



Derleme (Review)

Cilt 2- Sayı 2: 45-48 / Mayıs 2019

(Volume 2- Issue 2: 45-48 / May 2019)

TÜRKİYE'DE SAĞLIK SEKTÖRÜNDE VERİ MADENCİLİĞİ KULLANIM ALANLARI

Müberra TERZİ^{1*}

¹Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Akıllı Sistemler Mühendisliği, Samsun, Türkiye

Gönderi: 28 Aralık 2018; **Kabul:** 19 Şubat 2019; **Yayınlanma:** 01 Mayıs 2019

(Received: December 28, 2018; **Accepted:** February 19, 2019; **Published:** May 01, 2019)

Özet

Artan veri tabanları ile beraber hayatımıza giren big data ilerleyen dönemlerde veri madenciliği ifadesini oluşturmuştur. Veriler her geçen gün artmakta ancak işlenmediği ve anlamlı hale getirilmediği sürece hiçbir şey ifade etmemektedir. Bu nedenle insanlar verileri analiz ederek anlamlı hale getirmiş böylece işletmelere farklı bakış açıları geliştirmişlerdir. Ayrıca veri madenciliğini yaptıkları işlere yardımcı olması için bir araç olarak da kullanmışlardır. Bu derlemede veri madenciliği hakkında bilgi edinip veri madenciliği metodlarının ülkemiz sağlık sektöründe nasıl kullanıldığını ve hangi alanlarda kullanılabileceğine göz atmış olacağız.

Anahtar kelimeler: Veri Madenciliği, Türkiye, Sağlık


Data Mining Applications in Health Sector in Turkey

Abstract: Big data entering our lives with increasing databases formed data mining expression in the following periods. Data is increasing every day, but does not mean anything unless it is processed and made meaningful. For this reason, people have made data meaningful by analyzing the data so that they can develop different perspectives. They also used data mining as a tool to help with their work. In this review, we will learn about data mining and look at how data mining methods are used in our country's health sector and in which areas they can be used.

Keywords: Data mining, Turkey, Health

***Corresponding author:** Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Akıllı Sistemler Mühendisliği, Samsun, Türkiye

E mail: muberraterzi61@gmail.com (M. TERZİ)

Müberra TERZİ  <https://orcid.org/0000-0002-3939-4268>

Cite as: Terzi M. 2019. Data mining applications in health sector in Turkey. BSJ Health Sci, 2(2): 45-48.

1. Giriş

Gelişen teknoloji ve artan veri tabanı nedeniyle insanlar verileri anlamlı hale getirmek istediler. Bu nedenle ortaya çıkan veri madenciliği kavramı birçok kişi tarafından farklı tanımlanabilmektedir. Genel bir tanım olması açısından veri madenciliği; depolanan verilerden matematiksel ve istatistiksel metodlar yardımı ile anlamlı

bilgiler keşfedebilmektir (Altıntaş, 2010). Diğer kaynaklardaki veri madenciliği tanımlarına bakacak olursak; Tang ve MacLennan' a göre verilerin analiz edilmesi ve gizli kalıpların otomatik veya yarı otomatik araçların yardımı ile bulunmasıdır (Tang ve MacLennan 2005). Al-Baidhani'ye göre verilerdeki kalıpları bulmak için akıllı yöntemlerin kullanılma stildir (Al-Baidhani, 2017). Verileri analiz etme işlemine veri madenciliği

diyebilmemiz için veri setimizin büyük olması ve sistemin sorularımıza cevap verebilir nitelikte olması gerekmektedir. Aynı zamanda veri madenciliği problemleri modelleyerek gösteren bir sistem olarak da düşünülmektedir (Irmak ve ark., 2012).

Veri madenciliği firmalara benzer gruplardaki verilerini sınıflandırma ve bu verilerin gerçek hayata optimize edilerek sunulması konusunda destek olur (Savaş ve ark., 2012).

İnsanoğlu her gün verilerle beraber onları yorumlayarak işine yarayacak hale getirmeye çalıştı. Başlangıç aşaması ENIAC (ilk sayısal bilgisayar) olarak düşünülmektedir (Şimşek, 2012). Kavramsal olarak 1960'larda veri toplama süreci olarak başlayan süreçte bilgisayar, teyp, disk gibi teknolojik aletler kullanılmış ve geçmişe yönelik veriler toplanmıştır. Ancak bu veriler statik veri kavramına uymaktadır. (Dünyadaki hasta sayısı gibi) 1980'lerde veri erişimi haline gelen süreçte SQL, ilişkisel veritabanları gibi sistemler kullanılmaya başlanmış olup toplanan veriler de dinamik veri dağıtımları yapılabilmektedir. (Hastaların Türkiye'deki sayısı gibi) 1990'larda Veri ambarları ve karar destek sistemleri oluşmuş olup OLAP, Veri Ambarı teknolojileri kullanılmaya başlandı. Bu sistemler ile geriye dönük veriler daha fazladır. Günümüzde Veri Madenciliği adını alan bu sistem ile beraber İleri düzey algoritmalar, Çok işlemci bilgisayarlar kullanılmaya başlandı. Artık geçmişe dönük verileri kullanarak geleceğe yönelik tahminleme yapılabilmektedir (Kundereli, 2012).

Bu çalışmanın amacı, sağlık sektöründe veri madenciliği üzerine yapılan son yıllardaki çalışmalarını inceleyerek bir araya getirilmesidir.

2. Veri Madenciliği Metotları

Veri madenciliği için izlenmesi gereken bazı adımlar vardır. Bu adımları incelediğimizde karşımıza çıkan durumları şu şekilde özetleyebiliriz. Bunlar:

Problemin tanımlanması: Öncelikle yapılan veri madenciliği çalışmasının hangi kuruma ne amaçla yapılacak olduğunun belirlenmiş olması gerekmektedir (Şimşek, 2012). Veri madenciliğinin en zor ve en önemli adımı amacın tanımlanmış olmasıdır. Amacı yanlış belirlenmiş bir problemin sonucu da yanlış olacak ve bulduğumuz cevapla aradığımız cevap arasında fark olacaktır (Diler, 2016).

Verilerin hazırlanması: Elimizde ham veri seti bulunmaktadır. Bu ham veriden kullanılacak olan veri kümesi hazırlanır. Analiz etmede kullanacağınız değişkenler bu aşamada belirlenmelidir. Değişkenler üzerinde dönüşümler yapılabilir. Aynı zamanda elimizde bulunan ham veri seti modelleme araçlarını kullanmaya hazırlanmalı ve gereksiz yerler temizlenmelidir (Larose, 2005).

Modelin kurulması ve değerlendirilmesi: Problem

amacına ve verilerine göre en uygun modeli kurmak gerekmektedir. Buda çok sayıda model denemesi ile olur (Akpınar, 2000). Uygun modelleme teknikleri uygulanarak sonuçları optimize edecek olan ayarlar yapılır. Bir veri madenciliği uygulamasında aynı anda çok sayıda teknik kullanılabilir (Larose, 2005). Modelin test edilmesinde kullanılacak Simple Validation (basit geçerlilik), Cross Validation (çapraz geçerlilik), N-Fold Cross Validation (n katlı çapraz geçerlilik) gibi yöntemler vardır (Akpınar, 2000).

Modelin kullanılması: Kurulup geçerlilik bakımından test edilmiş olan model uygulama olabilir. Ya da herhangi bir uygulamanın alt basamağı olarak kullanılabilir (Savaş ve ark., 2012).

Modelin izlenmesi: Zaman geçtikçe kurulan model içerisindeki veriler değiştiğinden dolayı model sürekli izlenmeli ve gerektiği durumlarda tekrar düzenleme yapılmalıdır (Savaş ve ark., 2012).

İzlenen bu adımlarla birlikte kullanılan metotlar vardır (Albayrak, 2017). Veri madenciliği sınıflandırma, kümeleme ve birliktelik kuralları olmak üzere üç ana gruba ayrılmıştır. Bunlarında kendi içlerinde alt başlıkları olup Şekil 1'de başlıklar halinde verilmiştir.

Sınıflandırma verilerin niteliklerine göre başta belirlenmiş olan sınıflara atanmasına denir (Şimşek, 2012). Dört ana başlığa ayrılmış olup bunlar; Karar ağaçları, Mesafeye dayalı yöntemler, İstatistiğe dayalı yöntemler ve Diğer yöntemlerdir. Bunlarda kendi içlerinde alt başlıklara ayrılmış olup tabloda gösterilmiştir.

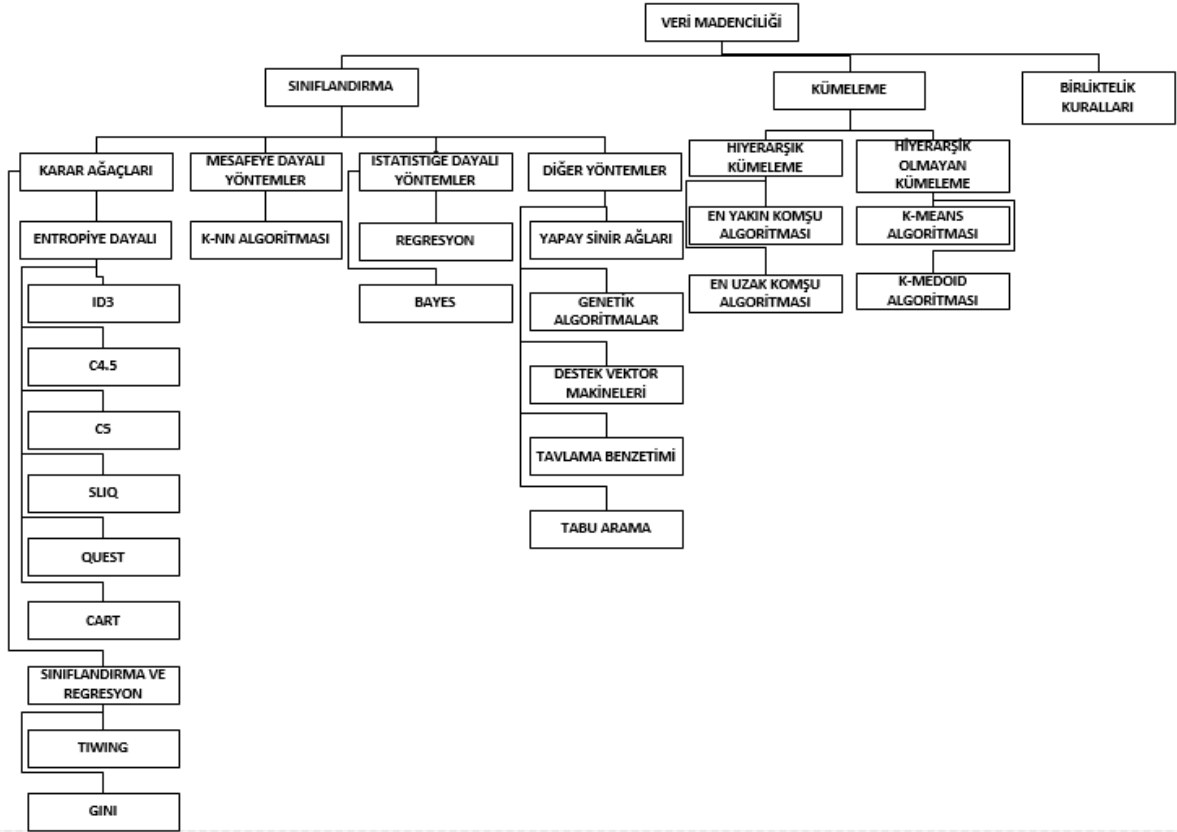
Kümeleme birbirine benzeyen özelliklerin belirlenerek benzer grupların oluşturulması işlemidir(Şimşek, 2012). Grup sayısının bizim belirlediğimiz hiyerarşik kümeleme ve grup sayısını bizim belirleyemediğimiz hiyerarşik olmayan kümeleme olmak üzere iki ana başlığa ayrılmıştır. Bu başlıkların alt başlıkları tabloda gösterilmektedir.

Birliktelik kuralları ise bir araya gelmiş olan verilerin özelliklerinin fark edilmesidir (Şimşek, 2012).

Veri madenciliği birçok alanda kullanılmaktadır. Bu alanlardan bazıları finans, pazarlama, sigortacılık, iletişim, borsa, sağlık, sanayi, bilim ve mühendisliktir (Albayrak, 2017). Derlemede bu alanlardan sağlık sektörü incelenmiştir.

3. Sağlık Sektöründe Veri Madenciliğinin Kullanım Alanları

Sağlık sektörü hem big data'nın olduğu hem de sürekli değişen verilerin olduğu bir alandır. Bu alanda çok fazla çalışma yapılmış olup bu çalışmaların bazılarını ve sağlıkçıların veri madenciliğinden istediklerine aşağıda ulaşacaksınız (Koyuncugil ve Özgülbaş, 2009).



Şekil 1. Veri madenciliğinin sınıflandırılması.

3.1. Hastalık Teşhisi ve Tanısına Yönelik Uygulamalar

Uçar'ın 2014 yılında yaptığı çalışmaya göre hastane sistemleri sağlık uzmanlarına yol gösterici ve verdiği kararları destekler şekilde olmalıdır. Uçar çalışmasını Ege Üniversitesi Tıp fakültesinde yapmış olup hipertansiyon ve diyabet hastalarında böbrek yetmezliğinin olup olmadığı kararını veren sağlıkçılara yardımcı olmak amaçlıdır. Veri madenciliği çalışmasını 74322 adet kaydı Naive Bayes algoritmasına göre eğiterek 7296 kaydı test amaçlı kullanmış ve %78 başarı elde etmiştir (Uçar, 2014) Erkuş 2015 yılında yaptığı çalışmada kalp ve damar hastalıklarına anjio yapmadan teşhis koymak için bir model geliştirmeyi hedeflemiştir. Bunun için kalp hastalığının tanısı konulan ve koyulmayan 604 adet veriden oluşan bir veri seti hazırlamıştır. Veri olarak biyokimya test sonuçları kullanılmıştır. Parametreler belirlenmiş ve bu parametreleri göre üç farklı veri grubu oluşturulmuştur. Model için farklı yöntemler denenmiş olup sonuçta en uygun parametre belirlenmiştir. Daha sonra parametrelere sınıflandırma algoritmaları uygulanmış olup başarı oranı tüm parametrelerde en yüksek olan algoritma HNB Algoritması olarak bulunmuştur. Gelecekte biyokimya sonuçlarına göre modelin otomatik olarak çalışıyor olması istenmektedir (Erkuş, 2015).

Fayez'in 2018 yılındaki çalışmasına göre Koroner kalp hastalığı ölümlerine sonuçlandırdığı için çok fazla dikkat çekilen bir konuydu. Veri madenciliğinin sınıflandırma tekniğini ve Phyton programlama dilini kullanarak koroner kalp

hastalığı teşhisinde maliyet ve zaman olarak daha iyi bir sistem tasarlayan Fayez çalışmasında yüksek oranda doğruluk elde etmiş. Çalışmasını farklı koroner kalp hastalığı verilerinde uygulayarak Random forest algoritmasında %99, Cleveland'da %94, SVM algoritmasında %58 sonuç elde etmiştir (Fayez, 2018). Altun'un 2018 yılında yaptığı çalışmada 2009 - 2017 yılları arasında beyin tümörü teşhisi konulan ve koyulmayan hastaların MR verileri kullanılmıştır. Çalışmaya göre MR Spektroskopide ölçülen kolin metabolit değerini çok önemlidir. Hazırlanan veri seti WEKA programı yardımı ile çözülmüştür. Sınıflandırmanın beş yöntemi denenmiş olup bunlardan kNN Algoritması %95.69 oranı ile en yüksek başarıyı elde etmiştir. Kümeleme yöntemlerinden iki tanesi kullanılmış olup bunlar arasından da %98.92 ile K Ortalama en yüksek başarıyı elde etmiştir. Derin öğrenmeye göre ise başarı oranı %93.54 olmuştur. Çalışmanın sonuçları kontrol ettirilmiş olup tümör tespitinde alternatif bir yöntem olabileceğine karar verilmiştir (Altun, 2018).

3.2. İlaçlara Yönelik Uygulamalar

Çiçek'in 2014 yılında yapmış olduğu çalışmaya göre diyabet hastaları üç ayda bir kan şekerini ölçtürmelidir. Yapılan çalışmanın amacı da bu ölçümlere göre ilaç dozlarının uygun bir şekilde ayarlanmasında doktorlara alternatif bir model tasarlamaktır. Parametreler belirlenip model oluşturulmuş ve model WEKA'da çözülmüştür. Model sonucunda görülmüştür ki veri madenciliği ilaç dozu ayarlamada kullanılacak bir

yöntemdir (Çiçek ve ark., 2014).

Çalış'ın 2010 yılında yaptığı çalışmaya göre yüksek tansiyon hastalarının ilaç dozlarına karar verirken baktığı bazı parametreler vardır. Çalış, yaptığı çalışmada; veri madenciliği tekniklerinden yararlanarak bu çalışma ile ilaç dozunu ayarlamayı hedeflemiştir. ANFIS ve Rough Set yöntemlerini kullanan Çalış parametrelerini sisteme girmiş ve dört tane ilaç için doz planlaması yapmıştır. Model sonucunda görmüştür ki ANFIS daha güvenilir sonuçlar vermektedir (Çalış, 2010).

Yukarıda verilen çalışmalar dışında birçok çalışma mevcuttur. Aynı zamanda veri madenciliği sağlık sektöründe farklı alanlarda da kullanılıyor olup iş zekası ile çözüme ulaştırabileceğimiz alanlar da vardır. Bunlardan bazıları sağlık çalışanlarının performansını izleme, hasta akış planları oluşturma, tıbbi tedavi süreçlerini optimize etme, hasta ilaç profilendirmesi ve ülkemizde ilaç kullanma haritalarının oluşturulmasıdır (Koyuncugil ve Özgülbaş, 2009).

4. Sonuç

Bu derlemede veri madenciliği hakkındaki bilgiler tek çatı altında toplanmaya çalışılmış, ülkemiz sağlık sektöründe veri madenciliği uygulamalarına ve kullanım alanlarına göz atılmıştır. Derlemedeki örnek uygulamalar ülkemizde yapılan çalışmalardan seçilmiş olup hastalık riski üzerine yapılan çalışmaların daha fazla olduğu gözlemlenmiştir. Hastalık riski üzerine olan çalışmalardan sonra ilaç kullanımlarıyla ilgili olan çalışmalar dikkat çekmektedir. Veri madenciliğinin sağlık sektöründeki çalışmaları hastalık riski ve ilaç dozu ile sınırlı olmayıp sağlık sektöründe çalışanların beklentileri derleme içerisinde verilmiştir. Çalışmaların sonuçları yüksek başarı oranı içermekte olup veri madenciliğinin sağlık çalışanlarına verileri yorumlamalarında yardımcı bir araç olabileceği düşünülmektedir.

Çıkar ilişkisi

Yazar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Kaynaklar

Akpınar H. 2000. Veri tabanlarında bilgi keşfi ve veri madenciliği.

İstanbul Üniv İşletme Fak Derg, 29(1): 1-14.

Al-Baidhani AHJ. 2017. Estimation of heart disease based on data mining using patients health database. Ulusal Tez Merkezi, 497861: 3-5.

Albayrak M. 2017. Bilimsel araştırmalarda veri madenciliği kullanımı. IJSSER, 3(3): 752-756.

Altıntaş YY. 2010. Veri madenciliğinin tıpta kullanımı ve bir uygulama: hemodiyaliz hastaları için risk seviyelerine göre risk faktörlerinin etkileşimlerinin incelenmesi. Ulusal Tez Merkezi, 269710: 1-3.

Altun S. 2018. Mr spektroskopisi temelli beyin tümörü teşhisinde veri madenciliği uygulamaları. Ulusal tez merkezi, 505759: 49-60.

Çalış Ç. 2010. Drug dosage planning of hypertension disease using data mining techniques. Ulusal tez merkezi, 266487: 1-55.

Çiçek G, Tufan K, Erol H. 2014. Applying data mining techniques to implement the clinical guidelines for the management of the patients with type 2 diabetes. Med Dose Adjust, 1-9.

Diler S. 2016. Veri madenciliği süreçleri ve karar ağaçları algoritmaları ile bir uygulama. Ulusal Tez Merkezi, 433080: 10-14.

Erkuş S. 2015. Veri madenciliği yöntemleri ile kardiyovasküler hastalık tahmini yapılması. Ulusal tez merkezi, 392880: 34-62.

Fayez MA. 2018. Diagnoses of coronary heart disease (Chd) using data mining techniques based on classification. Ulusal tez merkezi, 520268: 1-54.

Irmak S, Köksal CD, Asilkan Ö. 2012. Hastanelerin gelecekteki hasta yoğunluklarının veri madenciliği yöntemleri ile tahmin edilmesi. Inter J Alanya Fac Busin, 4(1): 101-114.

Koyuncugil AS, Özgülbaş N. 2009. Veri madenciliği: tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. Bilişim Teknoloji Derg, 2(2): 21-37.

Kumdereli ÜC. 2012. Tıp bilişimi ve veri madenciliği uygulamaları: eeg sinyallerindeki epileptiform aktiviteye veri madenciliği yöntemlerinin uygulanması. Ulusal Tez Merkezi, 318331: 46-48.

Larose TD. 2005. Discovering knowledge in data. Wiley Interscience, p. 5-8.

Savaş S, Topaloğlu N, Yılmaz M. 2012. Veri madenciliği ve Türkiye'deki uygulama örnekleri. İstanbul Ticaret Üniv Fen Bilim Derg, 21(1): 2-17.

Şimşek MU. 2012. Sosyal ağlarda veri madenciliği üzerine bir uygulama. Gazi Üniversitesi. Ulusal Tez Merkezi. 321573:5-21.

Tang Z, MacLennan J. 2005. Data mining SQL server 2005. Wiley Publishing, p. 2-5.

Uçar ED. 2014. Data mining using in diagnosis of the chronic renal disease in diabetic patients. Ulusal tez merkezi, 376344: 12-61.