Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları

Ali Serhan Koyuncugil[1], Nermin Özgülbaş[2]

[1]Sermaye Piyasası Kurulu, Araştırma Dairesi, ANKARA
[2]Başkent Üniversitesi Sağlık Bilimleri Fakültesi, Sağlık Kurumları İşletmeciliği Bölümü, ANKARA
askoyuncugil@gmail.com, ozgulbas@baskent.edu.tr

Özet— Büyük miktarda veri içerisinden, gizli kalmış, değerli, kullanılabilir bilgileri açığa çıkarmak ve stratejik karar destek sağlamak amacıyla kullanılan Veri Madenciliği; büyük miktarda veriyle ilgili sorun alanlarına yanıt bulması yanında sağlık verilerinin kullanımında yeni bir perspektif yaratmış ve kullanım yaygınlığı hızla artmaya devam eden bir yöntem haline gelmiştir. Bu makalenin amacı, sağlıkta Veri Madenciliğinin kullanımı konusunda bir altyapı oluşturmak ve sağlık profesyonellerine sağlık sektöründe Veri Madenciliği'nin kullanımı ile ilgili örnekler sunarak karar verme süreçleri açısından yeni bir bakış açısı kazandırmaktır. Bu amaçla makalede Veri Tabanlarında Bilgi Keşfi, Veri Ambarı, Veri Madenciliği, İş Zekası ve Veri Madenciliği Yöntemleri konularında kavramsal çerçeve verilerek; ülkemizdeki sağlık sektöründe öncelikli konu ve sorun alanları ile ilgili Veri Madenciliği uygulamalarına örnekler sunulmaktadır.

Anahtar kelimeler— Veri Madenciliği, Veri Tabanlarında Bilgi Keşfi, veri Ambarı, iş zekâsı, Veri Madenciliği yöntemleri, tıp, sağlık, sağlık veri, sağlık hizmetleri, elektronik hasta (tıp) dosyaları (kayıtları)

Data Mining: Using and Applications in Medicine and Healthcare

Abstract— Data mining which is define as extraction of hidden, valuable and useful knowledge from big amount of data and its use for providing strategic decision support has been giving a new perspective for usage of health data and is becoming a method which is improving in all application domains besides as an reply for problematique domains related with huge data. The objective of this paper is to form a basis and new point of view to health professionals for decision making process by presenting examples of Data Mining application in healthcare. For this purpose conceptual framework are given about Knowledge Discovery in Databases, Data Warehouse, Data Mining, Business Intelligence, and Data Mining methods and then examples of Data Mining applications about primary subjects and issues in our country's health sector are presented in this study.

Keywords— Data Mining, Knowledge Discovery in Databases, data warehouse, business intelligence, Data Mining methods, medicine, health, health data, healthcare services, electronic patient (medical) records

1. GİRİŞ

Sağlık sistemi politikalarının ve yönetsel kararlarının temeli veri ve veriden elde edilmiş bilgidir. Sağlık politika ve kararlarının amaçlara uygun ve etkin olabilmesi güvenilir, güncel ve doğru veriye bağlıdır. Sağlık bilgi sistemlerinin amacı büyük miktardaki sağlık verilerinden faydalı bilgi üretmektir. Bu bilgiler hasta düzeyinde daha iyi sağlık hizmeti sunumu, sağlık kurumlarının daha iyi yönetilmesi, kaynakların etkin kullanımı ve sağlık politikalarının oluşturulması amaçları ile kullanılmaktadır. Sağlık verileri hastaneler, diğer sağlık kurumları, sigorta şirketleri ve ilgili kamu kurumları başta olmak üzere birçok kuruluş tarafından toplanmaktadır. Günümüzde dijital verilerin hacmindeki artış beraberinde yeni sorun alanları da yaratmıştır. Bunların başlıcaları; çok miktarda, çok boyutlu ve karmaşık verileri işlemek için yöntem ya da sistemler geliştirmek; yeni türdeki verileri işlemek için yöntem ya da sistemler geliştirmek; dağılmış verileri işlemek için yöntem, protokol ya da altyapı geliştirmek; verilerin kullanımı ve güvenliği ile ilgili modeller geliştirmek olarak sıralanabilir.

Büyük miktarda verinin ilk çağrıştırdığı kavram "Veri Madenciliği"dir. Veri Madenciliği, pek çok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları geçerli tahminler yapmak için

kullanan bir süreçtir. Veri - 22 -amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar verme modelleri yaratmaktır. Veri Madenciliği, William Frawley ve Gregory Piatetsky-Shapiro [1] tarafından, "verideki gizli, önceden bilinmeyen ve potansiyel olarak faydalı enformasyonun önemsiz olmayanlarının açığa çıkarılması" biçiminde yapılan bilgi keşfi tanımını destekler. 1990'lı yıllardan itibaren büyük miktarda veri içerisinden, gizli kalmış, değerli, kullanılabilir bilgileri açığa çıkarmak ve stratejik karar destek sağlamak amacıyla kullanılan Veri Madenciliği; bu sorun alanlarına yanıt bulması yanında sağlık verilerinin kullanımında yeni bir perspektif yaratmış ve kullanım alanları hızla artmaya devam eden bir yöntem haline gelmiştir.

Bu makalenin amacı, Veri Madenciliği konusunda bir altyapı oluşturmak ve sağlık profesyonellerine sağlık sektöründe Veri Madenciliği'nin kullanımı ile ilgili örnekler sunarak karar verme süreçleri açısından yeni bir bakış açısı kazandırmaktır. Bu amaçla makalede sırasıyla Veri Tabanlarında Bilgi Keşfi, Veri Ambarı, Veri Madenciliği, İş Zekası ve Veri Madenciliği Yöntemleri konularında tanımlayıcı bilgilere yer verilmekte; ülkemizdeki sağlık sektöründe öncelikli konu ve sorun alanları dikkate alınarak Veri Madenciliği uygulamalarına örnekler verilmektedir.

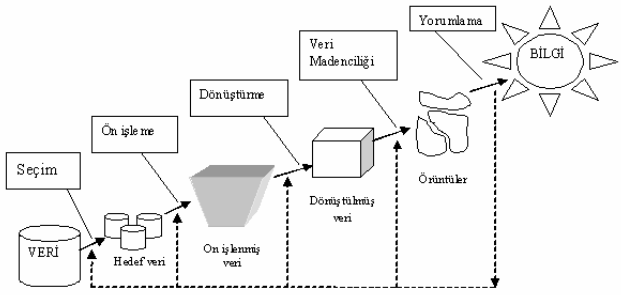
2. VERİ TABANLARINDA BİLGİ KEŞFİ

Veri Madenciliğinden bahsetmeden önce aktarılması gereken temel bir kavram Veri Tabanlarında Bilgi Keşfi (VTBK)'dir. Yaygın olarak KDD (Knowledge Discovery in Databases) kısaltmasıyla bilinen VTBK ve Veri Madenciliği ile aralarındaki ilişki aşağıda başlıklar halinde incelenmektedir.

Tarihsel olarak, veri içerisinde faydalı örüntüleri bulma kavramına Veri Madenciliği, bilgi aktarımı, enformasyon keşfi, enformasyon hasadı, veri arkeolojisi ve veri örüntü işleme gibi pek çok farklı isim verilmiştir. Veri Madenciliği terimi daha çok istatistikçiler, veri analistleri ve yönetim bilgi sistemleri toplulukları tarafından kullanılmaktadır. Aynı zamanda, veri tabanı alanında da popülerliğe ulaşmıştır. 'Veri tabanlarında bilgi keşfi' ifadesi 1989 yılında ilk KDD çalışma toplantısında ortaya atılmış [1] ve bilginin veri keşfi sürecinin nihai ürünü olduğuna vurgu yapılmak istenmiştir. Yapay zeka ve makine öğrenimi alanlarında popülerleşmiştir.

Yukarıdaki perspektiften bakıldığında, Veri Tabanlarında Bilgi Keşfi, veriden faydalı bilginin keşfedilmesi sürecinin tamamına atıfta bulunmakta ve Veri Madenciliği bu sürecin bir adımına karşılık gelmektedir. Veri Madenciliği, veriden örüntülerin aktarımı için özel algoritmaların uygulanmasıdır. Veri tabanlarında bilgi keşfi (VTBK) makine öğrenimi, örüntü tanıma, veri tabanları, istatistik, yapay zeka, uzman sistemler, veri görselleştirme ve yüksek performanslı hesaplama gibi araştırma alanlarının kesişimi olarak gelişmiş ve gelişimine devam etmektedir. Tek hedef, büyük veri setleri kapsamında, düşük düzeyde veriden, yüksek düzeyde bilgi aktarmaktır.

VTBK'nin Veri Madenciliği bileşeni, VTBK'nin Veri Madenciliği sürecinde veri içerisinde örüntüleri bulmada ağırlıklı olarak makine öğrenimi, istatistik ve örüntü tanıma gibi bilinen tekniklere güvenmektedir. Bu konuda doğal bir soru 'VTBK, örüntü tanıma veya makine öğrenimi veya ilgili alanlardan nasıl farklı olmaktadır?' olabilir. Cevap ise, bu alanların, VTBK'nın Veri Madenciliği adımında bazı Veri Madenciliği metotlarını sağlamalarıdır. Oysa VTBK, verinin nasıl depolanıp erişileceğinden, algoritmaların devasa veri setlerine nasıl ölçeklenebileceğine ve hala etkin olarak çalışmalarına, sonuçların nasıl yorumlanabileceği ve görselleştirilebileceğine ve bütün insan-makine interaksiyonunun kullanışlı olarak nasıl modellenip, desteklenebileceğine olmak üzere veriden bilginin keşfinin tüm süreçleri üzerine odaklanır. VTBK süreci, örneğin makine öğrenimi gibi herhangi bir tekniğin ilgi alanı içerisinde yer almanın ötesinde çok disiplinli bir faaliyet olarak görülmelidir. Bu kapsamda, (makine öğreniminin yanı sıra) yapay zekanın diğer alanları için de, VTBK'ye katkı sağlayacak açık fırsatlar vardır [2].



Şekil 1. VTBK sürecinin adımları [2, 3, 4]

2.1 Veri Tabanlarında Bilgi Keşfi Süreci

VTBK süreci, veritabanlarını kullanarak veritabanlarında istenilen seçim, ön işleme, alt örnekleme, dönüşüm, örüntülerin açığa çıkarılması için Veri Madenciliği yöntemlerinin (algoritmalarının) uygulanması ve açığa çıkarılan örüntülerin tanımlanması için Veri Madenciliği ürünlerinin yorumlanmasını ihtiva eder. VTBK sürecinin, VM bileşeni, veriden hangi örüntülerin aktarılıp, dikkate alınacağının algoritmik anlamda ifadesi olarak değerlendirilmelidir. VTBK sürecinin bütünü, Şekil 1'de de görüldüğü gibi, değerlendirme ve madenlenmiş örüntülerin hangilerinin yeni bilgi olarak değerlendirileceğinin olası yorumunu da içerir [4] .

VTBK süreci interaktif ve yinelemeli, kullanıcı tarafından kararların verilmesini gerektiren adımlardan oluşmaktadır. Brachman ve Anand, sürecin interaktif yapısına vurgu yapan pratik bir görünüm vermişlerdir [5]. Sürecin bazı temel adımlarının çerçevesi aşağıda verilmiştir [4]:

1. Adım: Uygulama alanı ve ilgili önsel bilgi ile ilgili bir anlayış geliştirmek ve müşterinin bakış açısından VTBK sürecinin hedefini tanımlamak.
2. Adım: Hedef veri kümesini yaratmak: Keşfin uygulanacağı veri kümesini seçmek veya değişkenlerin bir alt kümesi veya veri örnekleri üzerine odaklanmaktır.
3. Adım: Veri temizleme ve ön işleme: Eğer uygunsa gürültünün kaldırılması, model için gerekli enformasyonun toplanması, kayıp veri alanları için stratejilere karar vermeyi içeren temel operasyonlardır.
4. Adım: Veri indirgeme ve projeksiyon: Görev hedefine bağlı veriyi temsil edecek faydalı özellikleri bulmaktır. Boyut indirgeme veya dönüşüm yöntemleriyle göz önüne alınan değişken sayısı indirgenebilir veya verinin değişmez (invariant) temsili bulunabilir.
5. Adım: VTBK sürecinin hedefleri ile (1. Adımdaki), Veri Madenciliği yönteminin eşleştirilmesi: Özetleme, sınıflandırma, regresyon, kümeleme vb. Yöntemler uygulanmaktadır.
6. Adım: Veri Madenciliği algoritma(larının)sının seçimi: Açıklayıcı analizler, model ve hipotez seçimi: Tercih edilen Veri Madenciliği algoritmaları ve seçilen yöntemler veri örüntülerini araştırmak için kullanılır. Bu süreç, hangi modelin ve parametrelerin uygun olabileceğine ve Veri Madenciliği yönteminin VTBK sürecinin bütün kriterleriyle eşleşip eşleşmediğine karar verilmesini içermektedir.
7. Adım: Veri Madenciliği: Özel bir temsili form veya temsili küme içerisinde ilgilenilen örüntüler; sınıflandırma kuralları ve ağaçları, regresyon ve kümelemeyi içererek araştırılır.
8. Adım: Veri Madenciliği ile çıkarılan örüntülerin yorumlanması: Sonraki iterasyonlarda, Adım 1-7'den herhangi birine dönülmesi ihtimaliyle veri madenciliği ile çıkarılan örüntüler yorumlanır.
9. Adım: Keşfedilen bilgilerin birleştirilmesi: Keşfedilen bilgi sonraki çalışmalar için bir başka sistem altında toplanabilir veya basitçe dokümante edilip, raporlanarak ilgili birimlere iletilir. Bu aynı zamanda, önceden inanılan veya aktarılan bilgilerin doğruluğunu kontrol etme ve olası farklılıkların ayrıştırılmasını da içerir [2,3].

2.2. Veri Tabanlarında Bilgi Keşfi Sürecinin Veri Madenciliği Adımı

Bilgi keşfi hedefleri, sistemin kullanım amacına göre tanımlanır. Hedefleri ikiye ayırabiliriz:

- Doğrulama
- Keşif

Doğrulama ile, sistem kullanıcının hipotezlerini doğrulamak ile sınırlıdır. Keşif ile, sistem bağımsız olarak yeni örüntüler bulur. İleride keşif hedefi, bazı varlıkların gelecekteki davranışlarını tahmin etmek için sistemin örüntüleri bulmasında kullanıldığında tahmin ve kullanıcıya sunumda insanın anlayabileceği bir form için sistem kullanıldığında tanımlama olmak üzere iki alt gruba ayrılacaktır.

Veri Madenciliği gözlenen veriye model uydurmayı veya gözlenen verideki örüntüleri tanımlamayı gerektirmektedir. Model uydurma, bilgi çıkarımı rolünü üstlenmektedir: Modelin, kullanışlı veya ilginç keşifsel bilgiye işaret edip etmediği, tamamıyla interaktif VTBK sürecinin subjektif insan yargısına tipik olarak ihtiyaç duyduğu bir parçasıdır. Model uydurmada istatistiksel ve mantıksal olmak üzere iki temel matematiksel yapı kullanılmaktadır. Modelde, istatistiksel yaklaşım deterministik olmayan etkiye, mantıksal yaklaşım ise deterministik etkiye izin vermektedir.

Pek çok Veri Madenciliği yöntemi, makine öğrenimi, örüntü tanıma ve istatistikten denenmiş ve test edilmiş teknikleri temel almıştır: Sınıflandırma, kümeleme, regresyon vb. [2].

3. VERİ AMBARI

Veri Ambarları, Veri Madenciliği ile eşanlı olarak anılan ve Veri Madenciliği sürecinin gerçekleştirildiği veriyi sağlayan özel bir veri tabanıdır. Tanım olarak Veri Ambarı, pek çok farklı kaynaktan ve genellikle de farklı yapıda verinin depolandığı ve hepsinin de aynı birleşik çatı altında kullanılmasının ümit edildiği yapılardır. Ayrıca, Veri Ambarı pek çok farklı kaynaktan elde edilen veriyi aynı çatı altında analiz etme imkânı tanır [2].

Veritabanlarından geliştirilen ilgili bir alan da, işlem verilerini toplamak ve online analiz ve karar destek amaçlı kullanıma uygun hale getirmek için temizlemenin adı popüler iş trendi olarak atıfta bulunulan veri ambarlarıdır. Veri ambarcılığı, veri kümelerine VTBK aşaması için iki önemli yoldan yardımcı olur:

- Veri temizleme
- Veri erişimi [6].

3.1. Veri Temizleme

Organizasyonlar, sahip oldukları geniş kapsamlı veri ve veri tabanlarının, birleşik mantıksal görünümde olduğunu düşünmeye zorlandıklarından, haritalanmış verinin sonuçlarını bir tek isimlendirme eğiliminde olmalarının yanısıra, kayıp veriyi düzgün temsil etmek, ele almak ve gürültü ile hataları adres göstermek zorundadırlar [6].

3.2. Veri Erişimi

Genelde veriye erişim, özelde tarihi açıdan elde edilmesi zor olan veriye erişim yolları sağlayacak düzgün ve iyi tanımlanmış yöntemler yaratılmalıdır [6]. Öncelikle organizasyon ve bireyler verilerini depolama ve verilerine erişim problemlerini çözmelidirler. Doğal olarak bir

sonraki adım ‘ Bütün veri ile ne yapacağız ?’ sorusudur. Bu soru, doğal olarak VTBK fırsatını ortaya çıkarmaktadır.
Veri ambarlarını analiz etmenin popüler yaklaşımlarından birisi Online Analytical Processing (OLAP)’tır. OLAP araçları, birçok boyutta hesaplama özetleri ve tanımlamalarında SQL’den üstün olan, çok boyutlu veri analizi sağlamaya odaklanmışlardır. OLAP araçları interaktif veri analizi sağlama ve basitleştirmeyi hedeflemişlerdir. Ama VTBK araçlarının hedefi, süreci mümkün olduğunca otomatikleştirmedir [6].

4. VERİ MADENCİLİĞİ

Veritabanlarında bilgi keşfi, sıklıkla, büyük hacimde veri koleksiyonundan faydalı bilgiyi keşfetmeyi hedefleyen, Veri Madenciliği olarak anılmaktadır. Veritabanları günümüzde terabaytlarla ifade edilmektedir. Bu büyük hacimde verinin içinde stratejik önem taşıyan gizli enformasyon yatmaktadır. Ama bu kadar büyük hacimli veri içerisinde yer alan önemli bilginin nasıl açığa çıkarılacağı en önemli sorudur.

Bu önemli soruya en güncel yanıt, hem geliri artırırken hem de maliyetleri indirgeyen Veri Madenciliğidir.

Veri Madenciliği, pek çok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları geçerli tahminler yapmak için kullanan bir süreçtir [7]. Veri Madenciliği, en basit tanımıyla, veri tabanlarındaki ilişkili örüntüleri otomatik olarak belirlemedir. Veri Madenciliği sihir değildir. Yıllardır, istatistikçiler veri tabanlarını elle kazımakta, istatistiksel açıdan önemli ilişkiler aramaktadır. Veri Madenciliği, bu süreci otomatik olarak gerçekleştirmektedir.

Veri Madenciliği veri kümesi içerisinde keşfedilmemiş örüntüleri bulmayı hedefleyen teknikler koleksiyonunu betimlemektedir. Veri Madenciliğinin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar-verme modelleri yaratmaktır. Veri Madenciliği, William Frawley ve Gregory Piatetsky-Shapiro [1] tarafından, ‘ ... verideki gizli, önceden bilinmeyen ve potansiyel olarak faydalı enformasyonun önemsiz olmayanlarının açığa çıkarılması...’ biçiminde yapılan bilgi keşfi tanımını destekler [8].

4.1. İstatistiksel Perspektiften Veri Madenciliği veya İstatistiksel Öğrenme

İstatistik alanı bilim ve sanayideki problemlere kapı açmak için sürekli olarak onlara meydan okumaktadır. İlk zamanlarda bu problemler sık sık ziraat ve sanayi deneylerinden gelmekte ve göreli olarak dar kapsamlı olmaktaydılar. Bilgisayarların ve bilgi çağının gelişiyle istatistikî problemler hem boyut hem de karmaşıklık açısından patlamıştır. Veri depolama üniteleri, organizasyon ve araştırmadaki gelişmeler yeni bir alan olan Veri Madenciliğine işaret etmiştir; biyolojideki istatistiksel ve hesaplama problemleri ve ilaç biyoenformatiği yaratmıştır. Pek çok alanda hala çok büyük miktarda veri üretilmekte ve istatistikçilerin işi bunların tamamı hakkında akıl yürütmektir: önemli örüntü ve eğilimleri açığa çıkartmak ve ‘verinin ne söylediğini’ anlamaktır ki buna veriden öğrenme denilmektedir.

Veriden öğrenmedeki gelişmeler, istatistiksel bilimlerde bir devrime işaret etmiştir. Hesaplamanın böyle bir anahtar rol oynamasından beri, araştırmacıların bilgisayar bilimleri ve mühendislik gibi diğer alanlarda da bu yeni gelişmeleri gerçekleştirmesi çok fazla sürpriz olmamıştır.

Öğrenme problemleri kabaca denetimli ve denetimsiz olarak ikiye ayrılabilir. Denetimli öğrenmede hedef, girdi ölçülerinin sayısını temel alarak çıktı ölçüsünün değerini tahmin etmektir; denetimsiz öğrenmede ise çıktı ölçüsü yoktur ve hedef girdi ölçüleri kümesi arasındaki birliktelik ve örüntüleri betimlemektir. İstatistiki öğrenme bilim, sağlık, finans ve sanayinin pek çok alanında anahtar rol oynamaktadır. Öğrenme problemlerine bazı örnekler;

- Bir erkek hastanın, sağlık durumu, yaşam alışkanlıkları ve genetik faktörlere dayalı olarak prostat kanserine yakalanma riskinin tahmini. (Tahmin, bu hasta için demografik, diyet ve klinik ölçüleri temel alacaktır.)
- Şirket finansal performans ölçülerini ve ekonomik verileri temel alarak şirketin finansal kriz olasılığını tahmin etmek,
- Manyetik görüntüden el yazısı ile yazılmış rakam ve harfleri tanımlamak,
- olarak verilebilir.

Öğrenme bilimi istatistik, veri madenciliği ve yapay zeka, mühendislik alanı ve diğer disiplinlerin kesişiminde anahtar rol oynamaktadır. Veriden öğrenme dikkate alındığında, tipik senaryo; genellikle nicel (tahlil sonuçları) veya kategorik (prostat kanseri veya değil) çıktı ölçümü vardır ve içerik kümesi (demografik, diyet ve klinik ölçümler gibi) temel alınarak tahmin edilmek istenir. Nesne kümeleri için çıktı ve içerik ölçümlerinin gözlemlerinden verinin eğitim kümesi vardır. Bu verinin kullanımıyla tahmin veya öğrenme modeli kurulur, öyle ki yeni gözlenmemiş çıktıları tahmin etme imkanı tanır. İyi bir öğrenici böyle bir çıktıyı başarıyla tahmin eder.

Yukarıda yapılan tanımlama denetimli öğrenme problemini tasvir etmektedir. Öğrenme sürecine kılavuzluk edecek bir çıktı değişkeninin varlığı nedeniyle denetimli denilmektedir. Denetimsiz öğrenme probleminde ise sadece içerik gözlenir ve çıktı ölçümü yoktur. Görev sadece verinin nasıl organize edildiğini veya kümelendiğini betimlemektir [9].

4.2. Veri Madenciliği İle Diğer Analitik Yöntemlerin Karşılaştırılması

Veri Madenciliği ile yeni tanışanların özellikle de veri tabanı pazarlaması, geleneksel veri analizi ve istatistik

alanında çalışmış olanların Veri Madenciliği ile diğer analitik yöntemler arasındaki farkın ne olduğunu sıkça sormaları muhtemeldir. Veri Madenciliği sıkça aşağıdakiler gibi düşünülmektedir [4]:

- Büyük bir veri ambarı üzerinde SQL (Structured Query Language) sorgusu,
- Herhangi bir sayıda veritabanı veya veri ambarları üzerinde SQL sorgusu,
- İleri düzeyde enformasyon erişimi, örneğin akıllı ajanlar yoluyla,
- Çok boyutlu veritabanı analizi (Multidimensional Database Analysis-MDA),
- OLAP,
- Açıklayıcı veri analizi,
- İleri grafiksel görselleştirme,
- Veri ambarı üzerinde geleneksel istatistiksel işleme.

Bu yaklaşımların hiçbiri Veri Madenciliği değildir. Çünkü her birindeki temel eksiklik, bilginin keşfinin önceden önerilmiş bir hipotez olmaksızın gerçekleştirilmesidir [6].

4.2.1 İstatistiksel Analiz ile Veri Madenciliğinin Karşılaştırılması

İstatistiksel analiz ve Veri Madenciliğinin karşılaştırması ve farklılaştığı noktalar aşağıda yer almaktadır [10].

Tablo 1. İstatistiksel analiz ve Veri Madenciliğinin karşılaştırılması

İstatistiksel Analiz	**Veri Madenciliği**
İstatistikçiler genellikle bir hipotez ile başlarlar	Veri Madenciliği hipoteze gerek duymaz
Hipotezlerini eşleştirmek için kendi eşitliklerini geliştirmek zorundadırlar	Veri Madenciliği algoritmaları eşitlikleri otomatik olarak geliştirir.
İstatistiksel analizler sadece sayısal verileri kullanır.	Veri Madenciliği farklı tiplerde data kullanır (örneğin metin, ses) sadece sayısal veriyi değil.
Kirli veriyi analizleri sırasında bulur ve filtre eder.	Veri Madenciliği temiz veriye dayanır.
İstatistikçiler kendi sonuçlarını yorumlar ve bu sonuçları yöneticilere iletirler.	Veri Madenciliğinin sonuçlarını yorumlamak kolay değildir. Sonuçlarını analiz etmede ve yorumlamada ve bulguları yöneticilere iletmede mutlaka istatistikçiye ihtiyaç duyulmaktadır.

4.2.2 Veri Madenciliği, OLAP ve Veri Sorgusunun Kıyaslanması

- Hemen hemen ne arandığı biliniyor ve büyük veri tabanı ile çalışmak isteniyorsa veri sorgusu kullanılmalıdır.
- Büyük veri tabanlarında basit ilişkiler keşfedilmek isteniyorsa OLAP kullanılmalıdır.
- Veri içerisinde açıkça gözlenemeyen örüntü ve ilişkiler bulunmak isteniyorsa Veri Madenciliği kullanılmalıdır. Veri Madenciliği algoritmalarının göreli yavaşlığı nedeniyle, genellikle veri tabanının küçük veya örneklem olması gerekmektedir. Veri Madenciliği algoritmalarını büyük veri tabanlarında çalışacak biçimde ölçekleyebilmek, Veri Madenciliğinin güncel araştırma konularından birisidir.

SQL, OLAP ve Veri Madenciliği kullanımını, keşfedilmek istenen bilgi tipine göre sınıflarsak:

- Sığ Bilgi: Seçilen kayıtlara ait ortalama ve toplam değer gibi özet bilgiler için kayıt seçmek yeterlidir ki SQL bunu yapabilir.
- Çok boyutlu bilgi: Farklı özelliklerin, ortaya çıkma sıklığı hakkında bilgi. Veri küpü üzerinde OLAP bunu yapabilir.
- Gizli bilgi: Önceden tahmin edilemeyen örüntü ve ilişkiler Veri Madenciliği için başlangıç olabilir.
- Derin bilgi: Sadece önsel teknik veya meta-bilginin kullanımıyla keşfedilebilecek gizli örüntüler ve ilişkiler hakkında bilgi. Bu konu Veri Madenciliğinin araştırma sınırları içerisindedir [11].

5. VERİ MADENCİLİĞİ VE İŞ ZEKÂSI

İş zekâsı terimi, işte karar vermeyi destekleyen ve bilgi teknolojilerini temel alan bütün süreçler, teknikler ve araçlar için genel anlamda kullanılan bir ifadedir. Veri Madenciliği iş zekâsının yeni ve önemli bir bileşenidir. Şekil 2. farklı iş zekâsı çözümlerinin taktik ve stratejik iş kararları temelindeki potansiyel değerlerine göre mantıksal pozisyonlarını göstermektedir.

Şekil 2 Veri Madenciliği ve iş zekası

Genel olarak, piramitte aşağıdan yukarıya çıkıldıkça karar vermeyi destekleyen enformasyonun değeri artmaktadır [6].

6. VERİ MADENCİLİĞİ YÖNTEMLERİ

Veri Madenciliği yöntemlerini denetimli ve denetimsiz olmak üzere iki ana kategoriye ayırmak mümkündür. Veri Madenciliğinde iyi tanımlanmış veya kesin bir hedef olduğunda denetimli (supervised) ifadesi kullanılır. Elde edilmesi istenen sonuç için özel bir tanımlama yapılmamışsa veya belirsizlik söz konusu ise denetimsiz (unsupervised) ifadesi kullanılır [9].

Denetimli ve denetimsiz ifadeleri birbirinin tersine karşılık gelmektedir. Denetimli ve denetimsiz yöntemleri sürecin bütünü açısından değerlendirmek gerekirse;

- Denetimsiz yöntemler daha çok veriyi anlamaya, tanımaya, keşfetmeye yönelik olarak kullanılan ve sonraki uygulanacak yöntemler için fikir vermeyi amaçlamaktadır,
- Denetimli yöntemler ise veriden bilgi ve sonuç çıkarmaya yönelik kullanılmaktadır,

denilebilir. Bu nedenle denetimsiz bir yöntemle elde edilen bir bilgi veya sonucu, eğer mümkünse denetimli bir yöntemle teyit etmek, elde edilen bulguların doğruluğu ve geçerliliği açısından önem taşımaktadır.

Denetimli (Supervised) Veri Madenciliği yöntemleri:

- En yakın k komşuluk (k-Nearest-Neighbor)
- K-ortalamalar kümeleme (K-means clustering)
- Regresyon modelleri (Regression models)
- Kural çıkarımı (Rule induction)
- Karar ağaçları (Decision trees)
- Sinir ağları (Neural networks)

Denetimsiz (Unsupervised) Veri Madenciliği yöntemleri:

- Aşamalı kümeleme (Hierarchical clustering)
- Kendi kendini düzenleyen haritalar (Self organized maps)

olarak sınıflandırılabilir [9, 12].

Veri Madenciliği ile ilgili kullanılan pek çok yöntemin yanına hemen her geçen gün yeni yöntem ve algoritmalar eklenmektedir. Bunlardan bir kısmı onlarca yıldır kullanılan klasik teknikler diyebileceğimiz ağırlıklı olarak istatistiksel yöntemlerdir. Diğer yöntemler de genellikle istatistiği temel alan ama daha çok makine öğrenimi ve yapay zekâ destekli yeni nesil yöntemlerdir.

Veri Madenciliğinde kullanılan klasik yöntemlerin başlıcaları;

- Regresyon,
- K - En Yakın Komşuluk,
- Kümeleme

olarak sayılabilir.

Yeni nesil yöntemlerin başlıcaları ise;

- Karar Ağaçları,
- Birliktelik Kuralları,
- Sinir Ağları,

olarak sıralanabilir [8].

Ayrıca diğer Veri Madenciliği yöntemlerinin başlıcaları da;

- Temel Bileşenler Analizi,
- Diskriminant Analizi,
- Faktör Analizi,
- Kohonen Ağları,
- Bulanık Mantığa Dayalı Yöntemler,
- Genetik Algoritmalar,
- Bayesci Ağlar,
- Pürüzlü (Rough) Küme Teorisine Dayalı yöntemler,

olarak sıralanabilir [13].

Yukarıda sayılan yöntemlerin dışında birden fazla tekniği içine alan hibrid yöntemler ve zaman serilerine dayalı yöntemlerden de Veri Madenciliği yöntemi olarak faydalanılmaktadır [14].

Özet olarak, bilgi keşfine yarayan her yöntem Veri Madenciliği yöntemi olarak kullanılabilmektedir. Aşağıda yaygın kullanıma sahip başlıca veri madenciliği yöntemleri ve kısa tanımları verilmektedir [1], [2], [4], [6], [8], [9], [12], [13], [14] :

6.1. Regresyon Analizi

Yaygın kullanılan bir modelleme tekniğidir. Doğrusal, doğrusal olmayan ve lojistik modelleme alternatifleri imkanı vardır. Bağımsız değişken olarak anılan tahmin

edici değişkenlerin; bağımlı değişken denilen tahmin edilecek değişken değerini belirleyecek ağırlıklandırmaları içerecek bir bağımsız değişkenler birleşimidir.

6.2. K-En Yakın Komşuluk

Özellikle büyük veri tabanlarında kullanılan bir sınıflandırma tekniğidir. Sınıflandırılmak istenen nesnenin ait olduğu kümeyi, en yakınında yer alan K birim nesneden en fazla birime ait olanla aynı kümede sınıflandırması mantığına dayanmaktadır.

6.3 K-Ortalamalar Kümeleme Analizi

Segmentasyon, gruplama ve sınıflandırma yöntemidir. N birim nesnenin, K gruba ayrılması mantığına dayanır. Sınıf aralıkları belli olmadığında; bir benzerlik veya benzemezlik ölçütüne (metriğine) dayalı olarak, grup içinde homojen, gruplar arasında heterojen K adet küme yaratır.

6.4. Aşamalı (Hierarchical) Kümeleme Yöntemleri

K-ortalamalar kümeleme analizi gibi sınıf aralıklarının belli olmadığı durumlarda, kullanılan bir segmentasyon yöntemidir. K-ortalamalar Kümeleme Analizi'yle arasındaki en büyük fark, hiyerarşik kümeleme analizinde küme sayısının, uygulayıcı müdahalesi olmadan, gözlem değerlerinin farklılıklarına göre kendiliğinden oluşmasıdır. Dolayısıyla, küme sayısı analize başlarken belirlenememekte; analiz sonucunda belirlenebilmektedir. Analizin çıktısı olarak elde edilebilen Dendogram denilen şekiller de, analizi görsel olarak desteklemekte ve daha anlaşılır kılmaktadır.

6.5. Karar Ağaçları

Veri madenciliği denildiğinde, sinir ağları ile birlikte ilk akla gelen yöntemlerden olan karar ağaçları, yeni jenerasyon veri madenciliği yöntemlerindendir. Bir ağaç diyagramı biçiminde, her bir dal ve yaprağı bir sınıflandırma sorgusu olacak biçimde dallanan yöntem; nitel, nicel, sürekli, kesikli tüm değişkenlere uygulabilen algoritmaları, ağaç diyagramı şeklindeki görsel desteği, SQL sorgusuna kolay dönüştürülebilir yapısıyla en popüler segmentasyon yöntemlerinden birisidir. C 4.5, C5.0, C&RT ve CHAID en popüler yöntemlerdir.

6.6 Sinir Ağları

İnsan beyninin hesaplama mantığı baz alınarak oluşturulmuş (yapay) sinir ağları, karar ağaçları gibi yeni jenerasyon veri madenciliği yöntemlerindendir. Girdi ve çıktı arasında, küçük hesaplama birimlerinden elde edilen sonuçları birleştirerek sonuçlandıran bir modelleme yöntemidir. Karar ağaçları uygulama, anlama ve yorumlama açısından ne kadar kolaysa, sinir ağları da o derece zordur. Yalnızca model oluşturma, sonuçları yorumlama aşamasının ötesinde; doğru bir model kurabilmek için ağın eğitimindeki dengenin önemi oldukça büyüktür. Fazla eğitilmiş bir ağ, önceden gözlenmemiş bir gözleme yönelik tahmin kabiliyetini yitirirken; az eğitilmiş bir ağ ise yanlış tahmin verebilmektedir.

6.7. Birliktelik (İlişki) Kuralları

Gözlem değerleri arasındaki ilişkiyi, koşullu olasılık bazlı değerlendirmelerle özet olarak sunan ve uygulayıcı tarafından baştan tanımlanmış bir başarı oranının üzerindeki kuralları sıralayan bir yaklaşım izlenmektedir. Hesaplama mantığı nedeniyle hızlı sonuç vermesi ve çok büyük veri setlerine kolaylıkla uygulanabilmesi Birliktelik Kuralı Analizi'ni ticari veri tabanlarının madenciliğinde gittikçe popülerleşen bir araçtır haline getirmiştir.

6.8. Önemli Bileşenler Analizi

Çok fazla değişkenin etüd edilmesi gereken bir durumda, tüm değişkenleri içerecek bir modelin başarılı tahmin yapma kabiliyetinde oluşabilecek zafiyetin ötesinde, tüm bu değişkenleri gözlemlemek, veri toplamak ve değerlendirmek; zaman, insangücü ve maliyet açısından önemli bir yük getirmektedir. Bu noktada tolere edilebilir düzeyde açıklayıcılıktan fedakarlık, daha az sayıda değişkenle bir model kurmayı sağlayabilir. Önemli Bileşenler Analizi, çok sayıda değişken içerisinde, açıklayıcılığa önemli düzeyde katkı sağlayan daha az sayıda değişkenin kullanımına imkan tanıyan Doğrusal Regresyon Yöntemi'nin özel bir durumudur. Yöntem oldukça kullanışlı olmasına karşın, çok kesin Normal Dağılım varsayımlarının göz ardı edilmesi, yanlış sonuçlar elde etme riskini oluşturmaktadır.

6.9. Diskriminant Analizi

Bir sınıflandırma probleminde, sınıflamanın gerçekleştirilmesi ve oluşturulmuş bir sınıflamada gözlem atamalarının doğru yapılması Diskriminant Analizi'yle sağlanmaktadır.

6.10. Kendi Kendini Düzenleyen Haritalar

Bu yöntem K-ortalamalar Kümeleme Analizinin kısıtlı versiyonu olarak görülebilir; gözlemler iki boyutlu bir düzlemde sınıflandırılır. Kendi kendini düzenleyen haritalar, orijinal yüksek-boyutlu gözlemlerin iki boyutlu koordinat sistemine indirgenerek haritalandığı kısıtlanmış bir topolojik haritaya işaret etmektedir. Orijinal SOM algoritması çevrimiçidir (online) – gözlemler anında işlenir – ve toplu işlem (batch) versiyonu daha sonra önerilmiştir.

7. TIP ve SAĞLIK HİZMETLERİNDE VERİ MADENCİLİĞİ KULLANIMI ve UYGULAMALARI

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 profesyonellerinin en doğru ve güncel bilgiye ulaşması ve bu bilgiyi karar

destek sistemlerinden faydalanarak kullanması gerekmektedir.

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

Ülkemizde, Sağlık Bakanlığı yaptığı değerlendirmeler ile 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 tespitinde bulunmuş ve "Sağlıkta Dönüşüm Programı" kapsamında "Karar Sürecinde Etkili Bilgiye Erişim: Sağlık Bilgi Sistemi" başlığı ile çalışmalar başlatmıştır. Ulusal Sağlık Veri Sözlüğü, Minimum Veri Setleri, Sağlık Kodlama Referans Sunucusu ve sağlık verilerinin toplandığı Elektronik Sağlık Kaydı (ESK) veritabanı ve Karar Destek Sistemi bileşenleri bu çalışmaların kapsamını oluşturmaktadır [15]. Sağlık Bakanlığı'nın toplanan verilerin analiz amaçlı göstergelere dönüştürüldüğü, karar vermede yol gösterici modeller yaratacak Veri Madenciliği çözümlerine başvurması kaçınılmazdır. Ancak, burada doğru Veri Madenciliği çözümüne başvurmak doğru sonuçlara ulaşmak açısından çok önemlidir.

Makalenin bu bölümünde ülkemizdeki gerek kamu gerekse özel sağlık sektöründeki tüm sağlık profesyonellerine karar destek amaçlı perspektif sağlayacak Veri Madenciliği çözümlerine örnekler sunulmaktadır. Bu örmekler belirlenirken günümüzde özellikle sağlık yöneticilerinin ve profesyonellerinin öncelikli konuları dikkate alınmıştır.

7.1 Veri Ambarı Oluşturma

Sağlık işletmelerindeki tüm verilerin farklı amaçlarla kullanımında veriye erişim ve analiz edilebilir temiz veri sağlama en önemli sorun alanlarından birisini oluşturmaktadır. Ayrıca, büyük veri setlerinin içinde klinik ve demografik bilgiler gibi öncelikli verilerin derlenmesi zaman ve kaynak maliyeti yaratmaktadır [16,17].

Veri Madenciliği Çözümü: Veri Madenciliği yöntemlerinin uygulanmasında temel zorunluluklardan birisi veriye, kolay erişim ve analiz edilebilir temiz veri teminidir ki; bu sorun veri ambarlarıyla çözümlenmektedir.

Hastanelerdeki tüm veri, temizlenip konsolide edilerek bir tek üretici veri tabanına veya veri ambarına indirgenebilir.

Üstünlükler: Hastanelerin ürettiği büyük hacimde veri çoğunlukla bir hiyerarşi içerisinde düzenlenmeden rasgele tutulduğu için analiz edilebilirlikten uzak "Veri Çöplükleri" oluşmaktadır.

Veri Madenciliği altlığı hazırlayacak veri ambarlarıyla, Veri Madenciliği sürecinde ihtiyaç duyulan temiz, analiz edilebilir veriye erişim imkanı elde edilmektedir.

7.2 Elektronik Hasta Dosyalarının Oluşturulması

Hastanın hikâyesine yönelik tüm kayıtların; teşhis tedavi süreçlerinin; laboratuvar sonuçlarının; röntgen, MR gibi görüntü dosyalarının bir tek kayıt içerisinde zamana endeksli olarak hazırlanması verilerin değerlendirilebilmesinde ve hizmet sunumunda büyük önem taşımaktadır [18].

Veri Madenciliği Çözümü: Veri ambarı mantığına uygun olarak, kullanılabilir ve kaliteli verinin pek çok veri tabanından bir tek veri tabanına konsolide edilmesi veya tek merkezden erişimin sağlanabileceği bütünleşik bir yapı oluşturulması gerekmektedir.

Üstünlükler: Hastanın teşhis tedavi sürecinde, hekime karar-destek sağlayacak temiz veriye erişimin sağlanması ve kullanılacak Veri Madenciliği yöntemlerine uygun altyapının hazırlanmasıdır.

7.3 Veri Sorunlarının Çözümü

Sağlık verilerinin kullanımında önemli aşamalardan biri olan veri sorunun çözümü ve bunun için bir yaklaşım önermektir. Veri sorunları denildiğinde, temel anlamda:

- Kayıp veri,
- Tutarsız veri,
- Aykırı değer,
- Uç değer,

biçiminde özetlenebilir. Bu sorunların, aynı zamanda istatistiksel veri analizinin de sorunları olduğu açıktır. Dolayısıyla,

- Sorunlu kaydı (gözlemi) silme,
- Sabit değer (ortalama, mod, medyan vb.) atama,
- İmputasyon tekniklerinin kullanımı,

gibi çözüm yaklaşımları izlenebilmektedir. Yukarıdaki yaklaşımlar incelendiğinde en rasyonel çözümün imputasyon tekniklerinin kullanımı olduğu açıktır. Ancak, imputasyon yaklaşımı da çoğunlukla parametrik bir şablon modeli temel almakta ve statik bir çözüm sunmaktadır.

Veri Madenciliği Çözümü: Veri sorunun çözümü için parametrik olmayan, dinamik bir yaklaşım izleyip verinin kendi içerisindeki anomalilerin tespitinden hareket ederek veri sorunlarını gidermeyi amaçlayan Veri Madenciliğine dayalı profillendirme yaklaşımı kullanılmalıdır [19].

Üstünlükler: Veri Madenciliğine dayalı profillendirme yaklaşımı ile sağlık sektöründe uygulamalar gerçekleştirecek araştırmacılar, bir şablon model kullanmak yerine her farklı örüntü içerisinde yer alan veri sorunun, örüntünün özellikleri dikkate alınarak giderilmesi imkanına kavuşacaktır. Böylelikle, araştırmacılar daha az yanılgıya izin verecek bir çözüme ulaşmış olacaklardır.

7.4 Kronik Hastalıklar İçin Erken Uyarı Sinyallerinin Veri Madenciliği İle Tespiti

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 proaktif çözümler geliştirilmesi gerekmektedir [20, 21].

Veri Madenciliği Çözümü: Her bir kronik hastalığa yönelik sosyal, ekonomik, demografik, coğrafi vb. tüm değişkenler dikkate alınarak, hastalığın ortaya çıkışında etkisi olan değişkenlerin Önemli Bileşenler Analizi, Faktör Analizi veya Lojistik Regresyon ile belirlenmesi mümkündür. Akabinde, etkisi tespit edilen değişkenlerin, etkin olduğu sınır değerler dikkate alınarak; hastalığın ortaya çıkışına işaret edebilecek risk sinyalleri geliştirilebilir.

Üstünlükler: Yaygın olarak, genel kabule sahip varsayılan hipotetik kabul değerlerinin sınaması yerine, farklı gruplara yönelik norm değerler belirlemek mümkün olabilir. Böylece, farklı gruplara göre, farklı politikalar çözüm önerileri geliştirilebilir.

7.5 Laboratuar Testleri için Hata ve Suistimal Tespiti

Sağlık hizmetlerinin sunumunda ortaya çıkan hata ve suistimal arasındaki farkın ortaya konulması, risklerin minimize edilmesi ve gerekli önlemlerin bu ayrıma göre alınması hasta güvenliği açısından oldukça önemli bir konudur [22, 23].

Veri Madenciliği Çözümü: Büyük hacimli bir veri hesaba katılarak, ardışık bir süreç tasarlanması gerekmektedir. Öncelikle K-ortalamalar Kümeleme Analizi ile 'Normal Değerlerden' ayrışanlar tespit edilerek; ardından, anomali gösteren değerlerin özel bir sağlık durumu mu belirttiği, yoksa bir suistimale mi işaret ettiğinin belirlenmesi için Karar Ağaçları ve Birliktelik Kuralları yöntemlerinin kullanımıyla elde edilen bulgular, Kümeleme Analizi ile elde edilen 'Normal Değerler' ile kıyaslanarak; anomalinin özel bir sağlık durumuna veya suistimale işaret ettiği belirlenebilir. Kararsız kalınan durumlarda ise uzman görüşünden faydalanılarak, analiz süreci güncellenmiş veriyle yinelenir.

Üstünlükler: Başta sosyal güvenlik kurumları ve sigorta şirketlerinin ödemeleri olmak üzere tüm sağlık ödemelerinde suistimalle karşılaşılabilmektedir. Zaman zaman hastanın bilgisi dışında dahi, erkeklere gebelik testi, kadınlara prostat kontrolü gibi absürd örneklerle bile maalesef sistemin boşluklarından istifade edilerek gerçekleştirilen suistimaller olarak karşılaşılmaktadır. Karar ağaçları ve birliktelik kuralları yöntemleriyle, yalnızca mantıksal tutarlılık denetimi değil, aynı zamanda suistimal örüntüsünün detayları da elde edilerek, duruma özel çözüm önerileri geliştirmek mümkün olabilecektir.

7.6 Klinik Karar Destek Sistemlerinin Geliştirilmesi

Hastanın problemlerinin teşhis ve tedavisinde hekime yardımcı olacak bir veri ambarı yaklaşımlı veri bankası oluşturulmasına ve teşhis veya tedavi sırasında hekime veri, çözüm, risk ve önerileri otomasyona bağlı olarak sunabilecek sistemlere ihtiyaç duyulmaktadır [24, 25].

Veri Madenciliği Çözümü: Elektronik hasta kayıtları üzerinde çalışabilecek, hekimin isteğine uygun yöntemi teorik karmaşaya girmeden isteğe cevap verecek biçimde sunabilen sistemler tasarlanmalıdır. Bir başka deyişle, amaca uygun Veri Madenciliği yöntemlerinin kullanıcı grafik arayüzünün arkasına gizlendiği karar destek sistemleri geliştirilmesidir.

Üstünlükler: Akıllı sistem, birçok veriye anında ulaşabilen, değişkenleri anında ilişkilendirip analiz edebilen, kolay kullanım ve erken uyarı özelliklerine sahip olacaktır.

7.7 Hasta Odaklı Sağlık Hizmeti Sunumu ile Kalitenin Geliştirilmesi

Sağlık hizmetlerinde kalite birçok faktörden etkilenen ve hastanın algılama düzeyine göre değişkenlik gösteren bir kavramdır. Bu nedenle kalite göstergelerini tespit ederken, hastalık grupları, hastanın demografik özellikleri, hastanın sigorta durumu, klinik ve hizmet kalitesi gibi değişkenlerin bir arada düşünülmesi ve değerlendirilmesi gerekmektedir [26, 27, 28].

Veri Madenciliği Çözümü: Hasta ve yönetim görüşlerinin elde edileceği bir soru kağıdıyla elde edilecek değişkenler, Güvenilirlik ve Soru Analiziyle öncül indirgemeden sonra, anahtar değişken(ler) vasıtasıyla idari (kayıt) verisiyle birleştirilerek; Önemli Bileşenler Analizi, Faktör Analizi, Lojistik regresyon ve Karar Ağaçları algoritmaları uygulanabilir.

Üstünlükler: Kaliteyi etkileyen tüm değişkenler birlikte (çok boyutlu olarak) ele alınabilir, hasta, hastalık ya da hedef kümelere göre otomatik olarak kümelenerek her odak grup için kalite değişkenleri ayrı ayrı belirlenebilir.

7.8 Hizmet Sunumunu Optimize Etmek İçin Risk Analizleri

Ulusal, bölgesel ve hizmet verilen kurum bazında optimum hizmet bileşenini oluşturmak ve kaynak tahsisi için etkin planlama yapabilmek için hizmet sunumunun optimizasyonu gerekmektedir [29, 30, 31, 32, 33].

Veri Madenciliği Çözümü: Riskin teorik tanımı kayıp fonksiyonunun beklenen değeridir. Dolayısıyla, risk tanımlama da;

- Model tanımlama (ÖBA, Faktör Analizi, Regresyon Modeller, Sinir Ağları vb.)
- Risk göstergelerinin belirlenmesi esas alınabilir. (Karar ağaçları, Birliktelik Kuralları, K-ortalamalar Kümeleme Analizi'ni takiben Lojistik regresyon veya Sinir Ağları Modellemesi).

Üstünlükler: Hizmet sunumu için bileşenleri oluşturan tüm değişkenler birlikte (çok boyutlu olarak) ele alınabilir, risk faktörleri belirlenebilir.

7.9 Suistimallerin ve Fatura Yolsuzluklarının Tespiti

Her alanda olduğu gibi sağlıkta da suistimal oldukça yaygın ve çoğu kamu finansmanına dayalı hizmetlerde olduğu için ülke ekonomisine getirdiği yük oldukça fazladır. Ancak, ülkemizde şikayete dayalı bir sistem olduğundan, fatura yolsuzluğu, yeşil kart suistimali gibi suistimallerin tespiti zordur. Sosyal Güvenlik Kurumu'nun bu amaçla "örneklem" çekerek yaptığı denetimler de etkin çözüm olarak görülmemektedir. Bilişim teknolojilerindeki gelişmeler, klasik insan temelli teftiş yöntemlerinin yerini, gün geçtikçe daha fazla, otomasyona dayalı gözetim ve denetim sistemlerinin almasına yol açmaktadır. Önceleri insan eliyle yapılan hesaplama ve sorgulamalar yerini, bilişim teknolojileri destekli sistemlere bırakmaya başlamış ve neredeyse düşünüp karar veren akıllı algoritmalarla artık potansiyel riski algılamaya ve önlemeye yönelik sistemler gündeme gelmiştir [34, 35].

Veri Madenciliği Çözümü: Suistimaller aykırı veya uç değer olarak değerlendirilebilir. Ardışık olarak Hiyerarşik veya K-ortalamalar Kümeleme Analizi veya iteratif bir süreç söz konusu ise Birliktelik Kuralları algoritmaları kullanılabilir.

Üstünlükler: Yaygın olarak betimleyici istatistiklerin (ortalama, standart sapma, frekans dağılımları vb.) kullanımıyla aykırı değer tespiti subjektif olarak, gözleme dayalı belirlenmektedir. Ancak, objektif bir karar alma süreci için bilimsel anlamda geçerliliği herkes tarafından kabul edilebilir bir norm ortaya konulması gerekmektedir.

Betimleyici istatistiklerden uzman destekli suistimal belirleme büyük veri tabanları üzerinde pratikte imkansız olmasının yanısıra; Veri Madenciliği yöntemleriyle sayıların tartışmasız objektivitesi norm belirmede esas alınmaktadır.

7.10. Maliyete Etki Eden Faktörlerin Belirlenmesi ve Maliyetleri Minimize Edici Yol Haritalarının Belirlenmesi

Hizmet maliyetlerinin belirlenmesi sağlık hizmetini sunan ve satın alanlar açısından oldukça önemlidir. Sağlık hizmetleri sunucuları maliyetleri denetim altına alabilmek için maliyete etki eden faktörleri bilmeli ve maliyetleri minimize edici çözümler üretmelidir [36].

Veri Madenciliği Çözümü: Tüm potansiyel maliyet faktörleri içerisinden önemli düzeyde etkiye sahip olanlar belirlenerek; aralarındaki ilişkiler tanımlanabilir. Ayrıca, sadece bütünsel anlamda değil; alt gruplar bazındaki ayrışmalar da tespit edilebilir. Bu amaçla, Karar ağaçları, Önemli Bileşenler Analizi veya Faktör Analizi kullanılabilir.

Üstünlükler: Sabit ve değişken giderler yanında diğer değişkenlerin de birlikte değerlendirilmesi ve

değişkenlerin etki düzeylerinin belirlenebilmesi mümkün olmaktadır.

Karar ağaçlarının kullanımıyla, değişkenlerin etki düzeylerinin yanı sıra; yol haritaları çıkarmak da mümkün olmaktadır.

7.11. Finansal Performans ve Riskin Belirlenmesi ile Finansal Erken Uyarı Sistemlerinin Geliştirilmesi

Sağlık sektörüne ayrılan kaynakların hastane ve sağlık kurumlarının tümünde etkin kullanımı açısından finansal performansın ölçümü ve finansal risklerin belirlenmesi zorunludur [29]. Finansal performansın düşmemesi ya da finansal kriz yaşanmadan önlemler alınması için kullanılabilecek en pratik yöntem ise finansal erken uyarı sistemleridir [37, 38, 39, 40, 41].

Veri Madenciliği Çözümü: Finansal performansa yönelik norm belirlenmesi için profillendirme yaklaşımı izlenmesi gerekmektedir. Bu amaçla Karar Ağaçları kullanılabilir.

Üstünlükler: Yöneticilerin finansal değişkenler içinde kaybolmasını önleyerek, diğer istatistiki ve finansal yöntemlere göre çok daha objektif sonuçlar sunmaktadır. Ayrıca, temel değişkenleri ve riskleri kullanarak erken uyarı sinyallerine ulaşma imkanı sağlamaktadır.

7.12. Yönetsel Karar Destek Sistemlerinin Geliştirilmesi

Sağlık yöneticileri, sağlık kurumlarının daha etkin, verimli ve kaliteden ödün vermeden yönetimi için mevcut verileri en iyi şekilde kullanan ve karar verme sürecine destek olacak sistemlere ihtiyaç duymaktadırlar [42].

Veri Madenciliği Çözümü: İhtiyaç duyulan karar değişkenine ilişkin tanımlamalarda;

- Model tanımlama (ÖBA, Faktör Analizi, Regresyon Modeller, Sinir Ağları vb.)
- Verimlilik, kalite ve risk göstergelerinin belirlenmesi esas alınabilir. (Karar ağaçları, Birliktelik Kuralları, K-ortalamalar Kümeleme Analizi'ni takiben Lojistik regresyon veya Sinir Ağları Modellemesi).

Üstünlükler: Hastanelerde yönetsel amaçlı kullanılabilecek tüm değişkenler çok boyutlu olarak ele alınabilir, optimal değerler ve yol haritaları belirlenebilir.

8.SONUÇ

Bu makalede Veri Madenciliği konusunda bir altyapı oluşturmak ve sağlık profesyonellerine sağlık sektöründe Veri Madenciliği'nin kullanımı ile ilgili örnekler sunarak karar verme süreçleri açısından yeni bir bakış açısı kazandırmak amaçlanmıştır. Örnekler ülkemizde sağlık sektöründeki öncelikli konu ve sorun alanları dikkate alınarak sunulmuştur. Veri Madenciliği'nin sağlık sektöründe kullanımını bu örnekler ile sınırlamak mümkün değildir. Veri Madenciliği'nin sağlık sektöründeki diğer kullanım alanları olarak [37]:

- 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 profillendirilmesi ve Türkiye ilaç kullanım haritasının hazırlanması,
- Kronik hastalıklarda veri madenciliğine dayalı olarak ilaç kullanım alışkanlıkları ve risk tespiti,
- İlaç birim maliyetlerinin hesaplanması,
- İlaç inovasyon maliyetlerinin belirlenmesi
- Bioterörizme karşı sağlık veritabanı oluşturulması,
- Afet telafisinde önceliklerin ve minimum maliyetlerin belirlenmesi

gibi örnekler verilebilir.

Veri Madenciliği, sağlık profesyonellerinin en doğru ve güncel bilgiye ulaşmasını, en objektif ve optimum çözümleri kullanmasını sağlayacak bir karar destek aracıdır. Geleceğin sayısal karar verme ve iş zekası yöntemi olan Veri Madenciliğinin konunun uzmanı kişiler tarafından sağlık sektöründe kullanımı, sağlık hizmetlerinin daha etkin sunumu, kaynakların daha verimli kullanımı ve bilimsel, karşılaştırılabilir, şeffaf bilgi erişimi açısından önerilmektedir.

KAYNAKLAR

[1] G. Piatetsky-Shapiro, W. J. Fawley, "**Knowledge Discovery in Databases",** AAAI/MIT Pres, 1991.

[2] U. Fayyad, G. Piatetsky-Shapiro, P. Symth, P. "From Data Mining to Knowledge Discovery in Databases", AI Magazine, 17(3), 37-54, 1996.

[3] O. R. Zaine, "Principles of KDD". Ph. D. Thesis (Unpublished). University Of Alberta, Department of Computing Sciences, 1999.

[4] A.S. Koyuncugil, "Bulanık veri madenciliği ve sermaye piyasalarına uygulanması", Doktora tezi (basılmamış), Ankara Üniversitesi, Fen Bilimleri Enstitüsü, 2006.

[5] R. Brachman, T. Anand, "The Process of Knowledge Discovery in Databases: A Human-Centered Approach" Advances in Knowledge Discovery and Data Mining, ed. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy, AAAI/MIT Press 1996.

[6] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, A. Zanasi, "Discovering Data Mining: From Concept To Implemantation", Prentice Hall PTR, Upper Saddle River, New Jersey, 195, USA. 1997.

[7] Internet, Two Crows Corporation, Introduction To Data Mining And Knowledge Discovery, Third Edition, http://www.twocrows.com/intro-dm.pdf.

[8] A. Berson, S. Smith, K. Thearling, "Buildind Data Mining Applications for CRM",. Mcgraw Hill, 510, USA, 1999.

[9] T. Hastie, R. Tibshirani, J. Friedman, "The Elements Of Statistical Learning; Data Mining, Inference And Prediction", Springer Series In Statistics, 533, USA, 2001.

[10] L.T. Moss, S. Atre, "Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications",. Addison-Wesley Publishing, 576, USA, 2003.

[11] R. J. Roiger, M. Geatz, "Data Mining: A Tutorial Based Primer", Addison-Wesley Publishing, 2002.

[12] Internet, K. Thearling, www.thearling.com. Erişim Tarihi: 18.06.2005.

[13] Z. Chen, "Data Mining And Uncertain Reasoning: An Integrated Approach",. John Wiley & Sons, Inc., 370, Canada. 2001.

[14] B. Kovalerchuk, E. Vityaev, "Data Mining in Finance: Advances in Relational And Hybrid Methods". Kluwer Academic Publishers, 308, USA. 2001.

[15] Sağlık Bakanlığı, www.saglik.gov.tr. Erişim Tarihi: 18.05.2009.

[16] J. A. Lyman, K. Sclully, J. H. Harrison, "The Development Of Health Care Data Warehouses To Support Data Mining", Clinics In Laboratory Medicine, 28(1), 55-71, DOI: 10.1016/J.Cll.2007.10.00 3, 2008.

[17] S. N. Stolba, A. M. Tjoa, "The Relevance Of Data Warehousing And Data Mining In The Field Of Evidence-Based Medicine To Support Healthcare Decision Making", "Proceedings Of World Academy Of Science", Engineering And Technology, 11, ISSN 1307-6884, 2006.

[18] J. C. Prather, D. F. Lobach, L. K. Goodwin, J. W. Hales, M. L. Hage, W. E. Hammond, "Medical Data Mining: Knowledge Discovery In A Clinical Data Warehouse", Proc AMIA Annu Fall Symp. 101–105, 1997.

[19] A.S. Koyuncugil, N Özgülbaş, "Sağlıkta Veri Madenciliğiyle Anti-Fraud Uygulamalar: Veri Sorunlarının Çözümü İçin Veri Madenciliğine Dayalı Profillendirme Yaklaşımı", IV. Ulusal Tıp Bilişimi Kongresi, Antalya, 2007.

[20] K. M. Obenshain, "Application of Data Mining Techniques To Healthcare", Data *Infect Control Hosp Epidemiol*, 25, 690–695, DOI: 10.1086/502460, 2004.

[21] C.P. Subbe, M. Kruger, P. Rutherford, L. Gemmel, "Validation of A Modified Early Warning Score in Medical Admissions", QJM, 94(10), 521, 2001.

[22] W. Yang, S. Hwang, "A Process-Mining Framework for The Detection Of Healthcare Fraud And Abuse", Expert Systems with Applications, 31(1), 56-68, DOI:10.1016/J.Eswa.2005.09.003, 2006.

[23] S. Holt, "IBM's Health-Care System Helps Detect Fraud And Abuse", Infoworld, 19(35), 40, 1997.

[24] S. Ağabeydi, "Knowledge Management in Healthcare: Towards 'Knowledge-Driven", Decision-Support Services, 63(1), 5-18, 2001.

[25] A.S. Koyuncugil, N. Ozgulbas, "Donor Research and Matching System Based on Data Mining in Organ Transplantation", Journal of Medical Systems, DOI 10.1007/s10916-008-9236-7, 2008.
[26] Y. M. Chae, H. S. Kim, K. C. Tark, H. J. Park, S. H. Ho, "Analysis of Healthcare Quality Indicator Using Data Mining And Decision Support System", Expert Systems with Applications, 24(2), 167-172, 2003.
[27] J. J. Fried, "Data Mining For Quality Care", Commun Oncol, 3(1), 51, 2006.
[28] I. E. Allen, C. A. Seaman, "Data Mining For Quality", Quality Progress, 39 (2), 70-74, 2006.
[29] M. A. Fitzpatrick, "Using Data To Drive Performance Improvement In Hospitals", Health Management Technolog,; 27(12), 10, 2006.
[30] B. J. Vason, "Mine Data To Discover Infection Control Trends", Nursing Management, 35(6), 46, 2004.
[31] M. F. Wisniewski, P. Kieszkowski, B. M. Zagorski, W. E. Trick, M. Sommers, R. A. Weinstein, "Development Of A Clinical Data Warehouse For Hospital Infection Control", J Am Med Inform Assoc.10, 454-462, DOI 10.1197/Jamia.M1299, 2003.
[32] A. Milley, "Healthcare And Data Mining", Health Management Technology. 21(8), 44-46, 2000.
[33] S. E. Brossette, A. P. Sprague, J. M. Hardin, K. B. Waites, W. T. Jones, S. A. Moser, "Association Rules And Data Mining In Hospital Infection Control And Public Health Surveillance", Journal of The American Medical Informatics Association, 5(4), 1998.
[34] L. Sokol, B. Garcia, J. Rodriguez, M. West, K. Johnson, "Using Data Mining To Find Fraud In HCFA Health Care Claims", Top Health Inf, Manage, 22(1), 1-13, 2001.
[35] C. Dorn. "Data Mining Technology Helps Insurers Detect Health Care Fraud National Underwriter", Life & Health, 108(39), 34-36, 2004.
[36] M. Silver, T. Sakata, H.Su, C. Herman, S. B. Dolins, M. J. O'Shea, "Case Study: How To Apply Data Mining Techniques In A Healthcare Data Warehouse", Journal Of Healthcare Informatıon Management, 15(2), 2001.
[37] A. S. Koyuncugil, N. Ozgulbas, "Early Warning System for SMEs as a Financial Risk Detector" Data Mining Applications for Empowering Knowledge Societies. Hakikur Rahman, Ed, Idea Group Inc., USA, 221-240, 2008.
[38] N. Özgülbaş, A. S. Koyuncugil, "Sağlık Kurumlarında Finansal Performans Ölçümü: Kamu Hastanelerinin Veri Madenciliği ile Sınıflandırılması", İktisat, İşletme ve Finans Dergisi, Sayı:253, Nisan. 2007.
[39] A. S. Koyuncugil, N. Özgülbaş, "Hastaneler İçin Veri Madenciliği İle Finansal Erken Uyarı Sinyallerinin ve Yol Haritalarının Belirlenmesi" Sağlık ve Hastane İdaresi Kongresi, Antalya, 22-26 Ekim, 2008.
[40] N. Özgülbaş, A.S. Koyuncugil, "Financial Profiling of Public Hospitals: An Application by Data Mining", The International Journal of Health Planning and Management, Vol:22, 69-83, 2009.
[41] N. Ozgulbas, A. S. Koyuncugil, "Hospital Early Warning System For Detecting Financial Performance and Risk Indicators of Ministry Health Hospitals", International Conference on Healtcare Performance and Quality, Antalya, 19-21 Mart, 2009.
[42] A. S. Koyuncugil, N. Ozgulbas, "Detecting Road Maps for Capacity Utilization Decisions by Clustering Analysis and CHAID Decision Trees", Journal of Medical Systems, 10.1007/s10916-009-9258-9, 2009.