

TIBBİ VERİLER ÜZERİNDE BİRLİKTELİK KURALLARI MADENCİLİĞİ

Mehmet Ali ALAN¹

Özet

Bu çalışmada, hastane kayıtlarındaki veriler, veri madenciliği tekniği ile analiz edilerek başvuru hastane servisleri arasında birliktelik kurallarının olup olmadığı araştırılmaya çalışılmıştır. Bu amaçla 31.552 hastaya ait veriler işlenerek, Birliktelik Kuralları Madenciliği yapılmıştır. Yapılan çalışma sonucunda %75 güven seviyesinin üstünde 58 kural üretilebilmiştir. Üretilen bu kuralların hastane yöneticilerinin kararlarına katkı sağlayacağı değerlendirilmektedir.

Anahtar Kelimeler: Veri Madenciliği, Birliktelik Kuralları, Tıbbi Kural

Association Rules Mining on Medical Data

Abstract

In this study, it has been tried to investigate whether there are any association rules among hospital services applied to or not via analysing of data derived from hospital records with data mining technique. For this purpose, Association Rule Mining has been conducted using data belong to 31,552 patients. 58 rules which are above 75% confidence level have produced in consequence of the study. These rules are considered to contribute to decisions made by hospital administrators.

Keywords: Data Mining, Association Rules, Medical Rule

GİRİŞ

Verilerin en sağlıklı tutulduğu alanlardan birisi hastane otomasyon sistemleridir. Hastane otomasyon sistemleri, hastaneye gelen hastaların kişisel ve tıbbi geçmiş verilerini kayıt altına alıp, takip etme fırsatı sunmaktadır. Bu sistemler hem hastane çalışanlarına yönelik, hem de hastalara yönelik pek çok işin yürütülmesine imkân vermektedir.

Otomasyon sistemleri ile elde edilen kayıtlar, tıbbi veri tabanlarında muazzam miktarda verinin depolanmasını sağlamaktadır. Bu veri tabanlarının değerlendirilmesiyle, yeni bilgilerin keşfedilmesi, depolanan verilerin etkin kullanılması, hastane görevlerinin iyileştirilmesi açısından önem arz etmektedir.

Veri madenciliği yöntemlerinin temel amaçlarından biri, verilerde tutulan kalıpların net ve anlaşılır bir tanımını sağlamaktır. Ayrıca, tıbbi verilerin kaçınılmaz şekilde artması, veri madenciliği tekniklerinin kaliteyi iyileştirme ve sağlık hizmetlerinin maliyetini düşürme potansiyelini artıracaktır.

Veri Madenciliği, sağlık kuruluşundaki yeni eğilimleri ortaya çıkarmak için önemli bir rol oynamaktadır ve bu da tıbbi alanla ilişkili tüm taraflara yardımcı olmaktadır. Veri madenciliği teknikleri, bu sektörde hacimli veriler olduğundan, tıbbi veri madenciliğinde daha fazla işe yaramaktadır. Tıbbi verilerin hızla büyümesi nedeniyle, sağlık alanındaki karar destek ve tahmin sistemlerine yardımcı olmak için veri madenciliği tekniklerini kullanmak vazgeçilmez hale gelmiştir.

¹ Doç. Dr., Cumhuriyet Üniversitesi, Yönetim Bilişim Sistemleri Bölümü Öğretim Üyesi, alan@cumhuriyet.edu.tr

Birliktelik kuralları madenciliği, veri madenciliğinin en yaygın kullanılan yöntemlerinden biridir. Bu yöntemim hastane servislerinin konumlandırılmasında kullanılıp kullanılmayacağı sorusu bu çalışmanın temel sorunsalını oluşturmaktadır.

Bu çalışmada tıbbi veriler kullanılarak, hastaların başvurduğu servisler arasında birliktelik kurallarının olup olmadığı araştırılmaya çalışılmıştır. Böylelikle hastane yöneticilerine karar alma süreçlerinde destek sağlanmasına ve hasta memnuniyetinin artırılmasına çalışılmıştır.

I. VERİ MADENCİLİĞİ, BİRLİKTELİK KURALLARI VE APRIORI ALGORİTMASI

Veri madenciliği, veri tabanı sistemleri, istatistik, makine öğrenmesi, görselleştirme ve bilişim bilimini kapsayan, disiplinler arası bir alan ve disiplinler kümesinin bileşimidir (Han ve Kamber, 2006:29). Veri madenciliği, veri tabanlarındaki bilgileri keşfederek çalışılan alana yönelik karar destek sistemleri için gerekli ön bilgileri temin etmede kullanılmaktadır (Fayyad vd., 1996:38). Farklı bakış açılarından veri analiz etmeye dayalı bilgileri keşfetmenin mümkün olduğu veri madenciliği, reklam, bioformatik, veri tabanı pazarlaması, sahteciliğin tespiti, e-ticaret, sağlık, güvenlik, web, finansal tahmin vb. konu başlıkları da dâhil olmak üzere çok çeşitli alanlarda uygulama şansı bulabilmektedir (Jain vd., 2011:2793). Bunlara ilaveten veri madenciliği, önceden bilinmeyen örüntüleri keşfetmek için veri keşfetme bilimi ve teknolojisi olarak, veri tabanlarında bilgi keşfetmeye yönelik genel sürecin bir parçasıdır. Bilgisayar odaklı günümüz dünyasında, veri tabanlarındaki büyük çaptaki bilgiler, keşfedilecek örüntüleri de içerirler. Bu bilgilerin erişilebilirliği ve bolluğu, veri madenciliğini çok önemli ve gerekli hale getirmektedir (Rokach ve Maimon, 2008:1).

Veri madenciliğinde, Birliktelik Kuralları, Kümeleme, Karar Ağaçları, Diskriminant Analizi, Yapay Sinir Ağları, Genetik Algoritmalar vb. gibi çok sayıda teknik bulunmaktadır. Bu teknikler, bir yöneticinin kararlarını yönlendirebilecek bilgileri keşfetme ve bilgi sağlama amacıyla çeşitli alanlardan alınan bilgileri işleme amacıyla kullanılırlar (Wu ve Li, 2003).

Birliktelik Kuralları tekniği, bilgisayar bilimleri alanında geliştirilmiş, ancak daha çok market sepet analizi ve web tıklatma analizi gibi alanlarda uygulanmıştır. Genel olarak amaç, bir grup işlemde çoğunlukla birlikte meydana gelen parça gruplarının altını çizmektir (Giudici ve Figini, 2009: 90–91). Diğer bir ifade ile Birliktelik Kuralları, büyük veri kümelerindeki değişkenlerin özel değerleri arasında, ilişkinin ve birlikteliklerin tespit edilmesidir. Bu teknik, analizcilerin ve araştırmacıların büyük veri setlerindeki gizli örüntüleri açığa çıkarmalarını sağlar (Nisbet vd., 2009:126). Kullanışlılığı, kolay anlaşılması ve mümkün olan bütün örüntüleri ortaya çıkarması tekniğin güçlü yönüdür. Ancak mümkün olan bütün olasılıkları ortaya çıkarması aynı zamanda onun, bir zayıflığıdır. Çünkü karar vericiler, bütün bu olasılıklarını değerlendirecekleri büyük miktarda bilginin üstesinden gelmek durumunda kalırlar ki bu zor ve zaman alıcı bir durumdur (Kantardzic, 2003: 169).

Market Sepet Analizi olarak da adlandırılan Birliktelik Kuralları tekniği ile daha çok müşterilerin tüketim alışkanlıklarının analizine yönelik değerlendirmeler yapılmaktadır ve satın alma işlemlerinde birlikte oluşma eğilimi olan ürünlerin veya ürün gruplarının tanımlanmasına olanak sağlar (Giudici ve Figini, 2009: 175). Kısaca Birliktelik Kuralları ile müşterilerin tüketim alışkanlıkları ortaya konmakta ve müşterinin bir ürünü alırken, alması muhtemel diğer ürünleri de görmesi sağlanmaktadır.

Birliktelik Kuralları tekniğinde, Apriori, Fp-Growth, Tertius, Ais, Eclat gibi algoritmalar kullanılmaktadır. Bu çalışmada Apriori algoritması ile kurallar üretilmiştir.

Apriori algoritması, birliktelik kuralları madenciliğinde standart bir yaklaşım haline gelmiştir. İlk olarak Agrawal ve Srikant (1994) tarafından tanıtılmıştır. Algoritma, işlemleri içeren bir veri kümesiyle başlar ve en azından bir kullanıcı tarafından belirlenen eşişe sahip olan

sık ürün setleri oluşturmayı amaçlar. Apriori'nin algoritmik işleminde, k uzunluğundaki X kümesinin bir ögesi, k uzunluğuna sahip olan her X alt kümesinin de sık olması durumunda yaygın görülür. Bu değerlendirme, arama alanının önemli ölçüde azaltılmasına neden olur ve hesaplamalı olarak uygun bir zamanda kural keşfi sağlar (Nahar, 2013). Bu algoritma, doğrulama değerlerine göre kural bulur ve birinci dereceden mantıksal gösterimleri kullanır. Bu değerler, sınıf endeksi, sınıf doğrulama eşik değeri, doğrulama değeri, frekans eşik değeri, kayıp değerler ve veriler, gürültü eşik değeri, sayı kalıpları, tekrar kalıpları, ROC analizi ve çıktı değerleri gibi değişik seçenekleri içerir (Arora vd., 2013). Apriori, doğrulama değerlendirme fonksiyonunun en yüksek değerlerine sahip sonuçlarını arayan, bir tümevarımsal mantık programlama algoritmasıdır. Göreceli doğrulukla ağırlandırılmış en basit farklı doğrulama ölçülerini araştırır. Bir doğrulama ölçütü, bir kuralın umulmadık ve beklenen karşı örneklerin kısmını gösterir. Algoritmada beklenen ve gözlenen olasılığa sahip iki değer hesaplanır. Apriori, birinci dereceden kuralları ayıklar ve diğer programlarla Birliktelik Kuralları madencilik görevlerinde kullanılır (Nahar vd., 2013).

Apriori algoritmasının genel yapısı aşağıdaki gibidir:

Apriori Algoritması

```
 $F_1 = \{1 \text{ elemanlı sık görülen ürün kümeleri}\};$   
For (k= 1;  $F_k \neq \phi$ ; k++) do begin  
     $C_{k+1} = \text{apriori-gen}(F_k)$ ; // Yeni adaylar  
    for all işlemler t  $\in$  Database do begin  
         $C'_t = \text{altküme}(C_{k+1}, t)$ ; //t içindeki adaylar  
        For all adaylar c  $\in C'_t$  do  
            c.count++;  
        End  
         $F_{k+1} = \{c \in C_{k+1} \mid \text{c.count} \geq \text{minimum support}\}$   
    End  
End  
Answer  $U_k F_k$ ;
```

Birliktelik kurallarında başlangıçta kullanılan iki istatistik destek (*support*) değeri ve güven (*confidence*) seviyesidir. Bunlar sayısal değerlerdir ve bunları tanımlamak için bazı sayısal terimlerin tanımlanması gerekir. D işlemlerin veri tabanı, N de D deki işlemlerin sayısı olsun. Her D_i işlemi bir ürün kümesidir. $\text{Destek}(X)$ de X ürün kümesini içeren işlemlerin oranı olsun:

$$\text{Destek}(X) = \frac{|\{I \mid I \in D \wedge I \supseteq X\}|}{N}$$

I bir eleman kümesi ve $|\cdot|$ de kümenin eleman sayısını göstermektedir.

Bir birliktelik kuralının destek değeri, önceki ve sonrakinin her ikisinin de bulunduğu işlemlerin toplam işlem sayısına oranıdır. Güven değeri ise, öncekini içeren işlemlerin aynı

zamanda sonrakini de bulundurma oranıdır. $A \Rightarrow C$ birlikteliği için destek ve güven değerleri aşağıdaki gibidir (Webb, 2003):

$$\begin{aligned} Destek(A \Rightarrow C) &= Destek(A \cup C) \\ Güven(A \Rightarrow C) &= Destek(A \cup C) / Destek(A) \end{aligned}$$

Eğer destek değeri yeterince yüksekse (ve işlemler, gelecekteki işlemler ile aynı veri dağıtımından rasgele bir örneği temsil ediyorsa), o zaman güven seviyesi, kuralın birinci tarafını içeren herhangi bir gelecekteki işlemin, aynı zamanda, kuralın ikinci tarafını da içereceği ihtimalinin makul bir tahminidir (Webb, 2003:27-28).

Birliktelik kurallarında kullanılan bir değer hesaplama ise “lift” değeridir. Lift değeri şu şekilde hesaplanır:

$$Lift(A \Rightarrow C) = \frac{güven(A \Rightarrow C)}{destek(C)}$$

İşlem sonucunun 1’den küçük olması A’nın görülmesinin C’nin görülmesi üzerinde negatif korelasyona sahip olduğunu; 1’den büyük olması A’nın görülmesinin C’nin görülmesi üzerinde pozitif korelasyona sahip olduğunu ifade eder ki bunun anlamı, birinin görülmesi ile diğ erinin görülmesi ilişkilidir. Eğer lift değeri 1 çıkmış ise bu iki tarafın birbirinden bağımsız olduğu anlamına gelir (Taş, 2018:37-38).

II. LİTERATÜR ÖZETİ

Konuyla ilgili literatürde farklı tıbbi veri setleri üzerine yapılmış çalışmalar bulunmaktadır. Bunlardan Lakshmi ve Vadivu (2017), elektronik tıbbi sağlık kayıtlarını kullanarak, hastalıklar arasında, hastalıklar ve belirtiler arasında, hastalıklar ve ilaçlar arasındaki birliktelik kurallarını keşfetmişlerdir.

Ilayaraja ve Meyyappan (2013) farklı zaman dilimlerinde ve farklı bölgelerde yaşayan insanlarda tekrarlanan hastalıkla ilgili 29 hastalığa ait verileri girdi olarak alarak, hastalıkların sıklığı ve hastalıklar arasındaki birliktelik kurallarını Apriori algoritmasını kullanarak ortaya koymuşlardır.

Nahar vd. (2013), üç farklı birliktelik kuralı madenciliği algoritması (Apriori, Predictive Apriori ve Tertius) kullanılarak kalp hastalığı verileri üzerinde bir kural çıkarımı sunmuşlardır. Cinsiyete dayalı verileri sınıflandırarak, kural madenciliği temelli analiz yapmışlar ve hem erkek hem de kadınlar için kalp hastalığı için önemli risk faktörleri bulmuşlardır.

Krishnaiah vd. (2013), klinik tahminlerin idari, klinik, araştırma ve eğitimsel yönlerinde veri madenciliği tekniklerini uygulayarak, veri madenciliği tekniklerinin kaliteyi iyileştirme ve sağlık hizmetlerinin maliyetini düşürme potansiyeline vurgu yapmışlar, klinik tahminlerin çeşitli yönlerini iyileştirebilecek tıbbi veri madenciliği tekniklerini tartışmışlardır.

Tsumoto vd.(2011), bir hastane bilgi sisteminde depolanan verilere, veri madenciliği uygulamasını sunmuşlardır. Sonuçlar göstermiştir ki depolanan verilerin tekrar kullanılması hastane hizmetlerinin kalitesini artırmak için güçlü bir araç sunmakta ve hastane hizmetlerinin iyileştirilmesine yol açmaktadır.

Patil vd. (2010), 21 yaşın altındaki tip-2 diyabetik hasta hamile kadınların verilerini kullanarak, veriler arasında birliktelik kuralları çıkarmışlar, keşfedilen kuralların doktorların verileri daha iyi anlamalarına yardımcı olacağına vurgu yapmışlardır.

Isken ve Rajagopalan (2002), hastanelerde hasta akışının analitik modellerini oluşturmak amacıyla, hasta tipi tanımlarının geliştirilmesine rehberlik edecek, veri madenciliği tekniklerinden, K-means gibi kümeleme tekniklerin kullanma potansiyelini göstermişlerdir.

Stilou vd.(2001), Apriori algoritmasını diyabetik hastaların kayıtlarını içeren bir veritabanına uygulamışlar ve saklanan gerçek parametrelerden birliktelik kurallarını çıkarmaya çalışmışlar, izlenen metodolojinin, özellikle büyük veri hacimleri söz konusu olduğunda, teşhis prosedürü için iyi bir değer olabileceğini göstermişlerdir.

Brossette vd. (1998), bir, üç ve altı aylık zaman dilimlerinde, enfeksiyon kontrol verilerini kullanarak, birliktelik kuralları yöntemi ile analiz yaparak birinci veriden 2.000'den fazla kural, ikinci veriden 12.000'den fazla kural ve üçüncü veriden ise 20.000'den fazla kural üretmişlerdir.

III. VERİ SETİ VE YÖNTEM

Çalışmada bir kamu hastanesinden elde veriler, veri seti olarak kullanılmıştır. Excel kayıt formatında temin edilen 31.552 adet kayıttan, aynı hastaneye, en az iki farklı servise başvurmuş olan 4.350 hastanın verileri analize tabi tutulmuştur. Verilerde “Dahiye1”, “Dahiliye2” gibi farklı isimlerle yer alan ancak aynı servise ait olan poliklinik isimleri “Dahiliye” gibi tek isim altında toplanmıştır. Bu birleştirmelerden sonra birliktelik kuralına tabi tutulacak servis sayısı 20’ye inmiştir. Böylece 4.350 satır ve 20 sütundan oluşan veriler, Excel makroları kullanılarak gerekli dönüşümlere tabi tutulmuş ve veri ambarı oluşturulmuştur. Oluşturulan veri ambarı “Hastane.arff” adıyla kaydedilmiştir.

Veri ambarı, yönetim karar desteklerinde kullanılan konu odaklı, zaman varyantlı ve güncellenemez verilerin toplamıdır. Bir veri ambarı, meta, gerçek, boyutsal ve kümelenmiş bir veridir ve insanların bilgilendirilmiş kararları vermesini sağlayan, uygun bilgileri tedarik eden bir süreç yöneticisidir (Bose vd., 2009:190).

Bu çalışmadaki veri ambarının hazırlanması esnasında hastanın herhangi bir servise gelmiş olması durumunda ‘1’ değeri, başvurmadığı servisler için ise ‘0’ değeri atanmıştır. Çalışmada tüm servis adları değişken olarak tanımlanmış ve tüm değişken {1,0} değerlerini almıştır.

IV. UYGULAMA

Yapılan çalışmada Weka programının 3.7.2 sürümü kullanılmıştır. Weka programı, açık kaynak kodlu bir yazılımdır. Bu program pek çok sınıflandırma, kümeleme ve Birliktelik Kurallarına ait algoritmayı desteklemektedir.

Mevcut verilerin, Weka programınca da desteklenen Apriori algoritmasının uygulanması ile %75 ve üzeri güven seviyende 58 kural üretilebilmiştir. Tablo 1’de sunulan sonuçlara göre, Apriori algoritmasıyla üretilen kurallarda tablonun ikinci sütunu (Önceki) kuralın birinci koşulunu; üçüncü sütunu (Sonraki) ise ikinci koşulunu göstermektedir. Tablodaki dördüncü sütun güven (Confidence) seviyesini vermektedir. Tablonun beşinci sütununda ise lift değeri yer almaktadır. Aşağıda %75 ve daha üzeri güven seviyesi ile üretilen 58 kural sunulmuştur.

Tablo 1. Apriori Algoritmasıyla Üretilen Kurallar

No	Önceki	Sonraki	Güven	Lift
1	Cocuk_Acil_Servis=1 Kadin_Dogum=1 Kulak_Burun_Bogaz=1 14	Acil=1 14	1	2.8
2	Acil=1 Cocuk_Hastaliklari=1 Goz=1 12	Cocuk_Acil_Servis=1 12	1	3.45
3	Cocuk_Acil_Servis=1 Cocuk_Hastaliklari=1 Kadin_Dogum=1 Kulak_Burun_Bogaz=1 12	Acil=1 12	1	2.8
4	Acil=1 Cocuk_Hastaliklari=1 Kadin_Dogum=1 Psikiyatri=1 12	Cocuk_Acil_Servis=1 12	1	3.45

5	Acil=1 Çocuk_Hastalıkları=1 Psikiyatri=1 17	Cocuk_Acil_Servis=1 16	0.92	3.24
6	Acil=1 Çocuk_Cerrahisi=1 13	Cocuk_Acil_Servis=1 12	0.92	3.18
7	Acil=1 Cildiye=1 Kadın_Dogum=1 13	Cocuk_Acil_Servis=1 12	0.92	3.18
8	Acil=1 Çocuk_Hastalıkları=1 Kadın_Dogum=1 Kulak_Burun_Bogaz=1 13	Cocuk_Acil_Servis=1 12	0.92	3.18
9	Cildiye=1 Çocuk_Acil_Servis=1 Cocuk_Hastalıkları=1 12	Acil=1 11	0.92	2.27
10	Cocuk_Acil_Servis=1 Cocuk_Hastalıkları=1 Noroloji=1 12	Acil=1 11	0.92	2.57
11	Cocuk_Acil_Servis=1 Kadın_Dogum=1 Noroloji=1 12	Acil=1 11	0.92	2.57
12	Cocuk_Acil_Servis=1 Kulak_Burun_Bogaz=1 Psikiyatri=1 12	Acil=1 11	0.92	2.57
13	Cocuk_Acil_Servis=1 Cocuk_Hastalıkları=1 Kulak_Burun_Bogaz=1 18	Acil=1 16	0.89	2.49
14	Cocuk_Acil_Servis=1 Kadın_Dogum=1 Psikiyatri=1 18	Acil=1 16	0.89	2.49
15	Cildiye=1 Çocuk_Acil_Servis=1 Kadın_Dogum=1 14	Acil=1 12	0.86	2.4
16	Acil=1 Çocuk_Hastalıkları=1 Genel_Cerrahi=1 14	Cocuk_Acil_Servis=1 12	0.86	2.96
17	Cocuk_Acil_Servis=1 Cocuk_Hastalıkları=1 Göz=1 14	Acil=1 12	0.86	2.4
18	Cocuk_Hastalıkları=1 Fizik_Tedavi=1 Kadın_Dogum=1 14	Acil=1 12	0.86	2.4
19	Cocuk_Acil_Servis=1 Kadın_Dogum=1 Kulak_Burun_Bogaz=1 14	Cocuk_Hastalıkları=1 12	0.86	4.71
20	Acil=1 Çocuk_Acil_Servis=1 Kadın_Dogum=1 Kulak_Burun_Bogaz=1 14	Cocuk_Hastalıkları=1 12	0.86	4.71
21	Cocuk_Acil_Servis=1 Kadın_Dogum=1 Kulak_Burun_Bogaz=1 14	Acil=1Cocuk_Hastalıkları=1 12	0.86	4.8
22	Cocuk_Acil_Servis=1 Cocuk_Hastalıkları=1 Kadın_Dogum=1 Psikiyatri=1 14	Acil=1 12	0.86	2.4
23	Göz=1 Ortopedi=1 13	Cocuk_Acil_Servis=1 11	0.85	2.92
24	Göz=1 Ortopedi=1 13	Kadın_Dogum=1 11	0.85	5.04
25	Acil=1 Kulak_Burun_Bogaz=1 Psikiyatri=1 13	Cocuk_Acil_Servis=1 11	0.85	2.92
26	Cocuk_Hastalıkları=1 Kadın_Dogum=1 Noroloji=1 13	Acil=1 11	0.85	2.37
27	Cocuk_Hastalıkları=1 Kadın_Dogum=1 Psikiyatri=1 17	Cocuk_Acil_Servis=1 14	0.82	2.84
28	Acil=1 Çocuk_Hastalıkları=1 Kulak_Burun_Bogaz=1 20	Cocuk_Acil_Servis=1 16	0.8	2.76
29	Acil=1 Çocuk_Acil_Servis=1 Kulak_Burun_Bogaz=1 20	Cocuk_Hastalıkları=1 16	0.8	4.4
30	Cocuk_Acil_Servis=1 Cocuk_Hastalıkları=1 Psikiyatri=1 20	Acil=1 16	0.8	2.24
31	Acil=1 Kadın_Dogum=1 Psikiyatri=1 20	Cocuk_Acil_Servis=1 16	0.8	2.76
32	Göz=1 Noroloji=1 15	Acil=1 12	0.8	2.24
33	Cocuk_Cerrahisi=1 Cocuk_Hastalıkları=1 15	Cocuk_Acil_Servis=1 12	0.8	2.76
34	Acil=1 Cildiye=1 Çocuk_Acil_Servis=1 15	Kadın_Dogum=1 12	0.8	4.77
35	Acil=1 Fizik_Tedavi=1 Kadın_Dogum=1 15	Cocuk_Hastalıkları=1 12	0.8	4.4

36	Acil=1 Çocuk_Hastalıkları=1 Fizik_Tedavi=1 15	Kadin_Dogum=1 12	0.8	4.77
37	Cocuk_Hastalıkları=1 Genel_Cerrahi=1 Kadin_Dogum=1 15	Cocuk_Acil_Servis=1 12	0.8	2.76
38	Cocuk_Acil_Servis=1 Genel_Cerrahi=1 Kadin_Dogum=1 15	Cocuk_Hastalıkları=1 12	0.8	4.4
39	Acil=1 Cildiye=1 Çocuk_Hastalıkları=1 14	Cocuk_Acil_Servis=1 11	0.79	2.71
40	Acil=1 Çocuk_Hastalıkları=1 Noroloji=1 14	Cocuk_Acil_Servis=1 11	0.79	2.71
41	Acil=1 Çocuk_Acil_Servis=1 Noroloji=1 14	Cocuk_Hastalıkları=1 11	0.79	4.32
42	Cocuk_Acil_Servis=1 Fizik_Tedavi=1 Kadin_Dogum=1 14	Acil=1 11	0.79	2.2
43	Acil=1 Çocuk_Acil_Servis=1 Fizik_Tedavi=1 14	Kadin_Dogum=1 11	0.79	4.68
44	Acil=1 Çocuk_Acil_Servis=1 Noroloji=1 14	Kadin_Dogum=1 11	0.79	4.68
45	Acil=1 Çocuk_Hastalıkları=1 Noroloji=1 14	Kadin_Dogum=1 11	0.79	4.68
46	Cocuk_Acil_Servis=1 Noroloji=1 18	Acil=1 14	0.78	2.18
47	Cocuk_Acil_Servis=1 Kadin_Dogum=1 Psikiyatri=1 18	Cocuk_Hastalıkları=1 14	0.78	4.27
48	Cocuk_Hastalıkları=1 Kadin_Dogum=1 Kulak_Burun_Bogaz=1 17	Acil=1 13	0.76	2.14
49	Fizik_Tedavi=1 Kulak_Burun_Bogaz=1 16	Acil=1 12	0.75	2.1
50	Cocuk_Acil_Servis=1 Çocuk_Hastalıkları=1 Genel_Cerrahi=1 16	Acil=1 12	0.75	2.1
51	Acil=1 Çocuk_Acil_Servis=1 Genel_Cerrahi=1 16	Cocuk_Hastalıkları=1 12	0.75	4.12
52	Acil=1 Çocuk_Acil_Servis=1 Goz=1 16	Cocuk_Hastalıkları=1 12	0.75	4.12
53	Acil=1 Goz=1 Kadin_Dogum=1 16	Cocuk_Acil_Servis=1 12	0.75	2.59
54	Acil=1 Çocuk_Acil_Servis=1 Goz=1 16	Kadin_Dogum=1 12	0.75	4.47
55	Cocuk_Acil_Servis=1 Çocuk_Hastalıkları=1 Genel_Cerrahi=1 16	Kadin_Dogum=1 12	0.75	4.47
56	Acil=1 Çocuk_Acil_Servis=1 Çocuk_Hastalıkları=1 Kulak_Burun_Bogaz=1 16	Kadin_Dogum=1 12	0.75	4.47
57	Acil=1 Çocuk_Acil_Servis=1 Kadin_Dogum=1 Psikiyatri=1 16	Cocuk_Hastalıkları=1 12	0.75	4.12
58	Acil=1 Çocuk_Acil_Servis=1 Çocuk_Hastalıkları=1 Psikiyatri=1 16	Kadin_Dogum=1 12	0.75	4.47

Yukarıdaki tablodan da anlaşılacağı gibi üretilen ilk kurala göre, “Çocuk Acil Servis”, “Kadın Doğum” ve “Kulak Burun Boğaz” servislerine başvuran 14 hasta aynı zamanda “Acil” servise de başvurmuştur.

İkinci kurala göre “Acil” servise, “Çocuk Hastalıklarına” ve “Göz” servislerine başvuran 12 kişi aynı zamanda “Çocuk Acil” servisine de başvurmuştur.

Üçüncü kurala göre “Çocuk Acil Servisi”, “Çocuk Hastalıkları”, “Kadın Doğum” ve “Kulak Burun Boğaz” servislerine başvuran 12 kişi, aynı zamanda “Acil” servise de başvurmuştur.

Dördüncü kurala göre “Acil Servis”, “Çocuk hastalıkları”, “Kadın Doğum” ve “Psikiyatri” servislerine başvuran 12 kişi aynı zamanda “Çocuk Acil” servisine de başvurmuştur.

Yukarıdaki dört kuralın da lift değerleri farklı olsa da güven seviyesi 1’dir.

Beşinci kurala göre ise “Acil Servis”, “Çocuk Hastalıkları” ve “Psikiyatri” servislerine başvuran 17 kişiden 16’sı “Çocuk Acil” servisine başvurmuştur. Bu kuralın güven seviye ise 0.98’dir.

Geri kalan 53 kural için de yukarıdaki yorumlara benzer yorumlar yapılabilir.

SONUÇ VE ÖNERİ

Bütün dünyada sağlık maliyetlerinin artması, kıt kaynakların kullanımını optimize etmenin zorluğunu sürekli olarak ele almaya teşvik etmektedir. Kaynakları optimize etme yolunda atılan ilk adımlardan biri hastanelerde hasta akışını sağlıklı şekilde yönlendirmektir. Bu konuda veri madenciliği teknikleri hastane yöneticilerine rehberlik edebilir.

Verilerin en sağlıklı tutulduğu kaynaklardan biri, hasta kayıtlarının tutulduğu hastane veri tabanlarıdır. Bu veri tabanlarındaki verilerin değerlendirilmesi, bu verilerden üretilecek bilgi, hem hastane yönetiminin hem de hastaların işine yarayabilir. Hastane yöneticilerinin aldığı önemli kararlardan biri de hasta memnuniyetinin artırılmasıdır.

Bu çalışma ile hastane yöneticilerine, hastane servislerinin konumlandırılması aşamasındaki karar süreçlerine katkı sağlama amacı taşımaktadır. Çalışmanın amacını gerçekleştirebilmek için veri madenciliğinin Birliktelik Kuralları yönteminden yararlanılmıştır.

Birliktelik Kuralları madenciliği ile hastane veri tabanından alınan veriler yardımıyla, servisler arasındaki birliktelikler incelenmiş ve bu servisler arasındaki kurallar ve örüntüler tespit edilmeye çalışılmıştır. Böylelikle hastane servisleri arasındaki ilişkilerin ortaya konması ve hasta memnuniyetinin optimum bir seviyeye çekilmesi noktasında katkı sağlanmaya çalışılmıştır.

Çalışmada bir kamu hastanesine ait veriler, Birliktelik Kuralları madenciliği tekniği ile analiz edilmiştir. Bu bağlamda 31.552 hastaya ait veri, Weka Programının 3.7.2 sürümü kullanılarak Apriori algoritması ile analize tabi tutulmuştur.

Analiz sonucunda hastane servislerinin tercih edilmesinde, %75 ve üzeri güven seviyesinde 58 birliktelik kuralı tespit edilmiştir. Üretilen ilk kurala göre, “Çocuk Acil Servis”, “Kadın Doğum” ve “Kulak Burun Boğaz” servislerine başvuran 14 hasta aynı zamanda “Acil” servise de başvurduğu, 1 güven seviyesi ve 2.8 lift değeri ile anlaşılmıştır. Bu kurala benzer 57 kural daha tespit edilmiştir.

Bu sonuçlara göre hastane yönetimine, üretilen kurallara göre servisleri konumlandırma önerileri sunulmaktadır.

KAYNAKÇA

- Agrawal, R., & Srikant, R. (1994), *Fast algorithms for mining association rules*. In Proceedings of the 20th international conference on very large data bases, Santiago, Chile. Citeseer, :487–499.
- Arora, Jyoti; Bhalla, Nidhi and Rao, Sanjeev (2013), ‘A Review On Association Rule Mining Algorithms’ *International Journal of Innovative Research in Computer and Communication Engineering* 1.5: 1246-1251.

- Bose, I., Chun, L. A., Yue, L. V. W., Ines, L. H. W. and Helen, W. O. L. (2009), 'Business Data Warehouse: The Case of Wal-Mart', *Data Mining Applications for Empowering Knowledge Societies*, Ed. Hakikur Rahman, Information Science Reference, :189-198
- Brossette, Stephen E., Sprague, Alan P., Hardin, J. Michael, Waites, Ken B., Jones, Warren T., Moser, Stephen A., (1998), *Data Mining In Hospital Infection Control And Public Health Surveillance*, Journal Of The American Medical Informatics Association Volume 5 Number 4 Jul / Aug 1998, :373-381
- Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996). 'From data mining to knowledge discovery in databases' *AI magazine* 17.3:37-54.
- Giudici, Paolo and Figini, Silvia (2009), *Applied Data Mining For Business and Industry*, Second Edition, Wiley Publication, West Sussex.
- Han, Jiawei, and Micheline Kamber (2006). *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan Kaufmann.
- Ilayaraja, M. & Meyyappan, Thiru (2013), Mining Medical Data To Identify Frequent Diseases Using Apriori Algorithm, Proceedings Of The 2013 International Conference On Pattern Recognition, Informatics And Mobile Engineering, February 21-22, :194-199
- Isken, Mark W. and Rajagopalan, Balaji (2002), *Data Mining to Support Simulation Modeling of Patient Flow in Hospitals*, Journal of Medical Systems, Vol. 26, No. 2, April 2002, :179-197
- Jain, Yogendra Kumar, Vinod Kumar Yadav, and Geetika S. Panday. (2011) 'An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining' *International Journal on Computer Science and Engineering* 3.7 : 2792-2798.
- Kantardzic, Mehmed (2003) *Data Mining: Concepts, Models, Methods, and Algorithms*, John Wiley & Sons J. B. Speed Scientific School, University of Louisville IEEE Computer Society.
- Krishnaiah, V., Narsimha, G. & Chandra, N. S. (2013), *A Study On Clinical Prediction Using Data Mining Techniques*, International Journal of Computer Science Engineering and Information Technology Research (IJCEITR) ISSN 2249-6831 Vol. 3, Issue 1:239-248,
- Lakshmi K.S. , Vadivu, G. (2017), *Extracting Association Rules from Medical Health Records Using Multi-Criteria Decision Analysis*, 7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Procedia Computer Science 115 (2017) pp. 290–295, Elsevier, Cochin, India
- Nahar J., Imam T., Tickle KS., Chen YP (2013), *Association Rule Mining To Detect Factors Which Contribute To Heart Disease in Males And Females*, Expert Systems with Applications 40 (2013) 1086–1093
- Nisbet, Robert, John Elder IV, and Gary Miner, (2009), *Handbook of statistical analysis and data mining applications*. Elsevier Inc, Burlington.
- Patil, B. M., Joshi, R. C. & Toshniwal, D. (2010), *Association rule for classification of type -2 diabetic patients*, 2010 Second International Conference on Machine

Learning and Computing, 978-0-7695-3977-5/10 © 2010 IEEE, DOI
10.1109/ICMLC.2010.67, IEEE Computer Society

- Rokach, Lior and Maimon, Oded, (2008), *Data Mining with Decision Trees: Theory and Applications*, World Scientific New Jersey.
- Stiloul, S.& Bamidis, P.D.& Maglaveras, N.& Pappas, C.(2001), Mining Association Rules from Clinical Databases: An Intelligent Diagnostic Process in Healthcare, MEDINFO 2001, V. Patel et al. (Eds), Amsterdam: IOS Press, :1399-1403
- Taş, Yusuf (2018), *Birliktelik Kuralları Madenciliği Ve Bir Uygulama*, Cumhuriyet Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Sivas
- Tsumoto, Shusaku, Hirano, Shoji, Tsumoto, Yuko (2011), *Information Reuse in Hospital Information Systems: A Data Mining Approach*, IEEE IRI 2011, August 3-5, 2011, Las Vegas, Nevada, :172-176
- Webb, G.,I. (2003). Association Rules. In Nong Ye (Edt.), *The Handbook Of data Mining* (pp. 27-28). New Jersey: Lawrence Erlbaum Associates, Inc.
- Wu, Tong and Li, Xiangyang (2003), 'Data Storage and Management', *The Handbook of Data Mining*, Ed. Nong Ye, New Jersey: Lawrence Erlbaum Associates, Inc.: 393-407.