

Sağlık Hizmetlerinde Veri Madenciliğinin Kullanımı: Türkiye’de Lisansüstü Tezlerinin İncelenmesi

Use of Data Mining in Healthcare: Investigation of Graduate Thesis in Turkey

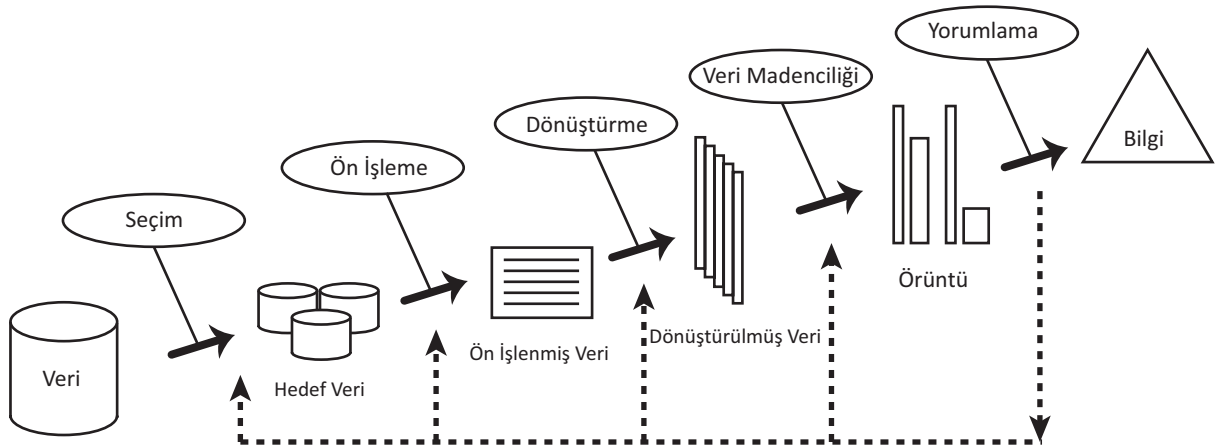
Sema DÖKME YAĞAR¹ *¹ Doktora Öğrencisi, Ankara Üniversitesi, Sağlık Bilimleri Fakültesi, Sağlık Yönetimi Bölümü

ARTICLE INFO	ÖZET
Makale Türü: Sistematiik Derleme Anahtar Sözcükler: Veri Madenciliği, Sınıflandırma, Kümeleme, Birliktelik Kuralları, Sağlık Keywords: Data Mining, Classification, Clustering, Association Rules, Health Sorumlu Yazar Sema DÖKME YAĞAR Adres: Ankara Üniversitesi, Sağlık Bilimleri Fakültesi, Sağlık Yönetimi Bölümü, Ankara, Türkiye E-mail: semadokme@ gmail.com	Büyüyen veri yığınları arasından ihtiyaç duyulan bilgilere ulaşmak her geçen gün zorlaşmaktadır. Günümüz rekabet koşullarında doğru bilgiye ulaşmak yöneticiler için oldukça önemli hale gelmiştir. Bu büyük veri yığınları arasından doğru ve hızlı bilgiye ulaşmak için bilgisayar tabanlı bilgi sistemleri olan veri madenciliği geliştirilmiştir. Bu yöntemin diğer alanlarda olduğu gibi sağlık alanındaki kullanımının da her geçen gün arttığı bilinmektedir. Bu çalışmada, veri madenciliği ile ilgili sağlık alanında Türkiye’de son yıllarda hangi konularda çalışıldığını ve daha çok hangi yöntemlerin kullanıldığını belirlemek amaçlanmıştır. Bu kapsamda veri madenciliği ile ilgili Türkiye’de sağlık alanında yapılan lisansüstü tezler bazı temel kriterler dikkate alınarak incelenmiştir. Çalışmalarda en çok tekniklerin karşılaştırıldığı, hekimlerin ve yönetimlerin karar vermelerine yardımcı olma konularında çalışmalar yapıldığı tespit edilmiştir. Yapılan tezlerin büyük bir çoğunluğunun “sınıflandırma” yöntemleri başlığı altında yer aldığı görülmüştür. Sınıflandırma yöntemlerinden de en çok karar ağaçları ve Bayes sınıflandırması tekniklerinin kullanıldığı gözlemlenmiştir. Örneklem grubu olarak neredeyse araştırmaların tümünde hastaların dikkate alındığı ve en fazla kanser hastaları üzerinde araştırmalar yapıldığı görülmüştür. Türkiye’de yapılan lisansüstü tezlere bakıldığında, klinik araştırmalar ile ilgili çalışmaların yapıldığı fakat idari konularda ve sağlık çalışanları üzerinde yeterli çalışmaların yapılmadığı belirlenmiştir.
	ABSTRACT
	It is getting harder every day to reach the information we need among the growing data stack. In today’s competitive conditions, reaching the right information has become very important for managers. Data mining, which is a computer-based information system, has been developed in order to reach accurate and fast information among this large data stack. It is known that the use of this method in the field of health is increasing day by day, as in other areas. In this study, in health-related fields of data mining in recent years in Turkey it aimed to determine which studied the issues and what methods are being used more. In this context, data mining postgraduate studies in health related fields in Turkey thesis have been analyzed taking into consideration some basic criteria. It has been found that most of the studies are compared with the techniques, and studies are conducted on helping physicians and administrations make decisions. It has been seen that most of the were made under the title of “classification” methods. Among the classification methods, it has been observed that decision trees and Bayes classification techniques are mostly used. As a sample group, it has been observed that almost all of the studies take into account patients and mostly studies are on cancer patients. Looking at the graduate theses held in Turkey, where the research related to clinical trials, but it was observed that no adequate studies on the health care workers and administrative matters.

1. GİRİŞ

Veri boyutu günden güne büyümektedir. Büyük, karmaşık, bilgi açısından zenginleştirilmiş veri kümelerini anlama ihtiyacı teknoloji, iş dünyası ve bilimin çeşitli alanlarında artmıştır. Verilerden bilgi bulmak için yeni teknikler de dahil olmak üzere bilgisayar tabanlı bilgi sistemi uygulama süreci olarak ifade edilen veri madenciliği yaklaşımı ortaya çıkmıştır (Jothi ve Husain, 2015). Veri madenciliği, veri tabanlarında daha önce bilinmeyen kalıpları, eğilimleri bulma ve tahmin modelleri oluşturmada kullanılan süreç olarak tanımlanabilir. Ayrıca, daha önce bilinmeyen örüntüleri ortaya çıkarmak için geniş veri depolarını kullanarak veri seçme, keşif ve model oluşturma süreci olarak ifade edilmektedir (Koh ve Tan, 2011). Kısacası veri madenciliği, büyük veri kümelerinde ilginç, beklenmedik veya değerli yapıların keşfedilmesidir. Bu nedenle, iki farklı yönü vardır. Bunlardan biri büyük ölçekli, “küresel” yapılarla ilgilidir ve amaç dağılımların şekillerini veya özelliklerini modellemektir. Diğer küçük ölçekli, “yerel” yapılarla ilgilidir ve amaç, anormallikleri tespit etmek ve gerçek ya da şans olayları olup olmadığına karar vermektir (Hand, 2007).

Veri madenciliğinin tarihsel gelişimine bakıldığında altı dönem olduğu görülmektedir. Sayım için ilk bilgisayarlar 1950’lerde kullanılmıştır. Veri tabanı ve verilerin depolanması kavramları 1960’larda ortaya çıkmış ve basit öğrenmeli bilgisayarlar 1960’ların sonunda geliştirilmiştir. İlişkisel Veri Tabanı Yönetim Sistemleri uygulamaları 1970’ler de kullanılmıştır. Büyük miktarda veri içeren veri tabanları 1980’lerde oluşturulmuş ve 1990’larda oluşan bu büyük veri tabanlarından faydalı verilerin nasıl çekileceğinin tartışılması üzerine konu ile araştırmalar yapılmıştır. Araştırmaların sonucunda 1992 yılında veri madenciliği için ilk yazılım gerçekleştirilmiştir. Veri madenciliği sürekli olarak gelişim göstermiş ve 2000’li yıllarda her alanda uygulanmaya başlanmıştır (Savaş vd., 2012).



Şekil 1. Veri Tabanlarında Bilgi Keşfi

Kaynak: Emre, İ. E. ve Erol, Ç. S. (2017). Veri Analizinde İstatistik mi Veri Madenciliği mi?. International Journal of Informatics Technologies, 10(2): 161-167.

Şekil 1’de görüldüğü üzere, veri madenciliği temel olarak beş aşamadan oluşmaktadır. Bunlar; veri seçimi, ön işleme, dönüştürme, veri madenciliği ve yorumlama/değerlendirmedir. Kısaca açıklamak gerekirse (Yıldırım vd., 2008);

- **Veri seçimi:** Çalışılacak veri tabanlarının dikkate alınarak ilgili verilerin bu kaynaklardan çekilmesi ve veri dosyasının oluşturulmasıdır.
- **Ön işleme:** Bu aşamada eksik veya hatalı veriler düzeltilip; anormal veriler kaldırılır.
- **Dönüştürme:** Elde edilen veriler kategorize edilmeye çalışılır, ilgili özellikler seçilir ya da boyut azaltma yoluna gidilir.
- **Veri Madenciliği:** Bu aşamada veriler için uygun olan bir algoritma seçilir ve bu algoritma, dönüştürülen (hazırlanan) verilere uygulanır.
- **Yorumlama/Değerlendirme:** Veri madenciliği sonucunda elde edilen bilgiler ve özellikler yorumlanır ya da değerlendirilir.

Veri madenciliği ile ilgili bahsedilmesi gereken bir diğer önem husus ise, veri madenciliğini etkileyen etmenlerdir. Bu etmenler “veri, donanım, bilgisayar ağları, bilimsel hesaplamalar ve ticari eğilimler” olmak üzere beş başlık altında incelenebilir. Veri, bu yaklaşımın günümüzde bu kadar önemli hale gelmesindeki temel etken olarak görülmektedir. Donanım, birkaç yıl önce üzerinde madencilik yapılamayan veriler ile çalışma imkânı sağlamıştır. Bu durumda gelişen belleklerin ve işlem hızı kapasitelerinin etkisi çok büyüktür. Bilgisayar ağları, özellikle günümüzde internetin çok hızlı hale gelmesi ile birlikte farklı algoritmaları kullanma ve dağıtık verileri analiz etme noktasında olanaklar sağlamıştır. Bilimsel hesaplamalar ile birlikte bilgi keşfi, teori, deney ve simülasyonun birbirine bağlanması kolaylaşmıştır. Bilim adamlarına göre, bilimin üçüncü yolu olarak simülasyonun ortaya çıkmasına olanak tanımıştır. Son olarak, ticari eğilimler de işletmelere rekabet etmede avantajlar sağlamış, daha hızlı hareket etmelerini, daha yüksek kalitede hizmet sunmalarını ve bunları minimum maliyet ve az sayıda insan gücü ile yapmalarını sağlamıştır (Savaş vd., 2012).

2. VERİ MADENCİLİĞİ MODELLERİ

Veri madenciliği modelleri genel olarak iki kısımda incelenmektedir. Bunlar; tahmin edici modeller ve tanımlayıcı modellerdir. Bilinen verilerin dikkate alınmasıyla bilinmeyen bir değer tahmin edilmeye çalışılıyorsa tahmin edici modeller kullanılıyor demektir. Regresyon ve sınıflandırma analizleri bu modellere örnek olarak verilebilir. Diğer yandan, tanımlayıcı modeller aracılığıyla da verilerdeki gizli ortak özellikler ve ilişkiler araştırılmaktadır. Birliktelik kuralları ve kümeleme analizi gibi yöntemler örnek olarak gösterilebilir (Yıldırım vd., 2008).

2.1. Sınıflama ve Regresyon

Gelecek veri eğilimlerini tahmin etmek amacı ile kullanılan analiz yöntemleridir. Kategorik değerleri tahmin etme noktasında sınıflama, süreklilik gösteren verileri tahmin etme noktasında ise regresyon kullanılmaktadır. Örneğin, bir bankanın kredi uygulamalarını güvenli veya riskli olmalarına göre kategorize etmek istendiğinde sınıflama modeli kurulabilir. Diğer yandan, mesleği ve geliri belli olan bireylerin bilgisayar ürünleri alırken yapacağı harcamaları tahmin etmek içinde regresyon modeli kurulabilir (Özekes, 2003). Sağlık açısından bakıldığında, belirli bir tedavi yönteminin belirli bir hastayı tedavi edip etmeyeceğini, zarar vereceğini veya herhangi bir etkisi olmayacağını tahmin etmek için sınıflama yöntemleri kullanılabilir. Aynı zamanda hastaneye bir yıl içinde yapılması beklenen başvuru sayısını tahmin etmek içinde regresyon modellerine başvurulabilir. *“Hangi hastaların iyileşme olasılığı daha yüksek (protokole göre)?”* veya *“Kimin başka bir kalp krizi geçirmesi daha olasıdır?”* gibi soruların cevaplanması noktasında da hem sınıflandırma hem de regresyon yöntemleri dikkate alınabilir (Kudyba, 2010).

Regresyon ve sınıflama modelleri kapsamında kullanılan birçok teknik yer almaktadır. Karar ağaçları ve yapay sinir ağları en çok kullanılan teknikler olarak örnek gösterilebilir. *Karar ağaçları*, sınıflama modelleri içerisinde en yaygın olarak kullanılan tekniktir. Nedeni ise veri tabanlarına kolay bir şekilde entegre edilebiliyor olması, kolay yorumlanabilir olması, kuruluşunun ucuz olması ve güvenilir olmasıdır. Adından anlaşılacağı üzere, ağaç şeklinde bir yapıya sahiptir ve karar düğümü, yaprak ile dallardan oluşur. Karar ağacı işlemleri kök düğümler ile başlar ve yaprağa ulaşana kadar devam eder. Her karar düğümünde bir test gerçekleşir. Gerçekleşen test sonucunda veri kaybı yaşanmadan dallara ayrılma meydana gelir. Her düğümde test gerçekleşebilir. Dalın ucunda test gerçekleşmediğinde bir karar düğümü meydana gelir. Elde edilen sonuçta bir sınıflama oluşuyorsa, o dalın neticesinde bir yaprak meydana gelir. Bu yaprak ile birlikte, veri üzerinde belirlenmek istenen sınıflar oluşur (Özekes, 2003).

Yapay sinir ağları (YSA), temel olarak insan beyni örnek alınarak oluşturulmuş bir teknolojidir. YSA, beynin bir işlevi yerine getirme yöntemini modellemek için tasarlanan bir sistem olarak tanımlanmaktadır. Bu yapıda yapay nöronlar basit bir şekilde kümelendirilmektedir. Yapılan bu kümelendirmeler tabakalar haline getirilip daha sonrasında birbiri ile ilişkilendirilmektedir. Bu noktada dış mekan ile bağlantılı olan nöronlar bulunmaktadır. Bunun nedeni ise, bazı nöronların girdileri iletmeleri bazı nöronların çıktılarını iletmeleri gerekliliğidir. Bu nöronlar dışında kalanlar ise gizli tabakalardadırlar, sadece ağ içinde bir bağlantıları mevcuttur (Ataseven, 2013).

2.2. Kümeleme

Verileri kümelere ya da sınıflara ayırma işlemine kümeleme denilmektedir. Makine öğrenimi, biyoloji ve istatistik gibi birçok alanda kullanılmaktadır. Sınıflama modelinde olduğu gibi belirli bir veri sınıfı yoktur. Sınıfları bulunmayan veriler gruplar halinde kümelere ayrılırlar (Özekes, 2003). Örneğin, bir hastalık / hasta segmentasyon modeli oluşturmak, benzer özelliklere sahip hastalıkların belirgin semptomlarını bulmak için kullanılabilir. *“Yüksek riskli hastaları nasıl tanımlayabilirim?”* ya da *“Hizmet kalitesini (veya hasta memnuniyetini) nasıl artırabilirim?”* gibi soruların cevabını aranıyorsa kümeleme yöntemleri dikkate alınabilir (Kudyba, 2010).

Literatüre bakıldığında kümeleme ile ilgili birçok teknik bulunmaktadır. Bölme yöntemleri ve hiyerarşik yöntemler en çok kullanılan teknikler arasında gösterilmektedir. *Bölme yöntemlerinde*, “n” veri tabanındaki nesne sayısını, “k” oluşturulacak küme sayısını ifade etmektedir. Bölme algoritması ile birlikte “n” adet nesne, “k” adet kümeye bölünür. Bölünme sonucunda oluşan kümelerdeki nesneler birbirlerine benzer, farklı kümedekiler ise farklı olurlar. K-means yöntemi, k-medoids yöntemi ve bunların varyasyonları en çok bilinen bölme yöntemleridir. *Hiyerarşik yöntemlerde*, veri nesneleri kümeler ağacı şeklinde gruplara ayrılırlar. Aşağıdan yukarıya (agglomerative hiyerarşik) ve yukarıdan aşağıya (divisive hiyerarşik) olmak üzere iki farklı şekilde meydana gelebilmektedir. Aşağıdan yukarıya doğru olan kümelemede ilk olarak her nesne kendi kümesini oluşturur. Daha sonra bu atomik kümeler bir araya gelmeye başlar ve tüm kümeler bir kümede toplanıncaya kadar devam eder. Yukarıdan aşağıya doğru olan kümelemede ise, tam tersi bir durum söz konusudur. Tüm kümeler tek bir kümede yer alırlar, bu nesnelerden her biri tek başına küme oluşturuncaya kadar kümeler daha küçük parçalara bölünürler (Özekes, 2003).

2.3. Birliktelik Kuralları

Birliktelik kuralları, büyük veri kümeleri arasında birliktelik ilişkileri kurulması ile ortaya çıkmaktadır. Şirketler, veri tabanlarındaki topladıkları ve depoladıkları verileri dikkate alarak birliktelik kuralları oluşturmaya çalışırlar. En basit örnek market sepeti uygulamasıdır. Market yöneticileri süt alan bireyin ekmek alma olasılığını düşünerek bir araştırma yapıp, sonucunda oranı yüksek bulursa raflarını süt ile ekmeği yan yana olacak şekilde tekrar düzenleyebilir. Bu durumda ekmek satışlarını artırabilecektir (Özekes, 2003). Sağlıkta ise ilişkilendirme yönteminin potansiyel bir kullanımı sağlıkta gruplama (bir araya getirme, ilişkilendirme) analizidir. Pandemiik sürveyansta ve belirli hastalıklar için tedavi protokollerinin tanımlanmasında oldukça yararlı olan bir yöntemdir. Viral veya bakteriyel gibi bir grup oluşturma noktasında hangi hastalıkların birlikte olduğunu bulmak için kullanılabilir. Aynı şekilde, “Araştırma için hangi ilaçları kullanmalıyım (tedavi)?” sorusunun cevabını bulmak için de birliktelik kuralları dikkate alınabilir (Kudyba, 2010).

Birliktelik kurallarının üretilmesi noktasında en yaygın olarak kullanılan teknik *apriori algoritması*dır. Bu teknikte güven ve destek ölçütlerini karşılaştırabilmek için eşik değerler belirlenir. Uygulama sonucunda elde edilen değerlerin eşik değerine eşit ya da yüksek olması beklenir. Sonrasında destek sayıları hesaplanır ve eşik destek sayısından küçük olanlar elenir. Bu elemeyi sonra ortaya çıkan ürünler ikiye bölünür ve değerlendirilmeye alınır, tekrar sayıları oluşturulur ve eşik değerleri ile karşılaştırılarak altında olanlar çözümlemeyi çıkarılır. Bu süreçte ürünler üçerli ve dörderli gibi gruplandırılmalar yapılarak analize dahil edilir ve eşik değerler ile karşılaştırmalar yapılır. Süreç sonunda ürün grupları belirlenir, kural destek ölçütlerine bakılarak birliktelik kuralları türetilir ve bu kuralların her birisi ile ilgili olarak güven ölçütleri hesaplanır (Coşlu, 2013).

3. SAĞLIK HİZMETLERİNDE VERİ MADENCİLİĞİ VE UYGULAMALARI

Günümüzde sağlık kuruluşları büyük miktarda veri üretilip toplayabilmektedir. Veri hacmindeki bu artış, gerektiğinde verilerin otomatik olarak alınmasını gerçekleştirebilir. Veri madenciliği tekniklerinin kullanımı ile bilgi elde etmek ve ilginç, faydalı kalıpları belirlemek mümkündür. Bu yolla elde edilen bilgiler, iş verimliliğini ve karar alma kalitesini artırmak için uygun sırayla kullanılabilir (Taranu, 2016). Son zamanlarda veri madenciliğinin sağlık hizmetlerindeki kullanımının giderek daha popüler hale geldiği görülmektedir. Çeşitli faktörler sağlık hizmetlerinde veri madenciliği uygulamalarının kullanımını motive etmiştir. Örneğin, sağlık sigortası sahtekarlığının ve istismarının varlığından dolayı, birçok sağlık sigortası kuruluşu suçluları bulmak için veri madenciliği araçlarını kullanmakta ve bu yaklaşımla birlikte kayıplarını azaltmayı hedeflemektedir. Son zamanlarda, sağlık sahtekarlığı ve kötüye kullanım tespitinde de başarılı veri madenciliği uygulamalarına ilişkin raporlar bulunmaktadır. Ayrıca veri madenciliği uygulamalarını kullanarak dolandırıcılık tespiti, ticari dünyada da yaygındır. Örneğin, hileli kredi kartı işlemlerinin tespitinde sıklıkla kullanılmaktadır (Koh ve Tan, 2011).

Sağlık kuruluşları tarafından üretilen veriler çok geniş ve karmaşıktır. Bu veriler, hastaneler, hastalar ve tedavi maliyeti gibi ayrıntıları içerir. Bu nedenle, bu karmaşık verilerden önemli bilgileri analiz etmek ve çıkarmak için güçlü bir araca ihtiyaç duyulmaktadır. Sağlık verilerinin analizi, hastane yönetiminin performansını artırarak sağlık hizmetlerini geliştirir. Veri Madenciliği teknolojilerinin sonucunda, sağlık kuruluşlarının etkili tedaviler sunabilmesini, benzer hastalık veya sağlık sorunlarına sahip hastaların gruplandırılması gibi fayda sağlamaktadır. Ayrıca, hastaların hastanede kalış sürelerini tahmin etmek, tıbbi teşhis ve etkili bilgi sistemi yönetimi için plan yapmak için de yararlı olabilir (Tomar ve Agarwal, 2013).

Sağlık hizmetlerinde veri madenciliği uygulamasının kullanımında büyük bir potansiyel vardır. Tedavilerin etkinliği, sağlık yönetimi, müşteri ilişkilerinin yöntemi, dolandırıcılık ve istismar, hastane sıralaması, hastane enfeksiyon kontrolü ve daha akıllı tedavi tekniklerinin belirlenmesi gibi uygulamalar örnek olarak verilebilir.

Tedavinin Etkinliği: Tıbbi tedavilerin etkinliğini değerlendirmek için veri madenciliği uygulamaları geliştirilebilir. Veri madenciliği, tedavinin nedenlerini, semptomlarını ve seyrini karşılaştırıp, hangi eylem yollarının etkili olduğunu gösteren bir analiz sunabilir. Örneğin, aynı hastalık veya durum için farklı ilaç rejimleri ile tedavi edilen hasta gruplarının sonuçları, hangi tedavilerin en iyi ve en uygun maliyetli olduğunu belirlemek için karşılaştırılabilir. Örneğin, bir Amerikan sağlık hizmeti şirketi olan United Healthcare, maliyetleri azaltmanın ve daha iyi ilaç sunmanın yollarını araştırmak için tedavi kaydı verilerini incelemiştir. Ayrıca, doktorlara tedavi uygulama kalıpları hakkında bilgi vermek, bunları diğer doktorların uygulamalarıyla ve endüstri standartlarıyla karşılaştırmak için klinik profiller geliştirmiştir. Benzer şekilde veri madenciliği, spesifik hastalıklar için başarılı standart tedavilerin belirlenmesine yardımcı olabilir. 1999 yılında Florida Hastanesi, tüm kampüs, klinik ve hasta kabullerinde standart bir bakım yolu geliştirmek amacıyla en iyi klinik uygulamalar girişimini başlatmıştır. Diğer yandan,

tedavilerle ilgili diğer veri madenciliği uygulamaları, tedavinin çeşitli yan etkilerini ilişkilendirmeyi, tanıya yardımcı olmak için ortak semptomları harmanlamayı, ana popülasyondan belirli ilaçlara farklı yanıt veren alt popülasyonları tedavi etmek için en etkili ilaç bileşiklerini ve proaktif adımları belirlemeyi içerir (Koh ve Tan, 2011).

Highmark kuruluşu, Medicaid ve Medicare geri ödemelerini almak için veri madenciliği teknolojilerini başarıyla kullanmıştır. Bunu başarmak için Highmark, hasta semptomlarını, sağlık geçmişini ve demografik bilgileri kullanarak bir karar ağacı modeli oluşturmuştur. Bu model, her hastanın sahip olacağı hastalık riskini tahmin etmek için kullanılmış ve 13 hastalık riskine göre hastaları sıralanmıştır. Hastaları hastalık riskine göre sıralamanın nedeni, yetersiz tanı konabilecek hastaları kolayca tanımlamaktır. Bu hastalar erken aşamada tespit edilirse, hizmet sağlayıcıları ve sigortacıların talepleri yeniden göndermeleri gerekmediğinden bakım maliyetleri azaltılabilir, çünkü Medicare ve Medicaid Hizmetleri Merkezleri'nden alınan geri ödemeler teşhise bağlıdır. Bu karar ağaçları, modellenen hastalık sayısının artması nedeniyle yıllık olarak revize edilmektedir. Modelin kullanımı, Highmark'ın sağlık geri ödemesinde milyonlarca dolar tasarruf sağlamıştır (Yoo vd., 2012).

Hastalık Teşhisi: Veri madenciliği araçlarının başarılı sonuçlar verdiği uygulamalardan biridir. Kalp hastalığı son on yılda dünyanın önde gelen ölüm nedenidir. Bazı araştırmacılar, sağlık uzmanlarına kalp hastalığı tanısında yardımcı olmak için istatistiksel ve veri madenciliği araçlarını kullanmaktadır. Örneğin bir çalışmada, kalp hastalığı tanısında tek veri madenciliği tekniğinin kullanımını gösteren kapsamlı bir araştırma yapılmış ve hastalığın teşhisi noktasında veri madenciliğinin az da olsa kullanıldığı tespit edilmiştir (Shouman vd., 2012). Aynı şekilde Cerrito (2020) tarafından hazırlanan kitapta veri madenciliği yöntemi ile sağlık çıktılarının nasıl iyileştirilebileceği üzerinde farklı çalışmalar yapılmıştır. Yapılan çalışmada veri madenciliği yöntemlerinin kanser, astım ve diyabet gibi birçok alanda faydalı olabileceğine vurgu yapılmıştır.

Hastane Enfeksiyon Kontrolü: Enfeksiyon kontrol verilerindeki bilinmeyen veya düzensiz paternleri keşfetmek için veri madenciliği teknikleri kullanılarak muayeneler için bir sistem inşa edilebilmektedir. İlişkilendirme kuralları, kamu gözetimi ve hastane kontrol verilerinden beklenmedik ve ilginç bilgiler üretmek için kullanılabilir (Tomar ve Agarwal, 2013).

Sağlık Yönetimi: Sağlık yönetimine yardımcı olmak amacıyla kronik hastalık durumlarını, yüksek riskli hastaları daha iyi tanımlamak ve izlemek, uygun müdahaleleri tasarlamak, hastane kabul ve taleplerinin sayısını azaltmak için veri madenciliği uygulamaları geliştirilebilir. Örneğin, daha iyi tanı ve tedavi protokolleri geliştirmek için Arkansas Veri Ağı, geri kabul ve kaynak kullanımını incelenmekte ve en iyi tedavi seçeneklerini belirlemek için verilerini mevcut bilimsel literatürle karşılaştırmakta, böylece tıbbi bakımı desteklemek için kanıt kullanmaktadır. Ayrıca, Grup Sağlık Kooperatifi, hangi grupların en fazla kaynak kullandığını belirlemek için hasta popülasyonlarını ve demografik özelliklerini kullanmakta ve tıbbi koşulların güçlendirilmesini hedeflemektedir. Bu yaklaşımla birlikte popülasyonun eğitilmesi, olumsuz koşulların önlenmesi veya yönetilmesi için yardımcı programlar geliştirmektedir. Grup Sağlık Kooperatifi, daha düşük maliyetlerle daha iyi sağlık hizmeti sunmak için çeşitli veri madenciliği çalışmalarına katılmıştır. Seton Tıp Merkezi'nde veri madenciliği ise hastanın kalış süresini azaltmak, klinik komplikasyonlardan kaçınmak, en iyi uygulamaları geliştirmek, hasta sonuçlarını iyileştirmek ve hekimlere bilgi sağlamak için kullanılır. Başka bir örnek olarak Blue Cross, daha iyi hastalık yönetimi yoluyla sonuçları iyileştirmek ve harcamaları azaltmak için veri madenciliği girişimleri uyguluyor. Örneğin, bilinmeyen astımlıları tanımlamak ve uygun müdahaleler geliştirmek için acil servis ve hastaneye yatış talepleri verileri, ilaç kayıtları ve doktor görüşmelerini kullanır. Veri madenciliği yüksek maliyetli hastaları tanımlamak ve anlamak için de kullanılabilir. Son zamanlarda, Sierra Sağlık Hizmetleri, tedavi kılavuzları, hastalık yönetimi grupları ve maliyet yönetimi dahil olmak üzere kalite iyileştirme alanlarını tanımlamak için veri madenciliğini kapsamlı bir şekilde kullanmıştır (Koh ve Tan, 2011).

Müşteri İlişkileri Yönetimi: Müşteri ilişkileri yönetimi ticari kuruluşlar (tipik olarak bankalar ve perakendeciler) ve müşterileri arasındaki etkileşimlerin yönetilmesinde temel bir yaklaşım olmakla birlikte, sağlık hizmetleri bağlamında da önemli bir boyutu oluşturmaktadır. Müşteri etkileşimleri çağrı merkezleri, doktor ofisleri, faturalandırma departmanları, yatan hastalar ve ayakta tedavi edilen hastalar aracılığıyla gerçekleştirilebilir. Ticari organizasyonlarda olduğu gibi sağlık sektöründeki bireylerin memnuniyet düzeylerini artırmak için tercihlerini, kullanım şekillerini, mevcut ve gelecekteki ihtiyaçlarını belirlemek için veri madenciliği uygulamaları geliştirilebilir. Bu uygulamalar ayrıca bir sağlık hizmeti müşterisinin satın alacağı diğer ürünleri, hastanın reçeteli tedaviye uyup uymadığını veya önleyici bakımın gelecekteki kullanımda önemli bir azalma sağlayıp sağlamadığını tahmin etmek için kullanılabilir. Örneğin Müşteri Potansiyel Yönetimi A.Ş., bireyin 25 ana tanı kategorisi, seçilen teşhisle ilgili gruplar veya belirli tıbbi hizmet alanları tarafından tanımlanan belirli sağlık hizmetlerini kullanma eğiliminin bir göstergesi olan bir Tüketici Sağlık Hizmetleri Kullanım Endeksi oluşturmuştur. Milyonlarca hastanın milyonlarca sağlık hizmeti işlemine dayanan bu endeks, belirli sağlık hizmetlerinden en fazla yararlanabilecek hastaları tanımlayabilir ve özel bakıma en çok ihtiyaç duyan hastaları bu ağlara erişmeleri için teşvik edebilir. Endeks, OSF Saint Joseph Tıp Merkezi tarafından stratejik zamanlarda en uygun hastalara doğru mesajları ve hizmetleri almak için kullanılmıştır. Sonuç, daha etkili ve verimli iletişimin yanı sıra gelir artışıdır (Koh ve Tan, 2011).

Dolandırıcılık ve İstismar: Sahteciliği ve kötüye kullanımı tespit etmeye çalışan veri madenciliği uygulamaları genellikle normlar oluşturur ve daha sonra doktorlar, laboratuvarlar, klinikler veya diğerleri tarafından alımlımadık veya anormal talep kalıplarını belirler. Diğer şeylerin yanı sıra, bu uygulamalar uygun olmayan reçeteleri veya tavsiyeleri, hileli sigorta ve tıbbi iddiaları gösterebilir. Örneğin, Utah Medicaid Dolandırıcılık Bürosu, olağandışı kalıpları tespit etmek ve dolandırıcılığı ortaya çıkarmak için milyonlarca reçete, operasyon ve tedavi kursu tarafından üretilen veri kümesini incelemiştir. Aynı şekilde sahtecilik ve kötüye kullanım tespiti sonucunda Relia Star Finansal Şirketi, yıllık tasarruflarda yüzde 20'lik bir artış bildirmiş, Wisconsin Doktor Servis Sigortası Şirketi önemli tasarruflar kaydetmiş ve Avustralya Sağlık Sigortası Komisyonu on milyonlarca dolarlık yıllık tasarruf kaydetmiştir. Sahteciliği ve kötüye kullanımı tespit etmek için veri madenciliğini kullanmanın bir başka başarılı örneği, bir yıldan az bir süre çalıştıktan sonra 1998 yılında 2.2 milyon dolar toplayan ve 1.400 şüpheliyi tespit eden Teksas Medicaid Sahtekarlık ve Kötüye Kullanım Tespit Sistemi'dir. Başarısının farkında olan Texas sistemi olağanüstü başarısı ve teknolojinin yenilikçi kullanımı sayesinde en iyi ödülleri kazandı (Koh ve Tan, 2011).

Ulusal Sağlık Dolandırıcılığıyla Mücadele Derneği'ne göre, ABD'deki ulusal sağlık sigortası dolandırıcılığının 2003 yılında 51 milyar dolar değerinde olduğu tahmin edilmektedir (bu, ülkenin yıllık sağlık harcamalarının % 3'üdür). Bununla birlikte, hükümet ve kolluk kuvvetleri ulusal sağlık sigortası dolandırıcılığını 170 milyar dolar olarak tahmin etmektedir (harcamaların yaklaşık %10'udur). Bu sorunu çözmek için, bir sağlık sigortası şirketi Highmark, potansiyel sahtekarlık vakalarını tanımlamak için çok sayıda iddialara, müşterilere ve sağlayıcı verilerine dayanan sınıflandırma modelleri oluşturdu. Ortaya çıkan sahtekarlık tespit sistemi, sahtekarlığı tespit edebilecek ve gerçekleşmeden önce durdurabilecek öngörücü modeller oluşturmak için gerçek zamanlı veya gerçek zamanlıya yakın analiz yapmayı amaçlamaktadır. Bu veri madenciliği çabasının bir sonucu olarak Highmark, sınıflandırma sistemi emek yoğun işlerden kaçınmak için otomatik hale getirildiğinden, sahtekarlıkla ilgili soruşturma ve karar vermenin daha hızlı gerçekleştirildiğini keşfetti. Dolandırıcılığı tespit etme yeteneği, daha sonra normal ve sahtekarlık iddiaları arasındaki sapmaları daha iyi modellemek için kullanılacak daha hızlı bir çalışma döngüsü geliştirilebilir. Bu sürekli güncellenen veri madenciliği döngüsü 2005 yılında 11,5 milyon dolar tasarruf sağladı ve araştırmacıları iş yükünü önemli ölçüde azalttı (Yoo vd., 2012).

Hastane Sıralaması: Hastanelerin sıralarını belirlemek ve hastanelerin çeşitli fonksiyonlarını analiz etmek için farklı veri madenciliği yaklaşımları kullanılmaktadır. Hastanelerin sıralaması, yüksek riskli hastalarla başa çıkabilme yeteneklerine göre yapılır. Yüksek kapasiteli hastane yüksek riskli hastayı öncelikli olarak ele alırken düşük kapasiteli hastane risk faktörünü dikkate almaz (Tomar ve Agarwal, 2013).

Daha Akıllı Tedavi Teknikleri: Veri madenciliğini kullanarak doktorlar ve hastalar farklı tedavi teknikleri arasında kolayca karşılaştırma yapabilir. Mevcut tedavilerin etkinliğini analiz edebilir, hangi tekniğin daha iyi ve uygun maliyetli olduğunu öğrenebilirler. Veri madenciliği ayrıca, belirli tedavinin yan etkilerini tanımlamalarına, tehlikeyi azaltmak için uygun karar vermelerine ve tedavi için akıllı yöntemler geliştirmelerine yardımcı olur (Tomar ve Agarwal, 2013).

Hastanelerdeki Departmanların Verimliliklerinin Değerlendirilmesi: Veri madenciliği ile birlikte, sağlık kuruluşlarında yer alan ve verimsiz oldukları düşünülen bölümlerin değerlendirilmesi yapılabilmektedir. Örneğin, ABD’deki sağlık politikalarının değişmesi ile birlikte acil sağlık hizmetleri kullanımında büyük bir yoğunluk oluşmuştur. Bu problem ile ilgili yapılan bir çalışmada acil sağlık hizmetlerinin verimliliği veri madenciliği yöntemleri ile incelenmiş ve sorunun kurum bazında çözülebilmesi için iki temel öneride bulunulmuştur. Bunlar; gelen hastaların en yakın olan kliniğe nakledilmesi ve hemşirelerin bu süreçte daha aktif bir şekilde kullanılması olmuştur (Nfodjo, 2020).

4. SAĞLIKTA VERİ MADENCİLİĞİNDE KARŞILAŞILAN ZORLUKLAR

Bilindiği üzere birçok sağlık verisi çeşitli sağlık kuruluşları tarafından üretilir ve saklanır. Ancak sağlıkla ilgili verilerle doğru kararların alınmasında ciddi engeller yaratabilecek çeşitli zorluklar vardır. Sağlık bakım verileriyle ilgili ilk zorluk, saklanan verilerin formatının her sağlık kuruluşuna göre farklı olmasıdır. Bugüne kadar depolanan veriler için standart bir format belirlenmemiştir. Örneğin, bir salgın hastalığının farklı coğrafi bölgelerde bir ülkeye yayıldığını varsayalım. Ülke sağlık bakanlığı, salgın durumun çözülebilmesi ve gerekli tüm adımları atmak için tüm sağlık kuruluşlarının analizlerini merkezi veri deposuyla paylaşmalarını şart koşturmaktadır. Veri formatlarının farklı olmasından dolayı merkezde toplan verilerin analizi normalden daha uzun sürebilir ve bu sebepten dolayı durumun kontrolden çıkmasına neden olabilir. Sağlık verileri, hastaların sağlık hizmetlerini iyileştirmede, anlamlı bilgileri elde etmede oldukça yararlıdır. Veri kalitesi çok önemlidir, çünkü hiçbir niteliği olmayan veriden anlamlı bilgiler çıkarılamaz. Dolayısıyla verilerin kalitesi de çok önemli bir zorluktur. Sağlık hizmetlerinde kaliteyi korumak için gerekli tüm adımlar atılmalıdır. Veri paylaşımı bir diğer önemli sorundur. Ne hastalar ne de sağlık kuruluşları özel verilerini paylaşmakla ilgilenmezler. Bu nedenle, salgın durumları daha da kötüleşebilir, büyük bir nüfus için daha iyi tedaviler sağlamayı planlamak mümkün olmayabilir ve sağlık sigortası şirketlerinde sahtekarlık ve kötüye kullanımın tespitinde zorluk olabilir. Diğer bir zorluk ise bir ülkedeki tüm sağlık kuruluşlarının verilerini paylaştıkları veri ambarını oluşturmak çok maliyetli ve zaman alıcı bir süreçtir (Ahmad vd., 2015).

5. YÖNTEM

Bu çalışma, veri madenciliği ile ilgili Türkiye’de sağlık alanında yapılan lisansüstü tezlerle yönelik bir literatür incelemesidir.

5.1. Amaç

Türkiye’de sağlık alanında veri madenciliği ile ilgili yapılan tezlerde hangi konuların çalışıldığını ve daha çok hangi yöntemlerin kullanıldığını belirlemek amaçlanmıştır. Ayrıca lisansüstü tezlerle ilişkin bazı tanımlayıcı özellikler de değerlendirilmiştir. Çalışma kapsamında ele alınan araştırma soruları şu şekildedir:

- Veri madenciliği ile ilgili yapılan tezlerde hangi yöntemler kullanılmıştır?
- Veri madenciliği ile ilgili yapılan tezlerde konuların dağılımı nasıl olmuştur?
- Tezlerin temel amaçları nelerdir?
- Tezlerin yıllara, yazım diline ve lisansüstü düzeylerine (yüksek lisansa veya doktora) göre dağılımı nasıldır?
- En çok hangi örneklem grubu üzerinde çalışılmıştır?
- Tezler hangi üniversitelerde ve anabilim dallarında yapılmıştır?

5.2. Tarama Stratejisi ve Dahil Edilme Kriterleri

Bu çalışmada literatürün incelenmesinde doküman analizi yöntemi kullanılmıştır. Bu yöntem ile ilgili araştırmalar belirli bir çerçevede incelenir, birbirleri ile ilişkilendirilir ve bütüncül bir resim elde edilmeye çalışılır (Şimşek, 2009).

Ocak 2008 – Aralık 2019 tarihleri arasında yapılan araştırmada, erişime açık olan lisansüstü tezler analiz edilmiştir. “Ulusal Tez Merkezi” veri tabanı üzerinden yapılan taramada “başlıklar” ve “özetler” dikkate alınmış ve farklı konularda yazılmış 1163 lisansüstü tez tespit edilmiştir. Tarama yapılırken bazı kriterler dikkate alınmıştır (Tablo 1). Tarama sonucunda 98 çalışmanın değerlendirilmesi gerektiğine karar verilmiştir.

Tablo 1. Tarama Kriterleri ve Sonuçları

Veri Tabanları	• Ulusal Tez Merkezi
Anahtar Kelimeler	• Veri Madenciliği • Kümeleme • Sınıflandırma • Birliklik Kuralı
Tarama Kriterleri	• Tezin erişime açık olması • Türkiye’yi konu edinmiş olması • Sağlık alanında yapılmış olması • Tezlerin araştırma türünde olması (çalışmaların derleme türünde olmaması)
Tarama Sonucunda Çalışmaya Dahil Edilen Lisansüstü Tez Sayısı	• 98

5.3. Veri Analizi

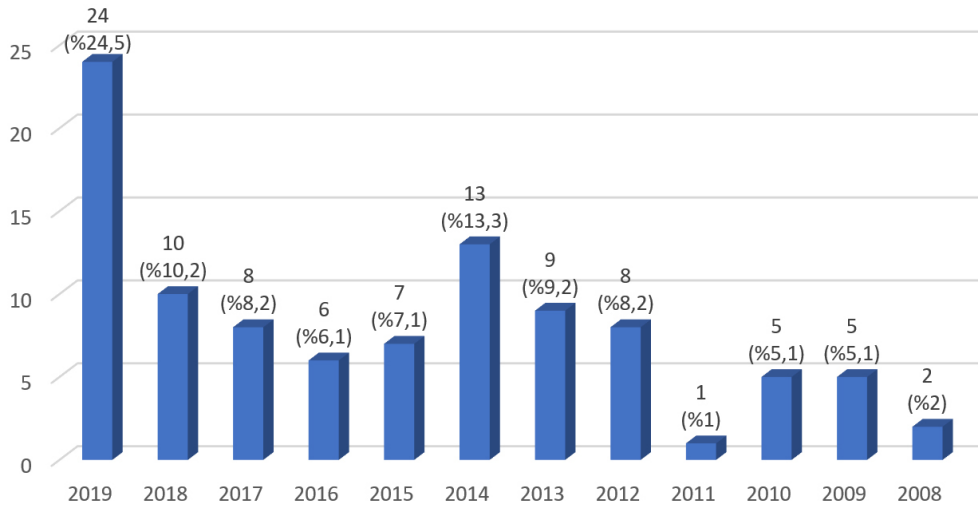
Doküman incelemesi sonucunda dikkate alınan lisansüstü tezlerden elde edilen veriler SPSS paket programı ile değerlendirilmiştir. Çalışma kapsamında tanımlayıcı istatistik yöntemler (frekans ve yüzde) kullanılmış ve sonuçlar tablo ve grafikler halinde sunulmuştur.

5.4. Sınırlılıklar

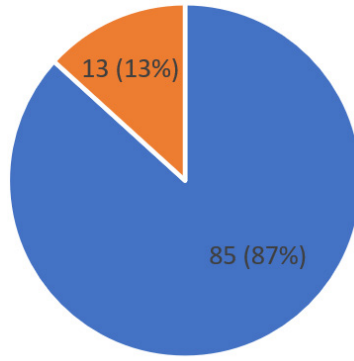
Çalışmaya sadece araştırma türünde yazılan lisansüstü tezlerin dahil edilmesi, sadece tam metnine ulaşımına izin verilen lisansüstü tezlerin dahil edilmesi ve sadece Türkçe ve İngilizce dillerinde yazılan tezlerin dahil edilmesi araştırmanın sınırlılıklarını oluşturmuştur.

6. BULGULAR

İncelenen tezlerin hangi yıllarda yapıldığı, türlerinin ne olduğu (yüksek lisans ya da doktora), hangi üniversitede ve anabilim dalında yapıldığı, yazım dillerinin ne olduğu, örneklemin, amacın, konunun ve yöntemlerin neler olduğu tablo ve grafikler ile değerlendirilmiştir.

*Grafik 1. İncelenen Lisansüstü Tezlerin Yıllara Göre Dağılımı*

Grafik 1’de incelenen lisansüstü tezlerin yıllara göre dağılımı gösterilmiştir. Veri madenciliği ile ilgili sağlık alanında yapılan çalışmaların yarısından fazlasının (%56,1) son 5 yılda (2019-2015) yapıldığı tespit edilmiştir.



■ Yüksek Lisans ■ Doktora

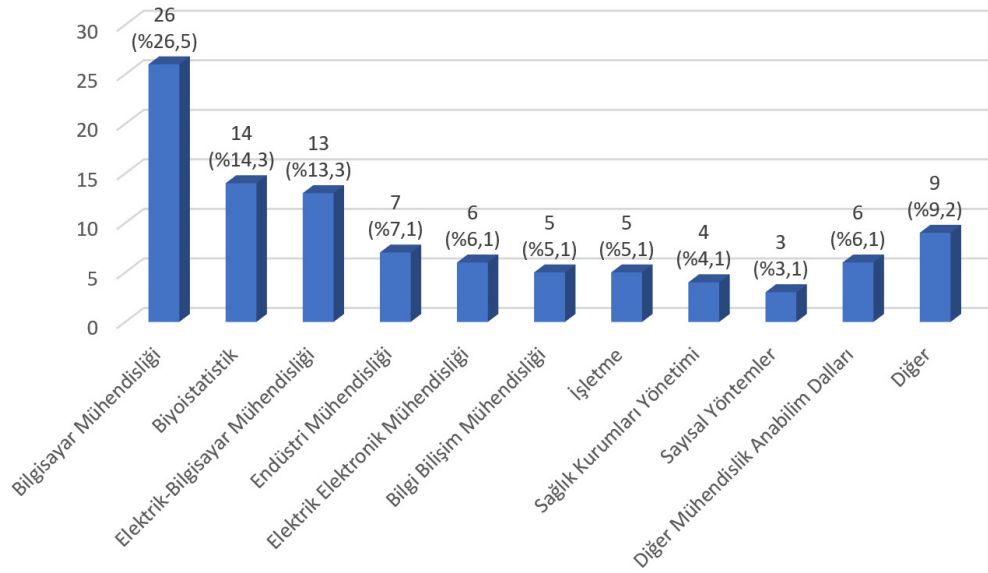
Grafik 2. İncelenen Lisansüstü Tezlerin Türlerine Göre Dağılımı

Çalışma kapsamında incelenen lisansüstü tezlerin türlerine göre dağılımı Grafik 2’de gösterilmiştir. Yüksek lisans düzeyinde yapılan tezlerin (%87) doktora düzeyinde yapılan tezlere (%13) göre daha fazla olduğu gözlemlenmiştir.

	Üniversiteler	n	%		Üniversiteler	n	%
1	Dokuz Eylül Üniversitesi	7	7.1	25	Düzce Üniversitesi	1	1
2	Marmara Üniversitesi	6	6.1	26	Karabük Üniversitesi	1	1
3	Altınbaş Üniversitesi	5	5.1	27	Gaziantep Üniversitesi	1	1
4	Bahçeşehir Üniversitesi	5	5.1	28	Çanakkale 18 Mart Üniversitesi	1	1
5	Gazi Üniversitesi	4	4.1	29	Kafkas Üniversitesi	1	1
6	Fırat Üniversitesi	4	4.1	30	Adnan Menderes Üniversitesi	1	1
7	Akdeniz Üniversitesi	4	3.1	31	Yaşar Üniversitesi	1	1
8	Çukurova Üniversitesi	3	3.1	32	Çankaya Üniversitesi	1	1
9	Ankara Üniversitesi	3	3.1	33	Van Yüzüncü Yıl Üniversitesi	1	1
10	Hacettepe Üniversitesi	3	3.1	34	Muğla Sıtkı Koçman Üniversitesi	1	1
11	Kırıkkale Üniversitesi	3	3.1	35	Sabancı Üniversitesi	1	1
12	Başkent Üniversitesi	3	3.1	36	Selçuk Üniversitesi	1	1
13	İstanbul Üniversitesi	3	3.1	37	Afyon Kocatepe Üniversitesi	1	1
14	Fatih Üniversitesi	2	2	38	Ağrı İbrahim Çeçen Üniversitesi	1	1
15	Atatürk Üniversitesi	2	2	39	Türk Hava Üniversitesi	1	1
16	Kahramanmaraş Sütçü İmam Üniversitesi	2	2	40	Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi	1	1
17	Orta Doğu Teknik Üniversitesi	2	2	41	Boğaziçi Üniversitesi	1	1
18	Erciyes Üniversitesi	2	2	42	Bülent Ecevit Üniversitesi	1	1
19	Cumhuriyet Üniversitesi	2	2	43	Trakya Üniversitesi	1	1
20	Sakarya Üniversitesi	2	2	44	Kadir Üniversitesi	1	1
21	Süleyman Demirel Üniversitesi	2	2	45	Pamukkale Üniversitesi	1	1
21	Karadeniz Teknik Üniversitesi	2	2	46	Ege Üniversitesi	1	1
23	Ankara Yıldırım Beyazıt Üniversitesi	2	2	47	Milli Savunma Üniversitesi	1	1
24	İstanbul Teknik Üniversitesi	2	2	TOPLAM		98	100

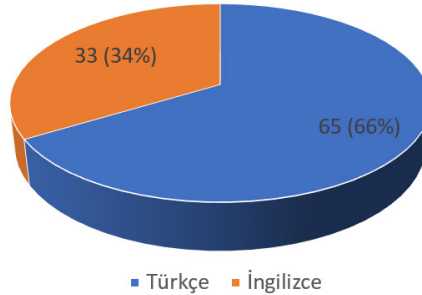
Tablo 2. İncelenen Lisansüstü Tezlerin Üniversitelere Göre Dağılımı

Tablo 2’de lisansüstü tezlerin üniversitelere göre dağılımı verilmiştir. Genel olarak bakıldığında 47 farklı üniversitede konu ile ilgili araştırma yapıldığı belirlenmiştir. İlk üç sırada Dokuz Eylül Üniversitesi (%7,1), Marmara Üniversitesi (%6,1) ve Altınbaş Üniversitesi’nin (%5,1) olduğu tespit edilmiştir. Üniversite yapılarına bakıldığında %81’ini kamu üniversitesi, %19’unun ise özel üniversitenin olduğu görülmüştür.



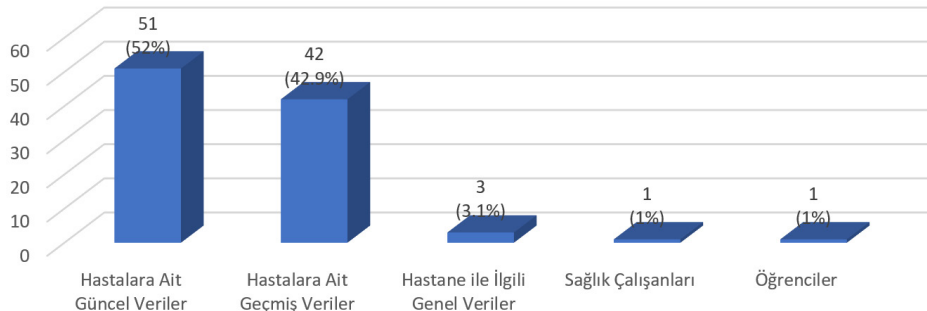
Grafik 3. İncelenen Lisansüstü Tezlerin Anabilim Dallarına Göre Dağılımı

Çalışma kapsamında incelenen tezlerin hangi anabilim dallarında yapıldığı Grafik 3’te gösterilmiştir. Konu ile ilgili en fazla bilgisayar mühendisliği (%26,5), biyoistatistik (%14,3) ve elektrik-bilgisayar mühendisliği (%13,3) anabilim dallarında çalışmalar yapıldığı tespit edilmiştir. “Diğer Mühendislik Anabilim Dalları” başlığı altında yer alan anabilim dalları; biyomedikal mühendislik (2), yazılım mühendisliği (2), biyomühendislik, nanoteknoloji ve mühendisliktir. “Diğer” başlığı altında yer alan anabilim dalları ise; medikal informatik, yönetim bilişim sistemleri, biyoteknoloji, tıp bilişimi, sağlık bilgileri, ekonometri, enformatik, teknoloji ve bilgi yönetimidir.



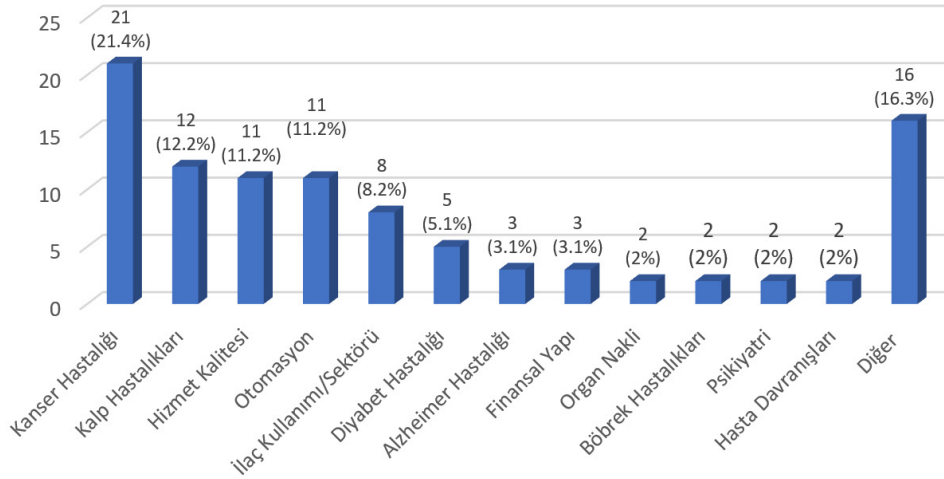
Grafik 4. İncelenen Lisansüstü Tezlerin Yazım Dillerine Göre Dağılımı

Grafik 4’te incelenen lisansüstü tezlerin yazım dillerine göre dağılımı verilmiştir. Türkçe dilinde (%66) yazılan tezlerin İngilizce dilinde (%34) yazılanlara göre daha fazla olduğu görülmüştür.



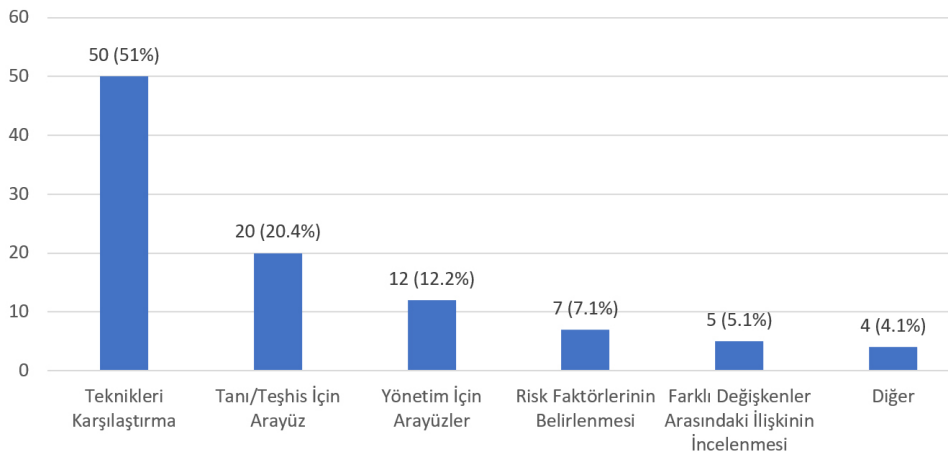
Grafik 5. İncelenen Lisansüstü Tezlerin Örneklemeye Göre Dağılımı

Grafik 5'e bakıldığında en fazla hastalara ait güncel veriler üzerinde çalışıldığı belirlenmiştir (%52). Aynı şekilde sağlık kuruluşlarından alınan hastalara ait geçmişe dönük verilerin araştırmalarda sıklıkla kullanıldığı tespit edilmiştir (%42,9).



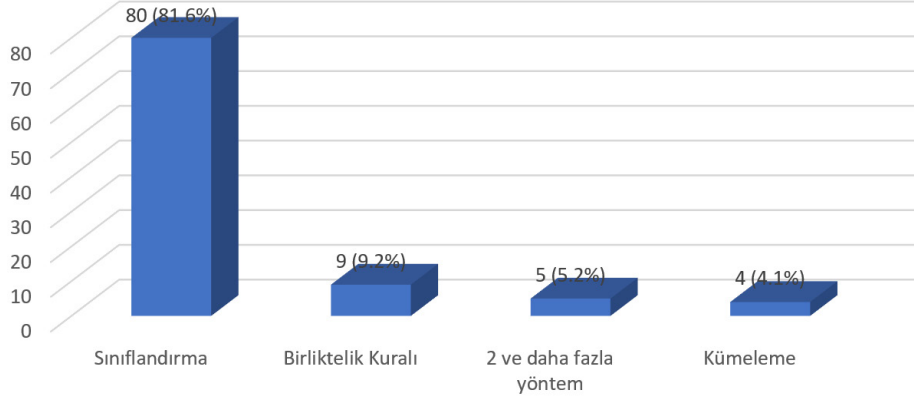
Grafik 6. İncelenen Lisansüstü Tezlerin Konularına Göre Dağılımı

İncelenen tezlerin konularına göre dağılımları Grafik 6'da gösterilmiştir. Bu verilere göre, en fazla kanser hastalığı (%21,4) ve kalp hastalıkları (%12,2) üzerinde araştırmalar yapılmıştır. Diğer başlığı altında ifade edilen bazı konular ise; sezaryen doğumları, sigorta şirketleri, ALS hastalığı, kadın hastalıkları ve doğum, akciğer hastalığı, üroloji, solunum hastalığı, hastane bazında ölüm tahminleri, karaciğer hastalığı, sigara kullanımı, acil servis yerleşiminin planlanması, mahremiyet ve bazı sağlık verileri dikkate alınarak veri madenciliği tekniklerinin karşılaştırılmasıdır.



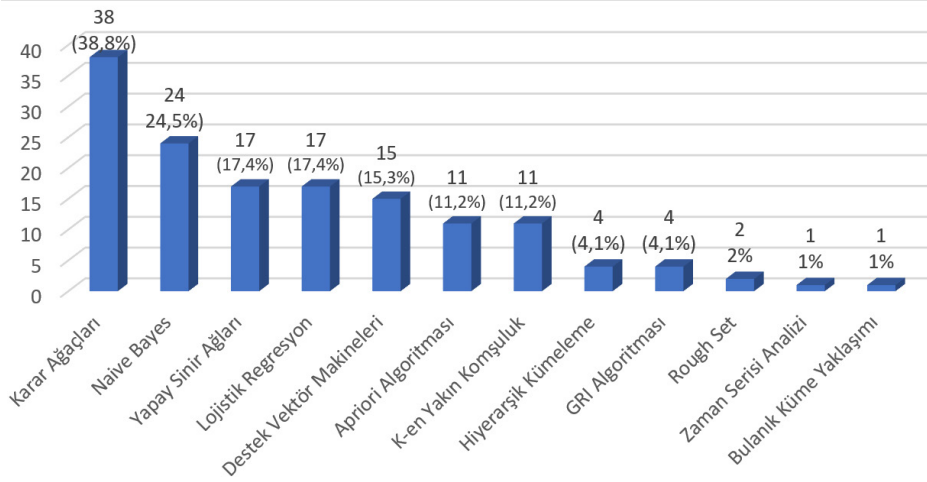
Grafik 7. İncelenen Lisansüstü Tezlerin Amaçlarına Göre Dağılımı

Grafik 7'de sağlık alanında yapılan lisansüstü tezlerin hangi amaçla yapıldığı verilmiştir. Çalışmaların %51'inde kullanılan teknikler karşılaştırılmış ve daha iyi sonuç verenler belirlenmiştir. Aynı şekilde %20,4'ünde hekimlerin karar vermelerine yardımcı olmak ve %12,2'sinde yönetimlerin karar vermelerine yardımcı olmak için çalışmalar yapıldığı tespit edilmiştir. Diğer başlığı altında ifade edilen amaçlar ise; var olan bir ilacın yeni kullanım alanlarının bulunması, hastalara ilişkin tıbbi laboratuvar verilerinin zamansal değişiminin izlenmesi, acil sağlık hizmetlerindeki teknolojiyi verimli kullanılması ve tip 2 diyabet hastaları için ilaç dozunun ayarlanmasıdır.



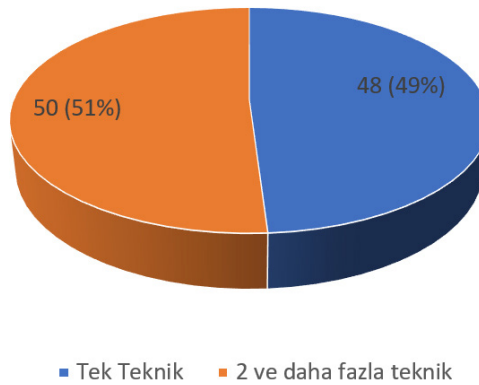
Grafik 8. İncelenen Lisansüstü Tezlerin Yöntemlere Göre Dağılımı

Çalışma kapsamında kullanılan teknikler dikkate alınarak veri madenciliği yöntemleri doğrultusunda bir sınıflandırma yapılmıştır (Grafik 8). Yapılan tezlerin büyük çoğunluğunu “sınıflandırma” yöntemleri çatısı altında yer aldığı tespit edilmiştir (%81,6). Aynı şekilde çalışmaların %9,2’sinde birlikte kuralı yöntemlerinin ve %4,1’inde ise kümeleme yöntemlerinin kullanıldığı gözlemlenmiştir.



Grafik 9. İncelenen Lisansüstü Tezlerin Tekniklere Göre Dağılımı

Sınıflandırma yöntemleri kapsamında dokuz farklı tekniğin çalışmalarda kullanıldığı belirlenmiştir (Grafik 9). Bunlar; yapay sinir ağları, karar ağaçları, lojistik regresyon, bayes sınıflandırması, rough set (kaba küme yaklaşımı), destek vektör makineleri, k-en yakın komşuluk, bulanık küme yaklaşımı ve zaman serisi yaklaşımıdır. *Birlikte kuralları* yöntemlerinde apriori algoritmasının ve GRI algoritmasının; *kümelemede* ise hiyerarşik kümelemenin kullanıldığı tespit edilmiştir.



Grafik 10. İncelenen Lisansüstü Tezlerin Teknik Sayılarına Göre Dağılımı

Grafik 10'da çalışmalardaki kullanılan teknik sayıları dikkate alınmıştır. Bu verilere göre, verilerin analiz edilmesinde “2 ve/veya daha fazla tekniğin” az da olsa “tek teknik” kullanımına göre daha fazla tercih edildiği tespit edilmiştir.

7. TARTIŞMA VE SONUÇ

Veri madenciliği uygulamaları ile birlikte sağlık alanında birçok önemli konuda araştırmalar yapılabilmektedir. Sağlığın daha iyi yönetilmesi, müşteri ilişkileri yönetiminin daha iyi hale getirilmesi, dolandırıcılık ve istismarın önlenmesi gibi konular örnek olarak verilebilir (Koh ve Tan, 2011). Aynı şekilde, uluslararası makaleleri dikkate alarak yapılan sistematik bir derleme çalışmasında veri madenciliği tekniklerinin klinik ve idari karar vermede daha çok kullanıldığına vurgu yapılmıştır (İslam vd., 2018). Türkiye’de yapılan lisansüstü tezler bakımından, yönetim ile ilgili karar vermede kullanılabilecek çalışmaların çok az olduğu tespit edilmiştir. Türkiye’de yapılan çalışmalarda daha çok klinik araştırmalara odaklanıldığı ve yönetim ile ilgili yapılan araştırmaların ise sayıca yetersiz olduğu görülmüştür. Aynı şekilde, sağlık kurumları yönetimi anabilim dallarında yapılan araştırmaların çok az olması da bu durumu kanıtlar niteliktedir (%4,1). Gelecekte yapılacak araştırmalarda maliyetlerin düşürülmesi noktasına odaklanılabilir. Literatüre bakıldığında, maliyetleri düşürmek ve daha rekabetçi hale gelmek amacı ile veri madenciliği yöntemlerinin kullanıldığı gözlemlenmiştir (Amarasingham vd., 2014; Bates vd., 2014; Haux vd., 2002).

Akıllı telefonlar, hasta merkezli sağlık hizmetleri için umut verici bir teknolojidir. Veri madenciliği tekniklerinin, hastaların alt grup ve bireysel düzeydeki ihtiyaçlarını karşılamak için mobil uygulamaları geliştirdiği iddia edilmektedir. Fallah ve diğerleri (2017) tarafından yapılan araştırmada takip ve izleme, tarama amaçlı erken teşhis ve tespit, sonuçların sınıflandırılması ve risk hesaplaması için mobil sağlık uygulamalarının veri madenciliği uygulamaları ile geliştirildiği belirtilmiştir. Yapılan çalışmada ise, hastane bazında yapılan çalışmalarda benzer uygulamaların çok olduğu gözlemlenmiş fakat mobil uygulamalar ile ilgili yapılan bir araştırmaya rastlanılmamıştır.

Diyabet hastaları ile ilgili yapılan bir sistematik derleme çalışmasında veri madenciliği tekniklerinin uygulanmasının, değerli bilgilerin elde edilmesinin, daha ileri bilimsel araştırma / deneyler için yeni hipotezlerin oluşturulmasının ve diyabet hastaları için sağlık hizmetlerinin iyileştirilmesinde kullanılmasının yararlı olduğuna vurgu yapılmıştır (Marinov vd., 2011). Türkiye’de sağlık alanında yapılan lisansüstü tezler bakımından ise, diyabet alanında yapılan çalışmaların çok az olduğu gözlemlenmiştir. Bu bağlamda daha fazla çalışma yapılması gerektiği söylenebilir. Benzer bir şekilde, ruh sağlığı alanındaki hastalıklarda da veri madenciliğinin büyük bir önem arz ettiği bilinmektedir. Alanso ve diğerleri (2018) tarafından yapılan bir çalışmada, 72 araştırma makalesinin dahil edildiği bir sistematik derleme yapılmış ve ruh sağlığı alanında risk faktörlerini belirlemek için veri madenciliği yöntemlerinin etkili bir şekilde kullanılabileceği sonucuna ulaşılmıştır. Yapılan çalışmada ise, ruh sağlığı özelinde yapılan çalışmaların yetersiz olduğu tespit edilmiştir.

Yöntemler açısından bakıldığında, en fazla kullanılan yöntemin sınıflandırma olduğu (%81,6) ve bu yöntemlerden en çok “karar ağaçları” (%38,78) ve “naive bayes” (%24,49) tekniklerinin kullanıldığı tespit edilmiştir. Diğer yandan, biriktirilmiş kuralları ve kümeleme yöntemlerinin çok az kullanıldığı gözlemlenmiştir. Farklı bulgular elde edebilmek ve yorumlama gücünü artırmak için sınıflandırma dışında kalan yöntemlerin kullanımının artırılması gerektiği söylenebilir.

Veri kullanımı noktasında, çalışmaların yaklaşık olarak %52’sinde hastalara ait güncel verilerin kullanıldığı gözlemlenmiştir. Diğer yandan, lisansüstü tezlerin yaklaşık olarak %43’ünde geçmişe yönelik verilerin kullanıldığı belirlenmiştir. Bu oranın güncel verilere göre düşük olmasında, hastaların geçmişe yönelik verilerinin kullanımı noktasında sağlık kuruluşlarının izin verme durumlarının sınırlı olması söz konusu olabilir. Proaktif çözümler üretebilmek için geçmişe dönük verilerin ve günümüz verilerinin birlikte dikkate alınması ve yorumlanması büyük önem arz etmektedir. Bu doğrultuda, geçmişe yönelik verilerin daha çok analize tabi tutulması ve güncel veriler ile karşılaştırılarak değerlendirmeler yapılması ulusal literatüre katkı sağlayabilir.

Yapılan lisansüstü tezlerin tanımlayıcı özelliklerine bakıldığında, yarıyıldan fazlasının son 5 yıl içerisinde yazıldığı (%56,1), büyük bir oranda yüksek lisans düzeyinde araştırmalar yapıldığı (%87), en fazla kamu üniversitelerinde çalışıldığı (%81), yazım dili olarak en fazla Türkçe’nin kullanıldığı (%66), konular açısından bakıldığında kanser hastaları üzerinde daha fazla araştırma yapıldığı (%21,4) ve çalışmalarda en çok tekniklerin karşılaştırıldığı (%51) gözlemlenmiştir.

Sonuç olarak, günümüzde hastaların müşteri olarak görülmeye başlanması, rekabet ortamının oluşması, kurumlar tarafından hizmet kalitesinin artırılması ve maliyetlerin düşürülmek istenmesi gibi birçok yaklaşımın önem kazanması ile birlikte veri madenciliği daha da değerli hale gelmiştir. Bu yöntemlerin, problemlerin çözümü noktasında önemli katkılar sağladığı yapılan çalışmalar ile kanıtlanmıştır. Türkiye’de yapılan lisansüstü tezlere bakıldığında ise, klinik araştırmalar ile ilgili araştırmaların yapıldığı fakat idari konularda yeterli çalışmaların yapılmadığı gözlemlenmiştir. Özellikle bu alanda yapılacak araştırmalar ile birlikte yönetimlerin ya da sağlık politikacılarının karar vermelerinde etkili karar destek sistemlerinin oluşturulabileceği söylenebilir. Aynı şekilde örneklem grubu olarak neredeyse araştırmaların tümünde hastaların dikkate alındığı görülmüştür. Bu sebeple, sağlık çalışanları ile ilgili yapılan araştırmaların artırılması gerekliliği de tespit edilmiştir.

KAYNAKLAR

- Ahmad, P., Qamar, S., & Rizvi, S. Q. A. (2015). Techniques of data mining in healthcare: a review. *International Journal of Computer Applications*, 120(15), 38-50.
- Alonso, S. G., de la Torre-Díez, I., Hamrioui, S., López-Coronado, M., Barreno, D. C., Nozaleda, L. M., & Franco, M. (2018). Data mining algorithms and techniques in mental health: A systematic review. *Journal of medical systems*, 42(9), 1-15.
- Amarasingham, R., Patzer, R. E., Huesch, M., Nguyen, N. Q., & Xie, B. (2014). Implementing electronic health care predictive analytics: Considerations and challenges. *Health Affairs*, 33, 1148–1154.
- Ataseven, B. (2013). Yapay sinir ağları ile öngörü modellemesi. *Öneri Dergisi*, 10(39), 101-115.
- Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Affairs*, 33, 1123–1131.
- Cerrito, P. (2020). *Cases on health outcomes and clinical data mining: studies and frameworks*. New York, Medical Information Science Reference.
- Coşlu, E. (2013). Veri madenciliği. In *Proceedings of Akademik Bilişim 2013 Conference* (pp. 573-585).
- Emre, İ. E., & Erol, Ç. S. (2017). Veri analizinde istatistik mi veri madenciliği mi?. *International Journal of Informatics Technologies*, 10(2), 161-167.
- Fallah, M., & Niakan Kalhori, S. R. (2017). Systematic review of data mining applications in patient-centered mobile-based information systems. *Healthcare Informatics Research*, 23(4), 262-270.
- Hand, D. J. (2007). Principles of data mining. *Drug Safety*, 30(7), 621-622.
- Haux, R., Ammenwerth, E., Herzog, W., & Knaup, P. (2002). Health care in the information society. A prognosis for the year 2013. *International Journal of Medical Informatics*, 66, 3–21.
- Islam, M. S., Hasan, M. M., Wang, X., & Germack, H. D. (2018). A systematic review on healthcare analytics: application and theoretical perspective of data mining. *In Healthcare* 6(2), 1-43.
- Jothi, N., & Husain, W. (2015). Data mining in healthcare—a review. *Procedia Computer Science*, 72, 306-313.
- Koh, H. C., & Tan, G. (2011). Data mining applications in healthcare. *Journal of Healthcare Information Management*, 19(2), 64-72.
- Kudyba, S. P. (2010). *Healthcare informatics: improving efficiency and productivity*. Florida, CRC Press.
- Marinov, M., Mosa, A. S. M., Yoo, I., & Boren, S. A. (2011). Data-mining technologies for diabetes: a systematic review. *Journal of Diabetes Science and Technology*, 5(6), 1549-1556.
- Nfodjo, D. (2020). *Utilization of the Emergency Department*. New York, Medical Information Science Reference.
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *İstanbul Ticaret Üniversitesi Dergisi*, 63-82.
- Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri madenciliği ve Türkiye’deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.

- Shouman, M., Turner, T., & Stocker, R. (2012). Using data mining techniques in heart disease diagnosis and treatment. In 2012 Japan-Egypt Conference on Electronics, Communications and Computers, pp. 173-177, Japan.
- Şimşek, H. (2009). Eğitim Tarihi Araştırmalarında Yöntem Sorunu. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 42(1), 33-51.
- Taranu, I. (2016). Data mining in healthcare: decision making and precision. *Database Systems Journal*, 6(4), 33-40.
- Tomar, D., & Agarwal, S. (2013). A survey on data mining approaches for healthcare. *International Journal of Bio-Science and Bio-Technology*, 5(5), 241-266.
- Yıldırım, P., Uludağ, M., & Görür, A. (2008). Hastane Bilgi Sistemlerinde Veri Madenciliği. *Çanakkale On Sekiz Mart Üniversitesi Akademik Bilişim*, 423-434.
- Yoo, I., Alafaireet, P., Marinov, M., Pena-Hernandez, K., Gopidi, R., Chang, J. F., & Hua, L. (2012). Data mining in healthcare and biomedicine: a survey of the literature. *Journal of medical systems*, 36(4), 2431-2448.