsiyet	E	Trigliserid	Y	Neut#	Y	Ckmb	Y	Sodyum	A
12	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Ckmb	Y		
13	Cinsiyet	E	Kolesterol	Y	Neut#	Y	Ckmb	Y		
14	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut#	Y	Ckmb	Y
15	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Lymp%	A		
16	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut#	Y	Ckmb	Y
17	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Lymp%	A		
18	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Ckmb	Y	Lymp%	A
19	Cinsiyet	E	Kolesterol	Y	Neut#	Y	Ckmb	Y	Lymp%	A
20	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Sodyum	A		
21	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut#	Y	Sodyum	A
22	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut#	Y		
23	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Ckmb	Y	Lymp%	A
24	Cinsiyet	E	Kolesterol	Y	Ckmb	Y	Lymp%	A	Sodyum	A
25	Cinsiyet	E	Trigliserid	Y	Neut#	Y	Ckmb	Y	Lymp%	A
26	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Wbc	Y		
27	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut%	Y		
28	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Wbc	Y	Neut#	Y
29	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Wbc	Y	Ckmb	Y
30	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut#	Y	Ckmb	Y
31	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Lymp%	A	Sodyum	A
32	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut%	Y	Neut#	Y
33	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Ckmb	Y	Sodyum	A
34	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Lymp%	A	Sodyum	A
35	Cinsiyet	E	Kolesterol	Y	Neut#	Y	Ckmb	Y	Sodyum	A
36	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut#	Y	Lymp%	A

Ek. A Apriori Algoritması Kullanılarak Elde Edilen Kurallar (devamı)

37	Cinsiyet	E	Trigliserid	Y	Neut%	Y	Ckmb	Y	Sodyum	A
38	Cinsiyet	E	Hdl	A	Sodyum	A				
39	Cinsiyet	E	Kolesterol	Y	Neut%	Y	Ckmb	Y		
40	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Wbc	Y	Lymp%	A
41	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut#	Y	Sodyum	A
42	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut%	Y	Ckmb	Y
43	Cinsiyet	E	Kolesterol	Y	Neut%	Y	Neut#	Y	Ckmb	Y
44	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Wbc	Y	Sodyum	A
45	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut%	Y	Lymp%	A
46	Cinsiyet	E	Kolesterol	Y	Sedimentasyon	Y	Wbc	Y		
47	Cinsiyet	E	Kolesterol	Y	Sedimentasyon	Y	Wbc	Y	Neut#	Y
48	Cinsiyet	E	Kolesterol	Y	Sedimentasyon	Y	Wbc	Y	Ckmb	Y
49	Cinsiyet	E	Kolesterol	Y	Sedimentasyon	Y	Neut#	Y	Ckmb	Y
50	Cinsiyet	E	Kolesterol	Y	Sedimentasyon	Y	Ckmb	Y	Lymp%	A
51	Cinsiyet	E	Trigliserid	Y	Akansekeri	Y	Ckmb	Y	Lymp%	A
52	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Sedimentasyon	Y	Lymp%	A
53	Cinsiyet	E	Trigliserid	Y	Sedimentasyon	Y	Ckmb	Y	Lymp%	A
54	Cinsiyet	E	Hdl	A	Neut#	Y				
55	Cinsiyet	E	Hdl	A	Neut#	Y	Sodyum	A		
56	Cinsiyet	E	Hdl	A	Akansekeri	Y	Sodyum	A		
57	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Akansekeri	Y		
58	Cinsiyet	E	Kolesterol	Y	Sedimentasyon	Y	Wbc	Y	Lymp%	A
59	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut#	Y	Akansekr	Y
60	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Akansekeri	Y	Ckmb	Y
61	Cinsiyet	E	Kolesterol	Y	Neut%	Y	Ckmb	Y	Lymp%	A
62	Cinsiyet	E	Kolesterol	Y	Neut#	Y	Akansekeri	Y	Ckmb	Y
63	Cinsiyet	E	Kolesterol	Y	Akansekeri	Y	Ckmb	Y	Lymp%	A
64	Cinsiyet	E	Hdl	A	Wbc	Y				
65	Cinsiyet	E	Hdl	A	Wbc	Y	Sodyum	A		
66	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut%	Y	Sodyum	A
67	Cinsiyet	E	Hdl	A	Wbc	Y	Neut#	Y		
68	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut%	Y		
69	Cinsiyet	E	Hdl	A	Wbc	Y	Neut#	Y	Sodyum	A
70	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut%	Y	Neut#	Y
71	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Akansekeri	Y	Lymp%	A
72	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Akansekeri	Y	Sodyum	A
73	Cinsiyet	E	Hdl	A	Neut%	Y				
74	Cinsiyet	E	Ldl	Y	Wbc	Y				
75	Cinsiyet	E	Hdl	A	Neut%	Y	Neut#	Y		
76	Cinsiyet	E	Hdl	A	Neut%	Y	Sodyum	A		
77	Cinsiyet	E	Hdl	A	Neut#	Y	Akansekeri	Y		
78	Cinsiyet	E	Ldl	Y	Kolesterol	Y	Wbc	Y		
79	Cinsiyet	E	Ldl	Y	Wbc	Y	Neut#	Y		
80	Cinsiyet	E	Hdl	A	Neut%	Y	Neut#	Y	Sodyum	A
81	Cinsiyet	E	Hdl	A	Neut#	Y	Akansekeri	Y	Sodyum	A
82	Cinsiyet	E	Ldl	Y	Kolesterol	Y	Wbc	Y	Neut#	Y

Ek. A Apriori Algoritması Kullanılarak Elde Edilen Kurallar (devamı)

83	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Sedimentasyon	Y	Neut#	Y
84	Cinsiyet	E	Hdl	A	Wbc	Y	Akansekeri	Y		
85	Cinsiyet	E	Ldl	Y	Ckmb	Y	Lymp%	A		
86	Cinsiyet	E	Hdl	A	Wbc	Y	Akansekeri	Y	Sodyum	A
87	Cinsiyet	E	Ldl	Y	Kolesterol	Y	Ckmb	Y	Lymp%	A
88	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Wbc	Y	Neut%	Y
89	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut%	Y	Ckmb	Y
90	Cinsiyet	E	Kolesterol	Y	Neut%	Y	Ckmb	Y	Sodyum	A
91	Cinsiyet	E	Hdl	A	Wbc	Y	Neut%	Y		
92	Cinsiyet	E	Ldl	Y	Wbc	Y	Lymp%	A		
93	Cinsiyet	E	Hdl	A	Wbc	Y	Neut%	Y	Neut#	Y
94	Cinsiyet	E	Hdl	A	Wbc	Y	Neut%	Y	Sodyum	A
95	Cinsiyet	E	Hdl	A	Wbc	Y	Neut#	Y	Akanseker	Y
96	Cinsiyet	E	Ldl	Y	Kolesterol	Y	Wbc	Y	Lymp%	A
97	Cinsiyet	E	Ldl	Y	Wbc	Y	Neut#	Y	Lymp%	A
98	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Sedimentasyon	Y	Wbc	Y
99	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut%	Y	Lymp%	A
100	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Akansekeri	Y	Lymp%	A
101	Cinsiyet	E	Kolesterol	Y	Sedimentasyon	Y	Wbc	Y	Sodyum	A
102	Cinsiyet	E	Kolesterol	Y	Wbc	Y	Neut%	Y	Akanseker	Y
103	Cinsiyet	E	Kolesterol	Y	Neut%	Y	Akansekeri	Y	Ckmb	Y
104	Cinsiyet	E	Hdl	A	Neut%	Y	Akansekeri	Y		
105	Cinsiyet	E	Ldl	Y	Wbc	Y	Ckmb	Y		
106	Cinsiyet	E	Ldl	Y	Neut#	Y	Ckmb	Y		
107	Cinsiyet	E	Hdl	A	Neut%	Y	Neut#	Y	Akanseker	Y
108	Cinsiyet	E	Hdl	A	Neut%	Y	Akansekeri	Y	Sodyum	A
109	Cinsiyet	E	Ldl	Y	Kolesterol	Y	Wbc	Y	Ckmb	Y
110	Cinsiyet	E	Ldl	Y	Kolesterol	Y	Neut#	Y	Ckmb	Y
111	Cinsiyet	E	Ldl	Y	Wbc	Y	Neut#	Y	Ckmb	Y
112	Cinsiyet	E	Pdw	Y	Lymp%	A				
113	Cinsiyet	E	Ldl	Y	Wbc	Y	Sodyum	A		
114	Cinsiyet	E	Ldl	Y	Kolesterol	Y	Wbc	Y	Sodyum	A
115	Cinsiyet	E	Ldl	Y	Wbc	Y	Neut#	Y	Sodyum	A
116	Cinsiyet	E	Ldl	Y	Wbc	Y	Lymp%	A	Sodyum	A
117	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Wbc	Y	Akanseker	Y
118	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut%	Y	Sodyum	A
119	Cinsiyet	E	Kolesterol	Y	Trigliserid	Y	Neut#	Y	Akanseker	Y

Ek B. Gri Algoritması Kullanılarak Elde Edilen Kurallar

Kural										
1	Cinsiyet	E	Akansekerı	Y	Hdl	A	Sodyum	A		
2	Cinsiyet	E	Lymp%	A	Pdw	Y				
3	Cinsiyet	E	Neut%	Y	Trigliserid	Y	Ckmb	Y	Yas	<5,5
4	Cinsiyet	E	Lymp%	A	Trigliserid	Y	Sodyum	A	Ckmb	Y
5	Cinsiyet	E	Wbc	Y	Kolesterol	Y				
6	Cinsiyet	K	Mcv	A	Akansekerı	Y	Sodyum	A		
7	Cinsiyet	K	Mcv	A	Akansekerı	Y				
8	Cinsiyet	E	Lymp%	A	Trigliserid	Y	Ckmb	Y		
9	Cinsiyet	E	Wbc	Y	Lymp%	A	Trigliserid	Y	Ckmb	Y
10	Cinsiyet	E	Wbc	Y	Neut#	Y	Kolesterol	Y		
11	Cinsiyet	E	Lymp%	A	Neut%	Y	Kolesterol	Y	Trigliserid	Y
12	Cinsiyet	E	Lymp%	A	Akansekerı	Y	Kolesterol	Y	Trigliserid	Y
13	Cinsiyet	E	Neut%	Y	Akansekerı	Y	Kolesterol	Y	Ckmb	Y
14	Cinsiyet	E	Sediment.	Y	Neut#	Y	Kolesterol	Y	Trigliserid	Y
15	Cinsiyet	E	Sediment.	Y	Lymp%	A	Yas	<6,5	Kolesterol	Y
16	Cinsiyet	E	Sediment.	Y	Wbc	Y	Lymp%	A	Kolesterol	Y
17	Cinsiyet	E	Lymp%	A	Neut%	Y	Kolesterol	Y	Ckmb	Y
18	Cinsiyet	E	Lymp%	A	Akansekerı	Y	Kolesterol	Y	Ckmb	Y
19	Cinsiyet	E	Neut%	Y	Ckmb	Y	Yas	<4,5		
20	Cinsiyet	E	Sediment.	Y	Lymp%	A	Kolesterol	Y	Trigliserid	Y
21	Cinsiyet	E	Sediment.	Y	Lymp%	A	Trigliserid	Y	Ckmb	Y
22	Cinsiyet	E	Sediment.	Y	Wbc	Y	Neut#	Y	Kolesterol	Y
23	Cinsiyet	E	Sediment.	Y	Lymp%	A	Kolesterol	Y	Ckmb	Y
24	Cinsiyet	E	Wbc	Y	Lymp%	A	Neut%	Y	Kolesterol	Y
25	Cinsiyet	E	Neut%	Y	Trigliserid	Y	Sodyum	A	Ckmb	Y
26	Cinsiyet	E	Wbc	Y	Neut%	Y	Neut#	Y	Kolesterol	Y
27	Cinsiyet	E	Lymp%	A	Neut%	Y	Trigliserid	Y	Ckmb	Y
28	Cinsiyet	E	Lymp%	A	Kolesterol	Y	Trigliserid	Y		
29	Cinsiyet	E	Lymp%	A	Kolesterol	Y	Ckmb	Y		
30	Cinsiyet	E	Wbc	Y	Lymp%	A	Kolesterol	Y		
31	Cinsiyet	E	Neut%	Y	Akansekerı	Y	Hdl	A		
32	Cinsiyet	E	Neut%	Y	Yas	<4,5				
33	Cinsiyet	E	Sediment.	Y	Wbc	Y	Kolesterol	Y		
34	Cinsiyet	E	Sediment.	Y	Neut#	Y	Kolesterol	Y	Ckmb	Y
35	Cinsiyet	E	Lymp%	A	Akansekerı	Y	Trigliserid	Y	Ckmb	Y
36	Cinsiyet	E	Neut%	Y	Kolesterol	Y	Ckmb	Y		
37	Cinsiyet	E	Wbc	Y	Neut%	Y	Kolesterol	Y		
38	Cinsiyet	E	Lymp%	A	Neut#	Y	Kolesterol	Y		

ÖZGEÇMİŞ

Kişisel Bilgiler

Ad Soyad : Fatih Elmas
Uyruk : T.C.
Doğum Yeri ve Tarihi: 16/04/1984
Medeni Hali : Bekâr
Telefon : 0 506 351 5415
E-posta : fatihcs@gmail.com

Eğitim

Lise	Şanlıurfa Anadolu Teknik Lisesi	2002
Lisans	Muğla Üniversitesi	2008
Yüksek Lisans	Muğla Sıtkı Koçman Üniversitesi	2014

İş Tecrübesi

Yıl	Yer	Pozisyon/görev
2002	Şanlıurfa	Tezgâhtar
2004	Muğla	Bilgisayar Teknikeri

Yabancı Dil(ler)

Dil (İngilizce)	Başlangıç	Orta	İleri
Yazma		X	
Konuşma	X		
Anlama	X		
Okuma		X	

Hobiler

1. Kitap Okuma 2. Gezmek 3. Yüzmek 4. Aşçılık ve yemek pişirme