



VERİ MADENCİLİĞİ UYGULAMASINA İLİŞKİN PAÜ HASTANESİNDE HASTA PROFİLİNİN BELİRLENMESİ*

THE DETERMINATION OF PATIENT PROFILE AT PAMUKKALE UNIVERSITY AS RELATER TO THE APPLICATION OF DATA MINING*

İrfan ERTUĞRUL^{1*}, Arzu ORGAN¹, Ayşegül ŞAVLI²

¹İşletme Bölümü, İktisadi ve İdari Bilimler Fakültesi, Pamukkale Üniversitesi, 20070, Denizli.

iertugrul@pau.edu.tr, aorgan@pau.edu.tr

²Sağlık Araştırma ve Uygulama Merkezi, Pamukkale Üniversitesi, 20070, Denizli.

asavli@pau.edu.tr

Geliş Tarihi/Received: 02.07.2012, Kabul Tarihi/Accepted: 29.08.2012

doi: 10.5505/pajes.2013.68077

*Yazışılan yazar/Corresponding author

Özet

Veri madenciliği, ham verilerden çok net olmayan, önceden bilinmeyen fakat potansiyel olarak kullanışlı bilginin elde edilmesidir. Bu da; kümeleme, veri özetleme, değişikliklerin incelenmesi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir. Veri madenciliği veri yığınlarından, hedefe yönelik sorulara cevap verebilecek gereksiz verilerden temizlenmiş çok değerli bilgilere ulaşılabilir yöntemidir. Veri madenciliği ile ilgili genel bilgiler verilerle çalışmaya başlanmıştır. Veri madenciliği ile pek çok alanda karar verme işlemlerine yardımcı olacak önemli sonuçlar elde edilebilir. Çeşitli yöntemlerle veriler incelenerek, ham verilerden anlamlı verilere ulaşılabilir. Veri madenciliğinin yaygın bir uygulama alanı vardır. Çalışmamız, bu uygulama alanlarından biri olan sağlık sektöründeki verileri kapsamaktadır. PAÜ Hastane Bilgi Yönetim Sisteminde yer alan veriler kullanılarak gelen hastaların profilini belirleyen bir uygulama çalışması yapılmıştır. Uygulama ile verilen bilgilerin sağlık profesyonellerinin en doğru ve güncel bilgiye ulaşmasını, en objektif ve optimum çözümleri kullanmasını sağlayacak bir karar destek aracı olduğu gösterilmeye çalışılmıştır.

Anahtar kelimeler: Veri madenciliği, Veri, Tıpta veri madenciliği, Hasta profili, Hastane bilgi yönetim sistemi (HBYS).

Abstract

Data mining is a way to obtain the potentially useful information, but previously unknown and unclear from the raw data-And it includes a certain number of technical approaches such as the clustering, data summarization, evaluation of changes and the detection of deviations. Data mining is a way of obtaining much valuable information from the data stack, which can reply the questions about the target and which was cleaned from redundant data. It was started to work by giving general information related to data mining. Thanks to data mining, it can be obtained important results that will help decision-making processes in many areas. By examining the data in many ways, meaningful data can be accessed by raw data. There is a common application area of data mining. Our study includes the data in health sector which is one of these application areas. Here, there is an application study on determining the patient profile through using the data in Pamukkale University hospital information management system-through this given data, health care professionals can obtain the most correct and the newest information and this given data is at the same time a decision support tool to provide the most objective and optimal solutions for them

Keywords: Data, Data mining, Data mining in medicine, Patient profile, The hospital information management systems(HIMS).

1 Giriş

Veri madenciliği büyük veri tabanlarından gizli, tahmini bilgi çıkarımı amaçlayan yeni ve güçlü bir teknolojidir [1].

Verilerin dijital ortamda saklanmaya başlanması ile birlikte, yeryüzündeki bilgi miktarının her geçen gün katlanarak arttığı günümüzde, veri tabanlarının sayısı da benzer, hatta daha yüksek bir oranda artmaktadır. Yüksek kapasiteli işlem yapabilmeye gücünün ucuzlaşmasının bir sonucu olarak, veri saklama hem daha kolaylaşmış, hem de verinin kendisi ucuzlamıştır. Veri tabanlarında saklanan veri, bir dağa benzetilirse, bu veri dağı tek başına değersizdir ve kullanıcı için çok fazla bir anlam ifade etmez. Ancak bu veri dağı, belirli bir amaç doğrultusunda sistematik olarak işlenir ve analiz edilirse, değersiz görülen veri yığnında, amaca yönelik sorulara cevap verebilecek çok değerli bilgilere ulaşılabilir [2]. Veri tabanı sistemlerinin artan kullanımı ve hacimlerindeki olağanüstü artış karşısında toplanan çok büyük miktarda ham veriyi özümleme ve yorumlama alanındaki yetersizlik, organizasyonları bu verilerden nasıl faydalanılabileceği düşüncesi ile karşı karşıya bırakmıştır. Bu bağlamda 1990' lı yıllara kadar daha çok verinin toplanması ve depolanması ile

ilgilenilmiş. 1990'dan itibaren ise veri analizi ön plana çıkmıştır[3].

Sağlık sistemi politikalarının ve yönetsel kararlarının temeli veri ve veriden elde edilmiş bilgidir. Sağlık politika ve kararlarının amaçlara uygun ve etkin olabilmesi güvenilir, güncel ve doğru veriye bağlıdır [4]. Bilgi sistemleri ve iletişim teknolojilerindeki gelişmeler sayesinde tıp ve sağlık alanındaki birçok veri sayısal ortamda saklanabilmekte ve kolaylıkla erişilebilmektedir. Hastane bilgi sistemleri hastalara ait demografik bilgiler, hastalık ve tedavi durumları, yapılan tetkikler, makbuz, faturalama ve idari işlere ait bilgileri içerir. Sağlık ve tıp, çağımızın en önemli bilimsel araştırma alanları olduğu için bu alandaki bilgi sistemleri de araştırmalar için en büyük veri kaynaklarıdır[5]. Uygun ve etkili bir karara ulaşmak için, karar verici birçok veriyi analiz etmek ve birçok faktörü dikkate almak zorundadır [6].

Sağlık bilgi sistemlerinin amacı büyük miktardaki sağlık verilerinden faydalı bilgi üretmektir. Bu bilgiler hasta düzeyinde daha iyi sağlık hizmeti sunumu, sağlık kurumlarının daha iyi yönetilmesi, kaynakların etkin kullanımı ve sağlık politikalarının oluşturulması amaçları ile kullanılmaktadır. Sağlık verileri hastaneler, diğer sağlık kurumları, sigorta

şirketleri ve ilgili kamu kurumları başta olmak üzere birçok kuruluş tarafından toplanmaktadır. Günümüzde dijital verilerin hacmindeki artış beraberinde yeni sorun alanları da yaratmıştır. Bunların başlıcaları; çok miktarda, çok boyutlu ve karmaşık verileri işlemek için yöntem ya da sistemler geliştirmek; yeni türdeki verileri işlemek için yöntem ya da sistemler geliştirmek; dağılmış verileri işlemek için yöntem, protokol ya da altyapı geliştirmek; verilerin kullanımı ve güvenliği ile ilgili modeller geliştirmek olarak sıralanabilir [4].

Hastalıkların hem tıbbi açıdan hem de hastane kaynak ve maliyetleri açısından ele alınarak doğru yönetilmesi gerekmektedir. Bu noktada bilgi sistemleri üzerinde çalıştırılacak klasik sorgulama yöntemleri yeterli kalmamaktadır. Veri madenciliği yöntemleri kullanılarak bu sistemlerdeki gizli ve önemli bilgiler keşfedilmektedir. Keşfedilen bu bilgiler hem tıbbi araştırmalar hem de yönetim planları için değerlendirilmelidir [5].

2 Veri Madenciliği

Bilgisayar biliminin yeni bir dalı olan veri madenciliği son yıllarda çok fazla ilgiye sahiptir [7].

Veri tabanı sistemlerinin artan kullanımı ve veri depolama ünitelerinin hacimlerdeki artış sorgulama ve raporlama araçlarının dev veri yığınları karşısında etkisiz kalmasına yol açmıştır. Bunun sonucunda veri tabanlarında bilgi keşfi adı altında yeni arayışlar ortaya çıkmıştır[8].

Veri tabanlarında bilgi keşfi, sıklıkla, büyük hacimde veri koleksiyonundan faydalı bilgiyi keşfetmeyi hedefleyen, veri madenciliği olarak anılmaktadır. Veri tabanları günümüzde terabaytlarla ifade edilmektedir. Bu büyük hacimde verinin içinde stratejik önem taşıyan gizli bilgi yatmaktadır. Ama bu kadar büyük hacimli veri içerisinde yer alan önemli bilginin nasıl açığa çıkarılacağı en önemli sorudur. Bu önemli soruya en güncel yanıt, hem geliri artırırken hem de maliyetleri indirgeyen Veri Madenciliğidir [4].

Veri madenciliği gömülü büyük hacimli veriden mantıklı bilginin hızla yayılmasına yönelik bir girişimdir. Veri madenciliği otomatik öğrenme, istatistik, yapay zeka, veri tabanları ve görselleştirme ile de örtüşmektedir [9].

Veri madenciliği ile ilgili farklı tanımlar bulunmakla beraber en fazla kabul gören tanımlar şu şekilde açıklanabilir;

- Veri madenciliği, büyük veri tabanlarından, çok net olmayan, üstü kapalı, önceden bilinmeyen ancak potansiyel olarak kullanışlı olabilecek bilginin çıkarılmasıdır [10].
- Pek çok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları geçerli tahminler yapmak için kullanan bir süreçtir[4].
- Gartner Grup tarafından yapılan tanımda ise veri madenciliği; istatistik ve matematik tekniklerle birlikte ilişki tanıma teknolojilerini kullanarak, depolama ortamlarında saklanmış bulunan veri yığınlarının elenmesi ile anlamlı yeni ilişki ve eğilimlerin keşfedilmesi sürecidir [8].
- Veri Madenciliği büyük miktardaki veri yığını içerisinde gelecekle ilgili tahmin yapmamızı sağlayacak, bağıntı ve kuralların bilgisayar programları kullanılarak aranmasıdır [11].
- Veri madenciliği; veri ambarlarında tutulan çok çeşitli ve çok miktarda veriye dayanarak daha önce keşfedilmemiş bilgileri ortaya çıkarmak, bunları

karar verme ve eylem planını gerçekleştirmek için kullanma sürecidir [12].

Veri Madenciliğinin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar-verme modelleri yaratmaktır. Veri Madenciliği, William Frawley ve Gregory Piatetsky-Shapiro tarafından, '... verideki gizli, önceden bilinmeyen ve potansiyel olarak faydalı bilgilerin önemsiz olmayanlarının açığa çıkarılması...' biçiminde yapılan bilgi keşfi tanımını destekler [4].

Veri madenciliği teknikleri elde edilen veri türüne ve elde edilen sonuçların kullanım amacına göre farklılıklar gösterir. Veri madenciliğinde kullanılan teknikler; Tanımlayıcı (Descriptive), Öngörülse (Predictive) olmak üzere ikiye ayrılır. Tanımlayıcı veri madenciliği, veri tabanındaki verinin genel karakterini, mevcut durumu ortaya çıkarmaya yönelik yöntemleri ön plana çıkarır [8]. Amaç; karar vermeye yardımcı edebilecek mevcut verilerin tanımlanmasını sağlamaktır [13]. Öngörülse veri madenciliği ise verileri geleceğe yönelik tahminler yapma, sonuç çıkarmaya yönelik işlemlerde kullanır [8]. Amaç, sonuçları bilinen verileri kullanarak bir model geliştirmek ve bu model yardımıyla sonuçları bilinmeyen veri kümeleri ile sonuçların tahmin edilmesini sağlamaktır [13].

Veri madenciliği teknikleri, ilişkisel veri tabanları, işlem veri tabanları ve veri ambarları gibi basit ve yapısal veri setleri için uygundur [14].

Veri madenciliği teknikleri kullanıldıkları veri yapılarına ve keşfedebildikleri örüntü biçimlerine göre kategorilere ayrılır. Birçok kaynak veri madenciliği teknikleri için farklı gruplandırmalar yapmıştır. Bunlardan en yaygın kabul göreni J. Han'ın ortaya sürdüğü kategorilerdir. J. Han kategorilerini kullanan kaynaklar bile, hangi algoritmanın hangi kategoriye ait olduğu konusunda net görüş birliğine sahip değildir. Bu kategorileri aşağıdaki gibidir [8].

- Tanımlama ve Ayrılma (Characterization and Discrimination),
- Birliktelik Analizi (Association Analysis)
- Sınıflandırma ve Öngörü (Classification and Prediction),
- Kümeleme Analizi (Cluster Analysis),
- Sıradışılık (istisna) Analizi (Outlier Analysis),
- Evrimsel Analiz (Evolution Analysis).

2.1 Veri Madenciliği Uygulama Alanları

Veri madenciliğinin geniş bir uygulama alanı mevcuttur. Müşterilerin satın alma eğilimlerinin ve alışkanlıklarının belirlenmesi, web sayfalarına erişimlerin analizi, ilaçların yan etkilerinin ortaya çıkartılması, hastalıklar ve şikayetler arasında bağlantıların kurulması, bankacılıkta sahteciliklerin tespit edilmesi, kredi borcu ödeme tahminleri gibi birçok alanda karar verme işlemlerine yardımcı olacak önemli sonuçlar elde edilebilmektedir [15].

Veri madenciliği verinin yoğun olarak üretildiği hemen her ortamda uygulama kullanım alanı bulabilir ve bulmuştur. Bazı uygulama alanları şu şekilde özetlenebilir [16].

- Bilimsel ve mühendislik verileri,
- Sağlık verileri,
- İş verileri,
- Perakendecilik – marketçilik verileri,
- Bankacılık, finans ve borsa verileri,
- Eğitim sektörü verileri,
- İnternet (Web) verileri,

- Doküman verileri,
- Taşımacılık ve ulaşım,
- Turizm ve otelcilik,
- Telekomünikasyon,
- Belediyeler.

2.2 Veri Madenciliğindeki Uygulamalarda Karşılaşılan Sorunlar

Günümüzde birçok kuruluş çok büyük veri tabanlarına sahiptir. Günlük kayıt sayısı birkaç milyon oranında olan, sınırsızca büyüyen veri tabanları vardır. Veri madenlerindeki sürekli veri akışları eşsiz fırsatlarla birlikte yeni zorluklarda getirmektedir [17].

Veri madenciliği sistemleri işlenmemiş veriyi tutmak için dinamik, bitirilmemiş, gürültü içeren ve büyük olma eğiliminde olan veri tabanlarına bel bağlamaktadır. Problemler saklanan bilginin yeterli ve konuyla ilgili olup olmaması sonucunda ortaya çıkmaktadır. Yani, veri madenciliğine girdi olarak kullanılan veri tabanlarının dinamik, eksiksiz, geniş ve net veri içermemesi ve verinin konu ile uyumsuzluğu durumunda sorunlar ortaya çıkmaktadır [12].

Veri madenciliği uygulamalarında karşılaşılan sorunlar şu şekilde belirtilebilir [12].

- Sınırlı bilgi,
- Gürültü ve eksik değerler,
- Kayıp veri,
- Belirsizlik,
- Boyut, güncellemeler ve konu dışı veriler.

3 Uygulama: 2008-2009-2010-2011 Yıllarında PAÜ Hastanesine Gelen Hastaların Profilini Belirleme

Bu çalışmada amaç hasta verileri üzerinden hastaneye gelen hasta kitlesini daha iyi tanımlamaya yardım edecek bilgiler elde edebilmektir.

Çalışma, hastanede kullanılan yazılım programının ilk olarak makbuz ve fatura tahsilatı işlemleri için tasarlanmış olmasından dolayı tüm veriler üzerinde yapılmamıştır. Verilerin daha güncel ve kullanım açısından daha tutarlı olmasından dolayı 2008, 2009, 2010 ve 2011 (2011 yılında ilk 11 aya ait veri incelenmiştir) yıllarını içine alacak şekilde bir çalışma yapılmıştır. Çalışmada ayaktan başvuru yapan hasta grubu üzerinde inceleme yapılmıştır.

Uygulamada kullanılan veriler ilgililerin bilgisi dâhilinde kullanılmıştır. Hastane bilgi sistemi çok kapsamlı bir yapıya sahiptir. Hastanın ilk girişinden çıkışına kadar yapılan tüm işlemler hastane bilgi yönetim sistemi veri tabanında yer almaktadır. Bu çalışmada sadece hastaların dosya ve ilk giriş bilgilerinin yer aldığı tablolar ile bazı tanım tablolarındaki bilgiler kullanılmıştır. İlgili tablolardan sorgulamalar yapılarak uygulamada kullanılan bilgiler elde edilmiştir.

Hastane bilgi yönetim sistemi verileri Microsoft SQL Server 2008 veri tabanı üzerinde tutulmaktadır. Çalışmada öncelikli olarak ilgili tablolardan belirli kriterlere göre sorgulama yapıp, sorgulama sonucu elde edilen veriler bir geçici tabloya kaydedilmiştir. Geçici tablodaki veriler incelenerek hatalı kayıtlar üzerinde veri temizleme ve düzenleme işlemleri yapılmıştır. Düzenlenen tablo üzerinde yer alan verilerden gerekli sorgulamalar yapılarak gelen hastaların profilini belirlemeye kullanılmak üzere bilgiler elde edilmiştir. Bu bilgiler Excel'e aktarılmış, tablo görünümünde özetlenmiş ve grafikler ile görsel hale getirilmiştir. Şekil 1'de sorgulama sonucu elde edilen verilerden örnekler gösterilmiştir.

dogtarih	dogyer	IIAd	IlceAd	cinsiyet	Refno	sevktarih	HastaTuru	PolKod
01.01.1974	2001	DENİZLİ	MERKEZ	Bayan	10 93	2008-04-28 ...	Sevкли	ACIL
01.02.1956	3507	İZMİR	BAYINDIR	Bay	10 49	2008-06-09 ...	Emekli	END
07.04.1972	6301	ŞANLIURFA	MERKEZ	Bay	10 94	2008-03-28 ...	Sevкли	ANEST ...
01.01.1963	4601	KAHRAMANMARAŞ ...	MERKEZ	Bay	10 58	2008-10-09 ...	Sevкли	BC
11.04.1964	0601	ANKARA	MERKEZ	Bay	11 88	2008-06-20 ...	Sevкли	ÜRO
05.09.1949	6401	UŞAK	MERKEZ	Bayan	11 73	2008-01-16 ...	Bağkurlu	GC
03.02.1945	3718	KASTAMONU ...	DOĞANYURT ...	Bayan	11 52	2008-05-05 ...	Emekli	ÜRO
21.12.1943	2001	DENİZLİ	MERKEZ	Bay	12 33	2008-01-02 ...	SSK	KALP
01.08.1964	2001	DENİZLİ	MERKEZ	Bayan	12 21	2008-02-08 ...	Emekli	HEMA
03.02.1945	3718	KASTAMONU ...	DOĞANYURT ...	Bayan	12 94	2008-04-21 ...	Emekli	ÜRO
01.01.1946	2002	DENİZLİ	ACIPAYAM	Bay	12 34	2008-05-12 ...	Emekli	DERM
22.08.1947	3401	İSTANBUL	MERKEZ	Bayan	12 21	2008-01-07 ...	SSK	UYKU
04.11.1965	0910	AYDIN	KUYUCAK	Bay	12 49	2008-01-14 ...	SSK	ÜRO
12.08.1987	6601	YOZGAT	MERKEZ	Bayan	12 82	2008-02-26 ...	Emekli	KARD
01.01.1981	7604	İĞDIR	TUZLUCA	Bay	12 99	2008-01-02 ...	Sevкли	NTIP
01.01.1972	3101	HATAY	MERKEZ	Bayan	12 69	2008-08-05 ...	Sevкли	DHL
01.01.1969	6401	UŞAK	MERKEZ	Bay	12 67	2008-06-16 ...	Sevкли	ÜRO
01.06.1933	2001	DENİZLİ	MERKEZ	Bayan	12 72	2008-02-04 ...	SSK	PC
01.01.1998	0401	AĞRI	MERKEZ	Bay	12 30	2008-01-07 ...	SSK	PC
07.02.1956	6401	UŞAK	MERKEZ	Bay	12 55	2008-01-11 ...	Bağkurlu	END
04.06.1943	2019	DENİZLİ	TAVAS	Bay	12 64	2008-01-29 ...	SSK	PC
03.03.1966	5512	SAMSUN	TERME	Bayan	12 84	2008-02-15 ...	SSK	HEMA
07.11.1982	0201	ADIYAMAN	MERKEZ	Bay	12 77	2008-04-03 ...	Sevкли	ORT
03.12.1988	2001	DENİZLİ	MERKEZ	Bayan	12 58	2008-01-08 ...	SSK	FT
01.07.1968	2011	DENİZLİ	ÇAMELİ	Bay	12 88	2008-01-15 ...	Bağkurlu	ENF
06.07.1948	2010	DENİZLİ	ÇAL	Bay	12 36	2008-01-14 ...	SSK	DERM
01.01.1999	2001	DENİZLİ	MERKEZ	Bayan	12 55	2008-01-04 ...	SSK	PC
01.01.1948	2001	DENİZLİ	MERKEZ	Bay	12 77	2008-01-08 ...	Emekli	PSK
01.01.1985	2001	DENİZLİ	MERKEZ	Bayan	12 59	2008-01-10 ...	Emekli	KDH
01.01.1956	1504	BURDUR	GÖLHİSAR	Bayan	12 97	2008-01-04 ...	Emekli	KDH
16.06.1940	2015	DENİZLİ	HONAZ	Bayan	12 94	2008-01-02 ...	Emekli	KALP

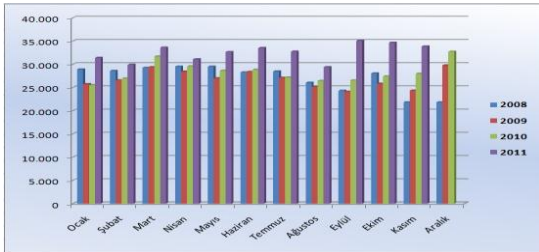
Şekil 1: Sorulama sonucu elde edilen veriler.

3.1 Aylara Göre Hasta Dağılımı

Temiz veriler üzerinde hastalar müracaat tarihlerine göre sorgulanıp aylara göre gruplamalar yapılmıştır. Sorğu sonucu elde edilen veriler Excel'e alınmıştır. Veriler Excel'de düzenlenip grafiksel hale dönüştürülmüştür. Sosyal güvenlik sisteminin tek çatı altına toplanması ve sevk zinciri olmadan hastaların direk istedikleri hastaneye müracaat etmeye başlaması(SSK'lı hastaların direk üniversite hastanesine başvurabilmesi ile hasta sayısı otuz bine yaklaşmıştır) ile 2008 yılının ilk 7 ayında hasta sayısının yaklaşık olarak 28.000-29.500 arasında değiştiği görülmektedir. Kasım ayında SGK'ya tabi hastaların aile hekiminden sevk olarak (sevk zincirine göre müracaat) müracaat etme zorunluluğunun getirilmesi ile hasta sayısında 2008 yılının Kasım ve Aralık aylarında gözle görülür bir şekilde düşüş olmuştur. Sevk zincirine göre müracaatların 2009 yılında kaldırılması ile hasta sayısının tekrar arttığı görülmektedir. Toplam hasta sayısını incelediğimiz zaman 2008 yılındaki hasta sayısı 2009 yılından 2.765 adet daha fazladır. Bazı polikliniklerde günlük poliklinik muayene sayısının artırılması 2010 yılındaki hasta sayısını bir önceki yıla göre 17.576 adet artırmıştır. 2011 yılında da hasta sayısı artmaya devam etmiştir. 2011 yılı Şubat ayından itibaren ek ödemelerin performans sistemine göre yapılmaya başlanması ile birlikte bazı polikliniklerde sınırsız hasta kabul edilmeye başlanması; hasta sayısının artmasında en büyük etken olarak belirtilebilir. Bir önceki yıla göre hasta sayısı bir ay eksik veri olmasına rağmen 18.441 adet artmıştır. Yıllara ve aylara göre hasta sayıları Tablo 1'de yer almaktadır. Şekil 2'de Tablo 1'de yer alan değerler grafiksel olarak gösterilmiştir.

Tablo 1: Aylara göre hasta sayıları.

	2008	2009	2010	2011
Ocak	28.925	25.736	25.531	31.413
Şubat	28.606	26.576	26.954	29.937
Mart	29.279	29.423	31.721	33.611
Nisan	29.507	28.467	29.587	31.096
Mayıs	29.518	27.001	28.631	32.674
Haziran	28.299	28.400	28.844	33.533
Temmuz	28.506	27.133	27.201	32.755
Ağustos	26.050	25.166	26.455	29.424
Eylül	24.331	24.124	26.540	35.065
Ekim	28.079	25.844	27.410	34.682
Kasım	21.843	24.366	27.998	33.852
Aralık	21.847	29.789	32.729	
	324.790	322.025	339.601	358.042



Şekil 2: Aylara göre hasta sayıları grafiği.

3.2 Doğum Yeri Göre Hasta Dağılımı

Temiz veri kümesi üzerinde hastaların doğum yeri bilgilerine göre gruplama yapılmıştır. Doğum yerine göre 2008, 2009, 2010 ve 2011 yıllarındaki hasta gruplarını incelediğimiz zaman hasta sayısı en fazla olan 15 ilin dört yılda da aynı sıralamaya sahip olduğu görülmektedir. Başvuran hastaların doğum yeri bilgileri ve yıllara göre dağılımı Tablo 2'de yer

almaktadır. En fazla hasta sayısına sahip ilk beş il içerisinde; Denizli'ye sınırlı komşusu olan "Afyonkarahisar, Aydın, Burdur, Uşak" illerinin yer aldığı görülmektedir. Veriler incelendiği zaman ilimize komşu olan illerden üniversite hastanesine müracaat eden hasta sayısının fazla olduğunu anlaşılmaktadır. Bu da hastanenin bir bölge hastanesi gibi hizmet verdiğinin göstergesidir. Tablo 2'de yer alan bilgiler kullanılarak Şekil 3, Şekil 4, Şekil 5 ve Şekil 6'da doğum yeri bilgisine göre hasta dağılımları grafiksel olarak gösterilmiştir. Tablo oluşturulurken ve grafikler çizilirken en fazla hasta sayısına sahip 20 il ayrı ayrı gösterilmiş, geriye kalan illerdeki hasta sayıları diğer başlığı altında toplanmıştır.

Tablo 2: Doğum yerine göre hasta sayıları.

	2008	2009	2010	2011
Denizli	217.291	217.273	225.311	238.207
Afyonkarahisar	15.210	13.935	14.846	15.386
Aydın	10.305	10.493	12.050	12.293
Burdur	12.078	10.903	11.338	11.784
Uşak	5.930	5.991	6.896	6.912
İzmir	5.227	5.085	5.923	6.121
Manisa	4.755	4.837	5.471	6.193
Muğla	4.073	4.151	4.676	5.294
Ankara	3.140	2.963	3.184	3.318
İstanbul	2.813	2.839	3.087	3.343
Konya	2.491	2.369	2.667	2.515
Isparta	2.452	2.323	2.330	2.336
Erzurum	1.865	1.691	1.843	2.121
Antalya	1.588	1.448	1.823	2.014
Ağrı	2.161	2.018	1.754	1.984
Almanya	1.343	1.395	1.438	1.337
Balıkesir	1.342	1.154	1.406	1.573
Muş	1.205	1.277	1.364	1.430
Adana	1.324	1.161	1.344	1.603
İçel	996	1.126	1.182	1.274
Diğer	27.201	27.593	29.668	31.004
	324.790	322.025	339.601	358.042

3.3 Hasta Cinsiyetine Göre Hasta Dağılımı

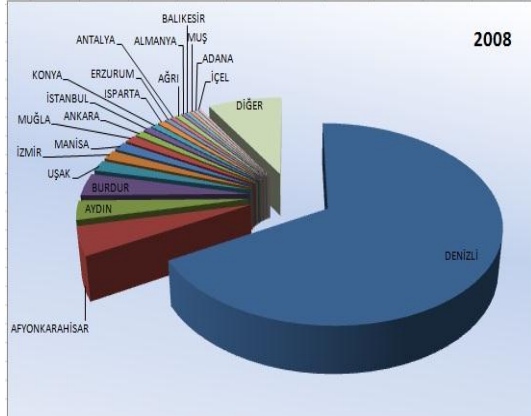
Hasta cinsiyetine göre hasta sayıları Tablo 3'te yer almaktadır. Tablo 3 incelendiği zaman dört yılda da bayan hasta sayısının daha fazla olduğu görülmektedir. Bu bilgi bayanların sağlıklarına daha çok dikkat edip düzenli olarak kontrollerini yaptırdıklarının işareti olarak kabul edilebilir. Ancak, 2009 yılında bayan hasta sayısı 2008 yılına göre azalmış, bay hasta sayısı 2008 yılına göre artmıştır. 2010 ve 2011 yıllarında bay ve bayan hasta sayıları önceki yıllara göre artmıştır. Yüzde olarak karşılaştırdığımız zaman 2008 yılında % 58 Bayan, % 42 Bay; oran 2009 yılında % 1 değişim gösterip % 57 Bayan, % 43 Bay; 2010 ve 2011 yıllarında da 2009 yılı ile aynı yüzdeye sahip olup % 57 Bayan, % 43 Bay şeklindedir. Bu değerler göz önüne alındığında hastaneye bayan hasta müracaatının daha fazla olduğu görüşüne varılabilir. Şekil 7'de cinsiyete göre hasta sayıları grafiksel olarak gösterilmiştir.

Tablo 3: Cinsiyete göre hasta sayıları.

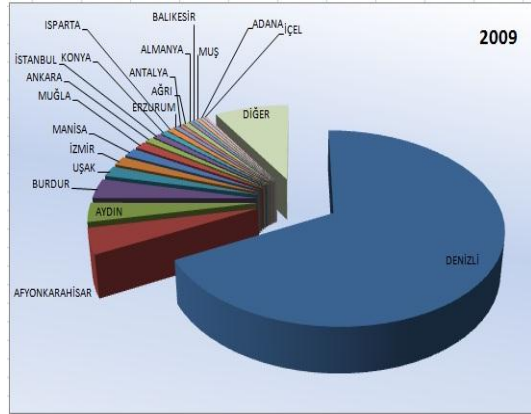
	2008	2009	2010	2011
Bay	135.969	138.057	147.318	154.810
Bayan	188.821	183.968	192.283	203.232
	324.790	322.025	339.601	358.042

3.4 Hasta Türüne Göre Hasta Dağılımı

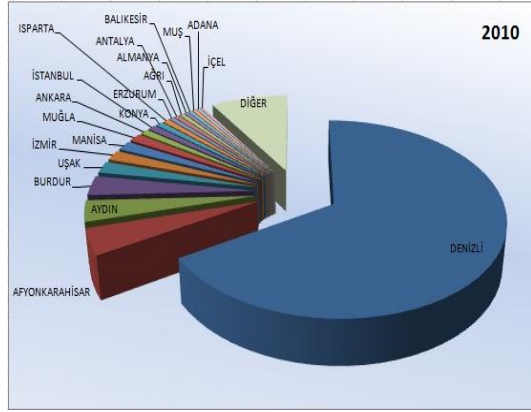
Hasta türü hastanın sosyal güvencesine göre gruplama yapılmasıdır. Yıllara göre hasta verileri hasta türüne göre sorgulanıp gruplama yapılmıştır. Elde edilen veriler Excel'e aktarılmıştır.



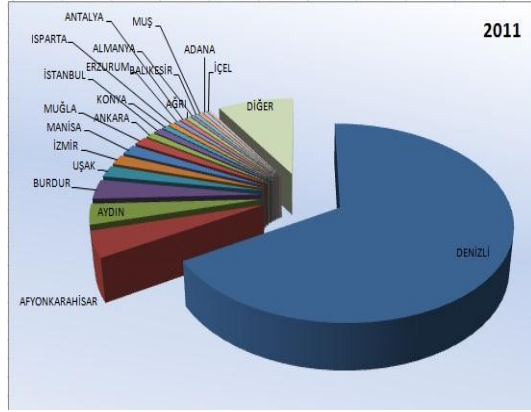
Şekil 3: 2008 yılı doğum yerine göre hasta sayıları grafiği.



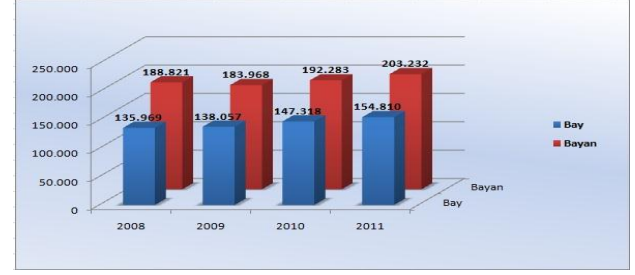
Şekil 4: 2009 yılı doğum yerine göre hasta sayıları grafiği.



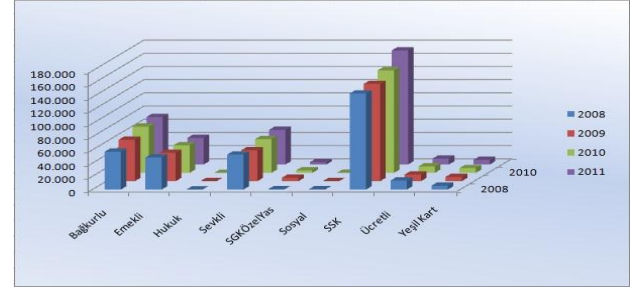
Şekil 5: 2010 yılı doğum yerine göre hasta sayıları grafiği.



Şekil 6: 2011 yılı doğum yerine göre hasta sayıları grafiği.



Şekil 7: Cinsiyete göre hasta sayıları grafiği.



Şekil 8: Hasta türüne göre hasta sayıları grafiği.

Tablo 4'te hasta türüne göre hasta sayısının yıllara göre dağılımı yer almaktadır. Şekil 8'de yer alan hasta türüne göre verilerin grafiksel olarak dağılımı incelendiği zaman dört yılda da SSK'lı hasta sayısının belirgin bir şekilde fazla olduğu görülmektedir. SSK'lı hastaları Bağkurulu, Sevklı ve Emekli hastalar izlemektedir. Başvuran SSK'lı hasta sayısının diğler hasta türlerinden fazla olması; şehirde sigortalı olarak çalışan kişi sayısının da bununla doğru orantılı olarak fazla olduğunun göstergesi olduğu kabul edilebilir. Denizli'nin bir sanayi kenti olması da bu sayının fazla olmasının açıklayıcısıdır.

Tablo 4: Hasta türüne göre hasta sayıları.

	2008	2009	2010	2011
Bağkurulu	57.330	62.808	70.059	72.103
Emekli	48.596	43.096	41.836	40.379
Hukuk	10	3	22	2
Sevklı	53.010	46.783	51.196	52.629
SGK Özel Yas	293	5.285	3.731	3.601
Sosyal	405	54	63	208
SSK	145.691	147.552	155.602	173.096
Ücretli	13.617	10.051	9.876	9.002
Yeşilkart	5.838	6.423	7.216	7.022
Toplam	324.790	322.025	339.601	358.042

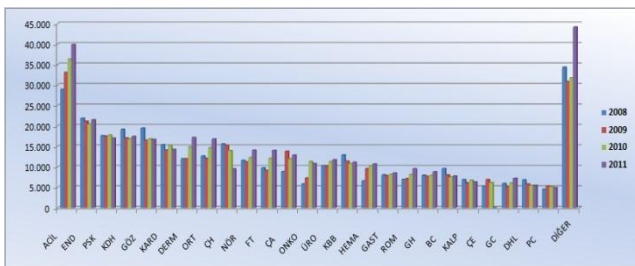
3.5 Polikliniğe Göre Hasta Dağılımı

Yıllara ve hastaların müracaat ettiği bölümlere göre dağılım Tablo 5'te yer almaktadır. Hastalar geldiği bölümlere göre incelendiği zaman ilk 5 bölüme ait bilgi 2008, 2009 ve 2010 yıllarında sadece iki bölümün sıralamada yer değiştirmiş olması ile aynı sıralamaya sahiptir. ACİL, END, PSK, KHD, GÖZ sıralamada en fazla hastaya sahip olan bölümlerdir. En fazla hasta üç yılda da Acil bölümüne gelmiştir. 2008 yılında 3. sırada Göz bölümü 5. sırada Psikiyatri bölümü yer alırken 2009 ve 2010 yıllarında 3. sırada Psikiyatri bölümü 5. sırada Göz bölümü yer almıştır. 2011 yılında ilk 5 bölüm yer değiştirmiştir; Acil ve Endokrinoloji bölümleri ilk iki sırada 3. sırada Kadın Hastalıkları ve Doğum bölümü yer almaktadır. Bu bölümleri Dermatoloji ve Psikiyatri bölümleri izlemiştir. İlk 5 sıralamasında yer alan bölümler incelendiğinde;

Tablo 5: Hastanın geldiği polikliniğe göre hasta sayıları.

	2008	2009	2010	2011
ACİL	28.923	33.052	36.332	39.875
END	21.856	21.135	20.455	21.478
PSK	17.617	17.528	17.763	16.987
KDH	19.177	17.065	16.857	17.405
GÖZ	19.452	16.473	16.829	16.694
KARD	15.415	14.121	15.205	14.283
DERM	12.004	11.976	14.871	17.174
ORT	12.675	12.066	14.652	16.774
ÇH	15.606	15.214	13.962	9.526
NÖR	11.558	11.193	12.227	14.062
FT	9.759	9.134	12.080	14.011
ÇA	8.834	13.782	11.990	12.878
ONKO	5.827	7.272	11.302	10.783
ÜRO	10.277	10.232	11.239	11.723
KBB	12.926	11.380	10.799	11.108
HEMA	6.561	9.566	10.228	10.673
GAST	8.045	7.833	8.156	8.500
ROM	6.934	7.143	8.088	9.522
GH	7.958	7.698	7.881	8.785
BC	9.581	8.028	7.472	7.766
KALP	6.876	6.093	6.707	6.304
ÇE	5.295	6.899	6.155	0
GC	5.927	5.165	6.027	7.173
DHL	6.843	5.803	5.460	5.494
PC	4.537	5.320	5.161	4.901
DiĞER	34.327	30.854	31.703	44.163
	324.790	322.025	339.601	358.042

Acil bölümünün 2009, 2010 ve 2011 yıllarında hasta sayısının artmış; diğer dört bölümün hasta sayısı 2009 yılında azalmış 2010 yılında ise Psikiyatri ve Göz bölümlerinin hasta sayısı bir önceki yıla göre artmıştır. 2011 yılında Endokrinoloji, Kadın Hastalıkları ve Doğum ve Dermatoloji bölümlerinin hasta sayısı bir önceki yıla göre artmış Psikiyatri ve Göz bölümlerinin hasta sayısı azalmıştır. Acile müracaatlarda hasta katılım payının alınmıyor olması, kişilerin acile başvurularında işlemlerin daha hızlı ilerlemesi nedeniyle durumu acil olmasa bile acile başvuru yapması, gündüz çalışan kişilerin mesai saatleri dışında acile başvuru yapabilmesi, Acil bölümünün hasta sayısının artmasının önemli nedenlerinden sayılabilir. Bölümlere göre hastaların dağılımı grafiksel olarak Şekil 9'da gösterilmiştir. Grafik çizilirken en fazla hasta sayısına sahip 25 bölüm ayrı ayrı geriye kalan bölümler toplanıp diğer başlığı altında gösterilmiştir.



Şekil 9: Hastanın geldiği polikliniğe göre hasta sayıları grafiği.

4 Sonuç ve Öneriler

Sağlık ve tıp, günümüzün en çok bilgi ihtiyacı olan araştırma alanlarıdır. Son yıllarda özellikle sağlık veri modelleri, standartlar ve kodlama sistemlerindeki yenilikler sayesinde hastanelerde ve sağlık merkezlerinde kullanılan bilgi sistemlerinde önemli gelişmeler yaşanmıştır. Bu gelişmeler daha çok ve çeşitli verinin saklanabilmesini sağlamış ve beraberinde bilgi keşfi ihtiyacını ortaya çıkarmıştır. Veri madenciliği, sağlık ve tıp alanındaki büyük veri tabanlarından değerli bilgileri ortaya çıkartarak, hem tıp açısından hem de hizmet kalitesinin artırılması açısından büyük katkılar sağlar[5].

Bu çalışma 2008, 2009, 2010 ve 2011 yıllarında PAÜ hastanesine müracaat eden hasta verileri kullanılarak hazırlanan bir veri tabanı sorgulama uygulama çalışmasıdır. Hastaların bazı genel özellikleri değerlendirilerek gelen hastaların profilleri belirlenmeye çalışılmıştır. Elde edilen bilgilerin gelen hastaları daha iyi tanımamıza ve tanımlamamıza yardımcı olacağı düşünülmüştür. Hasta hakkında edinilen bilgiler arttığında hastaya verilen hizmetin nasıl iyileştirilmesi gerektiği konusunda da daha doğru kararlar alınabileceği düşünülmüştür. Örneğin başka illerden gelen hastaların azımsanamayacak sayıda olduğu göz önüne alınıp yapılan tetkik ve tedavi randevularının düzenlenmesinin buna göre yapılması sağlanır hastanın aynı gün içerisinde işlemlerinin tamamlanması sağlanarak tekrar tekrar gidip gelmelerin önüne geçilebilir. Ayrıca şehirde kalacak yeri olmayıp aynı gün içerisinde işlemlerini bitiremeyen hastalar için misafirhaneler hazırlanabilir. Bunlara bağlı olarak ta hastaya verilen hizmet iyileşir ve hasta memnuniyeti artırılabilir.

Elde edilen veriler incelendiğinde 2009 yılı hariç toplam hasta sayısının bir önceki yıla göre hep arttığı görülmüştür. Hasta sayısındaki artış, başvuran hastaların cinsiyetine göre Bay/Bayan hasta yüzdesini pek değiştirmemiştir. 2008 yılında % 58 Bayan, % 42 Bay şeklinde olan hasta yüzdesi; son üç yılda % 57 Bayan, % 43 Bay hasta şeklindedir. Bayan hasta sayısının fazla olması değerlendirip ileride hastaneye ait alışveriş merkezi açılması düşünülürse bayanların ilgi alanlarına hitap eden mağazalara yoğunluk verilebilir. Hastalar 2008-2009-2010-2011 yıllarının hepsinde en fazla müracaatı Acil bölümüne yapmıştır. İkinci sırada en fazla hasta müracaat edilen bölüm Endokrinoloji bölümüdür.

İncelenen parametreler artırılıp, bundan sonraki yıllarda da incelemeler yapıp hasta profili üzerinde daha detaylı bilgiler elde edilip daha iyi tanımlama yapılabilir. Gelen hastalar daha iyi tanımlanıp, edinilen bilgiler ile hastalara verilen hizmet kalitesi artarsa hasta memnuniyeti de aynı oranda artıp hastaneye müracaat eden hasta sayısının da artması sağlanabilir. Hasta sayısının artması ile doğru orantılı olarak ta hastane gelirlerinde artış elde edilebilir.

Sağlık sektöründe gizli bilgilerin keşfedilmesi hayati önem taşımaktadır. Örneğin şikayetler ve hastalıklar arasındaki gizli kalmış ilişkilerin bulunması, tecrübe ile elde edilebilecek bilgilerin önceden keşfedilmesi, çok faktörlü hastalıklarda önemli faktörler arasındaki ilişkilerin ortaya çıkarılması gibi birçok alanda tıbbi veri madenciliğinin önemi tartışılmazdır. Tıp alanında bulunan mevcut veri oldukça fazla ve hayati öneme sahiptir. Hastane bilgi sistemleri sayesinde bu veriler düzenli olarak tutulmaktadır. Bu verilerden daha fazla yararlanmak mümkündür. Hastane bilgi sistemlerinden veya diğer tıbbi veri toplayan sistemlerden alınan veriler üzerinde yapılan veri madenciliği çalışmaları hem uzmanlar için hem

hastane yönetimi için hem de hastaların daha kaliteli bir hizmet almalarında etkin rol alabilir [18]. Veri madenciliği veritabanında depolanan büyük miktarda veriden değerli bilgi keşfetme sürecidir [19]. Veri madenciliği, sağlık profesyonellerinin en doğru ve güncel bilgiye ulaşmasını, en objektif ve optimum çözümleri kullanmasını sağlayacak bir karar destek aracıdır. Geleceğin sayısal karar verme ve iş zekâsı yöntemi olan Veri Madenciliğinin konunun uzmanı kişiler tarafından sağlık sektöründe kullanımı, sağlık hizmetlerinin daha etkin sunumu, kaynakların daha verimli kullanımı ve bilimsel, karşılaştırılabilir, şeffaf bilgi erişimi açısından önerilmektedir [4].

5 Teşekkür

Uygulama verilerinin kullanılmasına izin veren yetkili kişilere teşekkür ederiz.

*: Bu çalışma 12. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu'nda sunulan bildiridir. Çalışma, genişletilip, uygulama verileri de yeniden güncellenip makale haline getirilmiştir

6 Kaynaklar

- [1] Khan, A., Bawane, N.G., Bodkhe, S., "An Analysis of Particle Swarm Optimization with Data Clustering-Technique for Optimization in Data Mining.", (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, 2223-2226, 2010.
- [2] Özekes, S., "Veri Madenciliği Modelleri ve Uygulama Alanları", İstanbul Ticaret Üniversitesi Dergisi, Cilt:2, Sayı 3, 65-82, 2003.
- [3] Gürüler, H., Karahasan, M., İstanbullu, A., "Üniversite Öğrencilerinin Profilini Belirleme: Muğla Üniversitesi Veri Tabanları Üzerinde Bir Durum Çalışması", Muğla Üniversitesi, Sosyal Bilimler Enstitüsü Dergisi (İLKE), Bahar 2007, Sayı 18, 27-37, 2007.
- [4] Koyuncu, A.S., Özgülbaş, N., "Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları", Bilişim Teknolojileri Dergisi, Cilt: 2, Sayı: 2, 21-32, 2009.
- [5] Yıldırım, P., Uludağ, M., Görür, A., "Hastane Bilgi Sistemlerinde Veri Madenciliği", Akademik Bilişim 2008, Çanakkale Onsekiz Mart Üniversitesi, 429-434, 2008.
- [6] Ertuğrul, İ., Hacettepe Üniversitesi, İ.İ.B.F. Dergisi, Cilt: 25, Sayı:1, Ankara, 2007.
- [7] Wang, G., Wang, Y., "3DM: Domain-oriented Data-driven Data Mining", Fundamenta Informaticae 90, 395-426, 2009.
- [8] Arslan, E., "EMG İşaretlerinin İncelenmesi ve Veri Madenciliği Uygulaması", Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2008.
- [9] Mitra, S., Acharya, T., "Data Mining: Multimedia, Soft Computing, and Bioinformatics", 2003.
- [10] Dinçer, E., "Veri Madenciliğinde K-MEANS Algoritması ve Tıp Alanında Uygulanması", Kocaeli Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2006.
- [11] Ayık, Y.Z., Özdemir, A., Yavuz, U., "Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte ile İlişkisinin Veri Madenciliği Tekniği ile Analizi", Atatürk Üniversitesi, Sosyal Bilimler Enstitüsü Dergisi Cilt 10, Sayı 2, 441-454, 2007.
- [12] Akgöbek, Ö., Çakır, F. "Veri Madenciliğinde Bir Uzman Sistem Tasarımı", Akademik Bilişim' 09-XI. Akademik Bilişim Konferansı Bildirileri 11-13 Şubat 2009 Harran Üniversitesi, Şanlıurfa, 801-806, 2009.
- [13] Küçüksille, E., "Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi ve İMKB Hisse Senetleri Piyasasında Bir Uygulama", Süleyman Demirel Üniversitesi, Sosyal Bilimler Enstitüsü, Doktora Tezi, 2009.
- [14] Kholghi, M., Keyvanpour, M., "An Analytical Framework for Data Stream Mining Techniques Based on Challenges and Requirements", International Journal of Engineering Science and Technology (IJEST), Vol. 3 No. 3, 2507-2513, 2011.
- [15] Doğan, Ş., "Veri Madenciliği Kullanarak Biyokimya Verilerinden Hastalık Teşhisi", Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2007.
- [16] Baysakoğlu, A. "Veri Madenciliği ve Çimento Sektöründe Bir Uygulama" Akademik Bilişim 2005, Gaziantep, 2005.
- [17] Mala, A., Dhanaseelan, F.R., "Data Stream Mining Algorithms-A Review of Issues and Existing Approaches", International Journal on Computer Science and Engineering (IJCSE), Vol. 3, No. 7, 2726-2732, 2011.
- [18] Tuğ, E., "Genetik Algoritmalar ile Tıbbi Veri Madenciliği", Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 2005.
- [19] Yin, Y., Kaku, I., Tang, J., Zhu, J.M., "Data Mining: Concepts, Methods and Applications in Management and Engineering Design", 2011.