

# Hastanelerin Gelecekteki Hasta Yoğunluklarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi

*Predicting Future Patient Volumes of The Hospitals By Using Data Mining Methods*

**Sezgin IRMAK**

*Yrd. Doç. Dr., Akdeniz Üniversitesi Alanya İşletme Fakültesi, İşletme Enformatiği Bölümü, (sezgin@akdeniz.edu.tr)*

**Can Deniz KÖKSAL**

*Doç. Dr., Akdeniz Üniversitesi İ.İ.B.F., İşletme Bölümü, (candeniz@akdeniz.edu.tr)*

**Özcan ASILKAN**

*Yrd. Doç. Dr., Akdeniz Üniversitesi Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, (oasilkan@akdeniz.edu.tr)*

## ÖZET

### Anahtar Kelimeler:

*Veri Madenciliği,  
Zaman Serileri,  
Yapay Sinir Ağları,  
Sağlık Sektörü*

*Bu çalışmada hali hazırda işleyen bir hastane veritabanında bazı önemli veri madenciliği teknikleri ile hasta yoğunluklarının tahmin edilmesi uygulamaları yapılmış ve sonuçları karşılaştırmalı olarak aktarılmıştır. Kullanılan hastane veritabanında veri transferi, filtreleme ve veri ön-işleme faaliyetleri gerçekleştirilmiş sonrasında da zaman serileri ve yapay sinir ağları teknikleri kullanılarak birçok veri madenciliği tahmin modeli üretilmiştir. Üstel düzgünleştirme, ARIMA ve yapay sinir ağları yöntemleri önce kendi içlerindeki farklı modellerle kıyaslanmış sonrasında da her yöntemin en kestirimci modelleri birbirleriyle kıyaslanarak bu konuda en iyi sonucu veren model tespit edilmeye çalışılmıştır. Üstel düzgünleştirme yöntemlerinden Winters Additive modeli, ARIMA yöntemlerinden ARIMA(3,1,0)(1,0,0)<sub>12</sub> modeli ve yapay sinir ağları yöntemlerinden Prune yöntemi ile elde edilen model en iyi sonuçları vermiştir. Winters Additive üstel düzgünleştirme modeli ise bunlar arasında en kestirimci model olarak öne çıkmış ve gerçekleşen değerlere en yakın tahminleri üretmiştir.*

## ABSTRACT

### Keywords:

*Data Mining, Time  
Series, Artificial  
Neural Networks,  
Databases,  
Healthcare Industry*

*In this study, some techniques of data mining were used to predict the future volumes of patients in a currently active database of a hospital and the results were presented comparatively. Initially, data transfer, filtering and data pre-processing activities were performed in the hospital database. In order to predict the future volumes of patients, different models of exponential smoothing, ARIMA and neural network techniques were generated and the best models of each technique were determined. The best models of each technique then evaluated again to determine the best predictive model. Winters Additive exponential smoothing model, ARIMA(3,1,0)(1,0,0)<sub>12</sub> model, and ANN model trained by using Prune method have yield better results. The results of comparison showed that Winters Additive exponential smoothing model was the best predictive model for this data and this model made the closest predictions to actual values.*

## 1. GİRİŞ

Veri madenciliği kavramının günümüzde giderek yaygınlaştığı görülmektedir. İş dünyasında işletmeler daha fazla kar sağlamak amacıyla birçok farklı uygulama için veri madenciliği yöntemlerini kullanmaktadır. Bu yöntemlerle işletmelerin veritabanlarından, veri ambarlarından veya farklı biçimlerdeki veri setlerinden elde edilen bilgiler karar destek sistemlerinde ve stratejik planlama çalışmalarında girdi olarak kullanabilmektedir. Ayrıca veri madenciliği yöntem ve tekniklerinin geliştirilmesi çalışmalarına ek olarak, özellikle yığın veri içeren sağlık, bankacılık, perakendecilik vb. gibi farklı alanlardaki akademik çalışmalarda da veri madenciliği yöntemleri giderek daha sık kullanılmaya başlanmıştır.

Veri madenciliğinin kökleri yapay zekâ, makine öğrenimi, matematik, istatistik ve bilgisayar uygulamaları alanlarına dayanmaktadır. Veriden bilgi çıkarımı insanoğlunun yüzyıllar boyunca yaptığı bir iştir. Ancak günümüzde bilgisayar, bilgisayar ağları ve depolama sistemleri teknolojilerindeki gelişmelerle birlikte hayatı büyük miktarlarda veri kuşatmaktadır ve bu verinin büyük bir kısmı organizasyonel, ticari, bilimsel, stratejik ve benzeri amaçlarla büyük veritabanları veya veri ambarlarında saklanmaktadır.

Büyük, karmaşık ve bilgi zengini veri setlerinin anlaşılabilirliği ihtiyacı, işletme, bilim ve mühendisliğin bütün alanları için ortaktır. İşletmecilik dünyasında, şirket ve müşteri verisi şirketin stratejik varlıkları olarak algılanmaya başlamıştır ve bu verilerde saklı olan faydalı bilgilerin çıkarılması ve bu bilgiye dayalı hareket edebilme yeteneği, bugünün rekabetçi dünyasında giderek artan bir öneme sahiptir (Kantardzic, 2003: 1-2). Dolayısıyla işletmelerin gelecekte veri

madenciliği ve bununla bağlantılı konularda daha fazla yatırım yapmalarının bir zorunluluk haline geldiği görülmektedir. Bu zorunluluğa paralel olarak ise akademik alanda veri madenciliği alanıyla ilgili eğitim ve araştırma faaliyetlerinin yaygınlaşması bir ihtiyaç olarak ortaya çıkmaktadır.

Sağlık alanı yığın veri tutulan alanlardan biridir ve bu alan veri madenciliği faaliyetleri sonucunda elde edilecek faydalı bilgilerden en çok yararlanabilecek alanlardan biridir. Bu çalışmada bir hastane veritabanı kullanılarak veri madenciliği uygulamaları gerçekleştirilmiştir. Hastanenin gelecek aylardaki hasta yoğunluklarının tahmin edilmesinde üstel düzgünleştirme, ARIMA ve yapay sinir ağı teknikleri önce kendi içinde farklı modellerle karşılaştırılmış sonrasında da her tekniğe ait en iyi modeller kendi aralarında kıyaslanmıştır. Böylece en kestirimci model belirlenmeye çalışılmıştır.

## 2. VERİ MADENCİLİĞİ

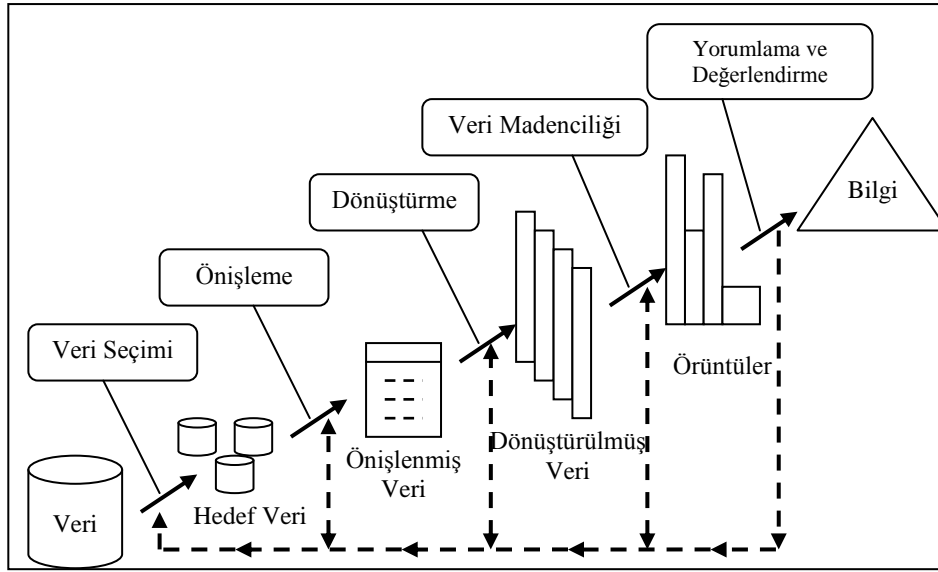
Veri madenciliği kavramını anlayabilmek için işin en başında kelimelerin yalın anlamlarından yola çıkılabilir. Madencilik yeryüzünün gizli ve kıymetli kaynaklarının açığa çıkarılması süreci olup, bu kelimenin veri kelimesi ile ilişkilendirilmesi ise veri yığınları içerisinde ilk bakışta fark edilemeyen kıymetli bilgilerin bulunması ve çıkartılması fikrini uyandırmaktadır (Giudici, 2003: 1).

Bununla birlikte veri madenciliğinin kendine özgü birtakım karakteristik özellikleri olmalıdır. Bir sistemin veri madenciliği sistemi olabilmesi için büyük miktarlarda veri ile çalışabilmesi, birleşik sorgulara cevap verebilir bir yapıda veri ve bilgi geri alma işlemlerini gerçekleştirebilmesi gerekmektedir (Han ve Kamber, 2006: 9).

### 2.1. Veri Madenciliği Süreci

Veri madenciliği sadece bir yöntem ve teknikler topluluğu olarak değil, probleme özgü tasarlanmış, ilgili yöntem, teknik ve uygulamaları da içine alan, sonuçları itibariyle probleme özgü olmak üzere ilişkileri, kuralları, örüntüleri, eğilimleri, vb. modelleyen ve gösteren bir süreç olarak algılanmalıdır (Kantardzic, 2003).

Şekil 1. Veri Madenciliği Sürecini Oluşturan Aşamalar



Kaynak: (Fayyad vd., 1996: 41)

Literatürde bu sürecin tanımlanması konusunda henüz bir ortak fikir oluşmadığı görülmektedir. Veri madenciliği bazı kaynaklarda veritabanlarında bilgi keşfi kavramı ile birbirinin yerine geçebilir şekilde kullanılmaktadır ve süreç veri madenciliği süreci veya veritabanlarında bilgi keşfi süreci olarak adlandırılmaktadır (Han ve Kamber, 2006: 7; Tsai ve Tsai, 2005; Kendall ve Kendall, 2008: 516). Bazı kaynaklarda ise süreç veri madenciliği süreci (Olaru ve Wehenkel, 1999; Brohman, 2006; Delavari, 2005) ve bu işlevi gerçekleştiren bir sistem de veri madenciliği sistemi (Li ve Khosla, 2005) olarak adlandırılmaktadır. Ancak birtakım kaynaklar veri madenciliğini veritabanlarında bilgi keşfi sürecinde bir aşama olarak konumlandırmaktadırlar (Tan, vd., 2006: 3; Fu, 1997; Dunham, 2003; Fayyad, vd., 1996) ve sürecin bütünü veritabanlarında bilgi keşfi kavramı ifade etmektedir. Birkaç kaynaktan ise bu kavram karmaşası içerisinde iki kavram birleştirilerek süreç bilgi keşfi ve veri madenciliği süreci (Microsoft Research, 2006; Fayyad, 1996) olarak isimlendirilmektedir.

İlgili literatürde en yaygın kullanılan gösterimlerden biri olan veri madenciliği süreci Şekil 1’de sunulmuştur. Veri madenciliği sürecinde öncelikle araştırılmak istenilen veri seti içerisinde hedeflenen verinin seçimi gerçekleştirilir. Bu veri seti veritabanı veya veri ambarı olabileceği gibi yalnızca metin dosyaları içerisinde tutulan veri de olabilir. Seçilen hedef veri üzerinde veri ön işleme işlemleri gerçekleştirilerek hatalı veriler, eksik değerler gibi sorunlar halledilir. Veri madenciliği modelleri gerçekleştirilmeden önce eğer ihtiyaç varsa veri üzerinde dönüştürme işlemleri yapılır yeni değişkenler elde edilebilir veya mevcut değişkenlerin değerleri yeniden düzenlenebilir. Daha sonra veri madenciliği algoritmaları kullanılarak modeller elde edilir. Elde edilen modeller yorumlanarak bilgiye dönüştürme işlemi gerçekleştirilmiş olur.

### 3. HASTA YOĞUNLUKLARININ TAHMİNİNDE KULLANILAN VERİ MADENCİLİĞİ TEKNİKLERİ

#### 3.1. Yapay Sinir Ağları

Yapay sinir ağları (Artificial Neural Networks – ANN) insan beyninin bilişsel öğrenme sürecinin benzetimini yapmak amacıyla ilk olarak 1940’ların başında ortaya atılmıştır. Yapay sinir ağları, birincil fonksiyonları deneme yanılmaya dayalı olarak problem uzayının bir modelini yapılandırmak olan bilgisayar-tabanlı yöntemlerdir. Modelin yapılandırılması süreci kavramsal olarak bir kısım veri setinin yapay sinir ağına gönderilmesiyle başlar ve yapay sinir ağı bir çıktı değeri tahmin eder. Bu tahmin değeri, gerçek değer ile bir geribildirim biçimi ile karşılaştırılır. Eğer tahmin doğru ise ağ başka bir faaliyet göstermez. Eğer tahmin yanlış ise yapay sinir ağı, kestirimin kalitesini arttırmak için hangi iç parametrelerin ne şekilde düzeltilmesi gerektiğini belirlemek için kendi kendisini analiz eder. Bu parametre ayarlamaları yapıldıktan sonra yapay sinir ağı veri setinin başka bir bölümünü alır ve süreci tekrar eder. Bu süreç içerisinde zamanla yapay sinir ağı büyük oranda doğru bir modele yakınsamaya başlar (Marakas, 2003: 130-131).

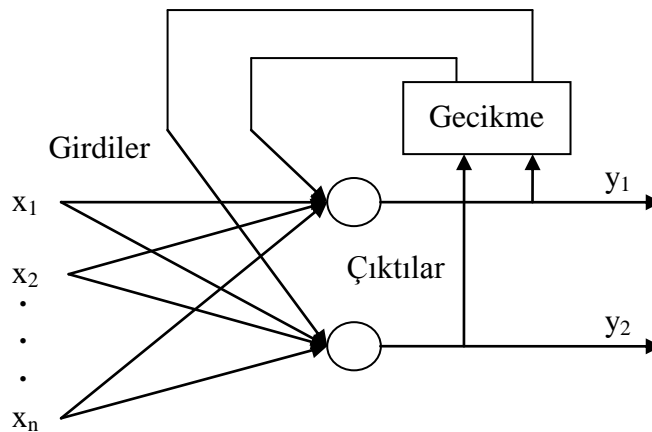
Yapay sinir ağları yüksek oranda parametrelerle ifade edilen istatistiksel model sınıflarından biridir ve son yıllarda oldukça çok ilgi çekmiştir. Yüksek oranda parametrelerle ifade edildiğinden dolayı yapay sinir ağları oldukça esnekler ve bu yüzden fonksiyonlardaki nispeten küçük aykırılıkları doğru bir şekilde modelleyebilirler. Diğer yandan böylece bir esneklik ciddi bir aşırı-uyum (overfitting) tehlikesini beraberinde getirir (Hand vd., 2001: 391).

Yapay sinir ağları geniş bir yelpazede ve birçok alanda problemlerin çözümünde kullanılabilen çok güçlü ve esnek sayısal araçlardan biridir (Pham vd., 2006: 307). Yapay sinir ağlarının kullanılma amaçları arasında sınıflandırma, örüntü tanıma ve eşleştirme, örüntü tamamlama, optimizasyon ve kontrol sayılabilir (Tsetsekas vd., 2006: 2987). Bu alanda yapılan güncel çalışmalar yapay sinir ağlarının güçlü örüntü sınıflandırma ve kestirim kapasitelerini kanıtlamışlardır (Zhang, 2004: 1). Bu ağların yapısı kestirimci değişkenler ve bağımlı değişken arasındaki çok karmaşık ilişkileri yakalamayı desteklemektedir. Yapay sinir ağlarının bu gücü diğer sınıflandırıcı yöntemlerde genellikle mümkün olmamaktadır (Shmueli vd., 2007: 167).

#### 3.1.1. Yapay Sinir Ağları Mimarisi

Gizli katmana sahip olmayan ve girdi ve çıktı katmanları olmak üzere iki katmana sahip yapıya yapay sinir ağları literatüründe perceptron adı verilmektedir. Bir yapay sinir ağı ise perceptron modelinden daha karmaşık bir yapıya sahiptir (Tan vd., 2006: 247-251). Yapay sinir ağları, katmanları arasındaki bağlantıların tiplerine göre genel olarak, ileri beslemeli (feedforward) ve yinelemeli (recurrent) olmak üzere iki kategoride sınıflandırılırlar.

Şekil 2. Yinelemeli (Recurrent) Yapay Sinir Ağı Mimarisi

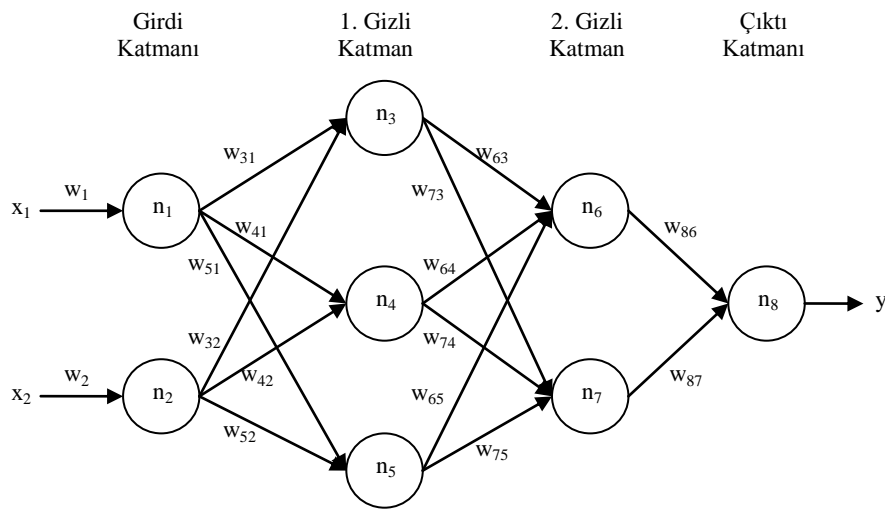


Yapay sinir ağında döngüsel yol oluşturan bir geri besleme (feedback) hattı varsa, bu ağ yinelemeli ağ olarak adlandırılır. Bu yapıda genellikle senkronizasyonu sağlamak amacıyla bir gecikme (delay) bileşeni de kullanılmaktadır. Şekil 2’de bir yinelemeli yapay sinir ağı mimarisi görülmektedir (Katardzic, 2003: 200).

Yinelemeli yapay sinir ağları dinamik yapılarıdır ve mimarileri statik olanlara göre geri besleme içermelerinden dolayı temel olarak farklıdır. Yinelemeli ağlar, ileri beslemeli olanlara göre daha küçük bir mimariye sahip olmaktadır (Goh ve Mandic, 2003: 1095).

Yapay sinir ağlarında en yaygın kullanılan mimari, yapılarındaki esneklik, iyi temsil yetenekleri ve geniş sayıdaki öğrenme algoritmaları ile ileri beslemeli ağlardır (Ma ve Khorasani, 2003: 361). İleri beslemeli yapay sinir ağları (Feedforward Neural Networks) bir girdi katmanı, bir veya daha fazla gizli katmanı ve bir de çıktı katmanı olan ağlardır. Girdi katmanı, eğitim seti girdilerini alarak ağırlıklandırır ve çıktılarını kendisinden sonra gelen gizli katmana eş zamanlı olarak gönderir. Bir gizli katmanın çıktıları, kullanılan gizli katman sayısına bağlı olarak bir sonraki gizli katmanın girdileri olabilir. Gizli katman sayısı analizcinin takdirine göre seçilebilir, ancak uygulamada genellikle tek gizli katman kullanılmaktadır. Son gizli katmanın ağırlıklandırılmış çıktıları, çıktı katmanının girdileridir ve bu katman çıktı olarak yapay sinir ağının verilen eğitim seti için yaptığı kestirimi verir (Han ve Kamber, 2006: 328).

Şekil 3. İleri Beslemeli Yapay Sinir Ağı

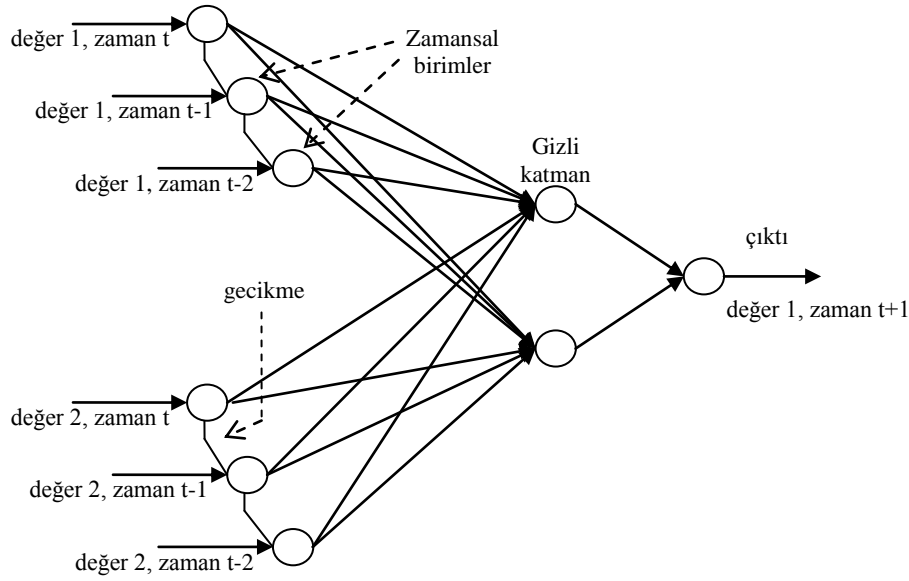


Şekil 3’de iki gizli katmanı olan bir ileri beslemeli yapay sinir ağı modeli görülmektedir. Bazı kaynaklarda ileri beslemeli yapay sinir ağına, çok katmanlı perceptron (multilayer perceptron) adı verilmektedir (Giudici, 2003: 111; Wang ve Fu, 2005: 25). Bishop (2005: 116-117) çok katmanlı perceptron kavramı ile ileri beslemeli yapay sinir ağı kavramlarının her ikisini birden birbirinin yerine geçebilir şekilde kullanmıştır. Bazı kaynaklarda bu yapıya çok katmanlı ileri beslemeli yapay sinir ağı (multilayer feedforward neural network) adı verilmektedir (Han ve Kamber, 2006: 328; Tan vd., 2006: 251). Bazı kaynaklarda ise sadece ileri beslemeli yapay sinir ağı (feedforward neural network) adı verilmektedir (Arulampalam ve Bouzerdoum, 2003: 561; Kantardzic, 2003: 200; Tang ve MacLennan, 2005: 248; Benardos ve Vosniakos, 2007: 365).

### 3.1.2. Zaman Serilerinde Yapay Sinir Ağları Kullanımı

Yapay sinir ağları, zaman serisi analizleri için kolayca adapte edilebilmektedir. Şekil 4’de zaman serileri için adapte edilmiş bir yapay sinir ağı modeli görülmektedir. Yapay sinir ağı, zaman serisi verisi üzerinde en eski veri noktasından başlayarak öğrenir. Daha sonra öğrenme ikinci en eski veri noktası ve girdi setindeki bir sonraki zaman noktası şeklinde ilerleyerek devam eder. Ağın öğrenmesi ileri beslemeli yapay sinir ağı biçimindedir ve geri yayılım algoritması her adımda seride yer alan bir sonraki değeri kestirmeye çalışır (Berry ve Linoff, 2004: 245).

Şekil 4. Zaman Gecikmeli Yapay Sinir Ağı



Geleneksel zaman serisi yöntemleri olasılıksal istatistik kavramlarına dayanmaktadır, ancak son yıllarda yapay sinir ağları kavramı zaman serisi tahmin yöntemleri arasına entegre olmaya başlamıştır. Yapay sinir ağları ve geleneksel zaman serisi yöntemleri karşılaştırmalarının birçok çalışmada yapıldığı görülmektedir (Tseng vd., 2002: 72). Bazı çalışmalarda kullanılan test setlerinin tahminlerinde yapay sinir ağlarının en iyi sonuçları verdiği gösterilmiştir (Kohzadi vd., 1996; Hill vd., 1996; Prybutok vd., 2000; Ho vd., 2002; Zou vd., 2007). Maier ve Dandy (1996) ise, ARIMA modellerinin kısa-dönem tahminlerde, yapay sinir ağlarının ise uzun-dönem tahminlerde kullanımını önermektedir.

### 3.2. Zaman Serileri

Bir zaman serisi veritabanı, zaman içerisinde tekrarlayan ölçümlerden elde edilen değerleri içerir. Bu değerler tipik olarak eşit zaman aralıklarında ölçülür (günlük, haftalık ve yıllık gibi). Zaman serisi veritabanları, hisse senedi piyasası analizleri, ekonomik tahminler, bütçe analizi, süreç ve kalite kontrolü ve iş yükü tahminleri gibi uygulamalarda kullanılmaktadır (Han ve Kamber, 2006: 489). Zaman serisi analizi, zamansal veri içerisindeki örüntü ve trendleri keşfetmek için güçlü bir tekniktir. Ancak bu veri madenciliği tekniği için kavramsal model eksikliği analizcinin yapılandırılmamış veri ile uğraşmasına neden olur. Bu türdeki veri düşük seviyede bir düzenlenmişliğe sahiptir ve yönetimi zordur (Zubcoff vd., 2009: 977).

Zaman serileri, değişkenlerin gün, hafta, ay, mevsim veya yıl gibi herhangi bir zaman dilimine göre dağılımını gösteren serilerdir. Çeşitli değişkenler için düzenlenmiş zaman serileri için özel tahmin teknikleri geliştirilmiştir. Zaman serisi analizlerinin kullanıldığı en önemli alanlar ekonomi ve işletme alanlarıdır. Tahmin edilecek değişkenlerin geçmiş değerlerinin çeşitli yöntemlerle incelenmesine dayanan zaman serisi analizleri altı grupta toplanmaktadır (Orhunbilge, 1999: 3-4):

1. Zaman serilerinin bileşenlerine ayrılması yöntemi (Decomposition methods)
2. Üstel düzgünleştirme yöntemleri (Exponential smoothing methods)
3. Otoregresif yöntemler (Autoregressive methods)
4. Hareketli ortalama yöntemleri (Moving average methods)
5. Bileşik otoregresif hareketli ortalama yöntemleri (Autoregressive integrated moving average methods)
6. Zaman serilerinde regresyon yöntemleri (Regression methods in time series)

Mevcut çalışmada zaman serisi analizlerinde üstel düzgünleştirme yöntemleri ve bileşik otoregresif hareketli ortalama (Box-Jenkins) yöntemleri kullanıldığından burada bu iki yönteme ve ayrıca zaman serilerinde yapay sinir ağlarının kullanılmasına değinilecektir.

#### 3.2.1. Üstel Düzgünleştirme Yöntemleri

İleriye dönük kestirim sistemlerinde, zaman serilerinin kestirimleri ilerleyen periyotlar için her periyotta yapılmaktadır. Dolayısıyla kestirim denklemi ve zaman serisi parametrelerinin tahminleri her periyodun sonunda en yeni gözlemi de hesaba katarak düzeltilmelidir. Bu düzeltme işlemi parametrelerdeki zaman içerisinde olabilecek değişimleri hesaba katacak şekilde yapılmalıdır. Buna ek olarak, bu değişiklikler, parametre tahminleri düzeltilindiğinde zaman serisi

gözlemlerine eşit olmayan ağırlıklar uygulanması gerekliliğini gösterebilir. Üstel düzgünleştirme, gözlenen zaman serisi değerlerini eşit olmayan bir biçimde ağırlıklandırılan bir kestirim yöntemidir (Bowerman ve O'Connell, 1993: 379).

Üstel düzgünleştirme yöntemleri özellikle verinin gürültülü olduğu zamanlarda veriyi düzgünleştirmek için kullanışlı olabilmektedir (Kirkup, 2002: 379). Üstel düzgünleştirme, zaman serisinin trend ve mevsimsel faktör gibi bileşenleri zaman içerisinde değişken olduğunda etkin bir tahmin yöntemidir. Bu yöntem, gözlenen zaman serisi değerlerini eşit olmayan bir şekilde ağırlıklandırmaktadır. Daha yakın gözlemler daha yüksek ağırlıklandırılırken, daha uzak gözlemler daha düşük ağırlıklandırılmaktadır. Eşit olmayan ağırlıklandırma bir veya daha fazla düzgünleştirme sabiti (smoothing constant) kullanılarak gerçekleştirilir ve bu sabit her gözleme ne kadar ağırlık atanacağını belirlemektedir (Bowerman vd., 2005: 345).

Mevsimsel zaman serileri için en yaygın bilinen ve kullanılan yöntemler Winters (1960) tarafından öne sürülen yöntemlerdir. Bunlardan biri toplamsal mevsimsellik için olan Toplamsal (Additive) Holt-Winters yöntemi, diğeri ise çarpımsal mevsimsellik için önerilen Çarpımsal (Multiplicative) Holt-Winters yöntemidir (Koehler vd., 2001: 269).

### 3.2.2. Bileşik Otoregresif Hareketli Ortalama (Box-Jenkins) Yöntemi

George E. P. Box ve Gwilym M. Jenkins 1970 yılında otoregresif modeller ile hareketli ortalama yöntemlerinin bileşimi ARIMA (Autoregressive Integrated Moving Average Model) yöntemini geliştirmişlerdir. Literatürde Box-Jenkins yöntemi adıyla da geçen bu teknik tanımlama, tahmin ve testlerden oluşan üç aşamalı bir yöntemdir. Bileşik otoregresif hareketli ortalama yöntemi, otoregresif ve hareketli ortalama modellerinin bileşimi olduğu için bu modeller hakkındaki tüm bilgilerin bir arada kullanılmasına dayanmaktadır (Orhunbilge, 1999: 189).

Box-Jenkins modelleri, mevsimsel ve mevsimsel olmayan modeller şeklinde ikiye ayrılmaktadır. Mevsimsel olmayan Box-Jenkins modelleri genel olarak ARIMA(p,d,q) şeklinde gösterilmektedir. Burada p, otoregresyon (AR) modelinin derecesi, d, fark alma işlemi sayısı ve q, hareketli ortalama (MA) modelinin derecesi olmaktadır. Mevsimsel Box-Jenkins modelleri ise genellikle ARIMA(p,d,q)(P,D,Q)<sub>s</sub> biçiminde ifade edilmektedir. Burada P, mevsimsel otoregresyon (SAR) modelinin derecesi, D, mevsimsel fark alma işlemi sayısı, Q, mevsimsel hareketli ortalama (SMA) modelinin derecesi ve s, periyot olmaktadır (Kadılar, 2005: 185). Bir bileşik otoregresif hareketli ortalama modelinde, bir değişkenin gelecekteki değeri, geçmiş gözlemlerin ve rassal hataların doğrusal bir fonksiyonu olarak varsayılır

## 4. VERİ MADENCİLİĞİ İLE HASTA YOĞUNLUKLARININ TAHMİNİ

Bu çalışmada veri madenciliği yazılımı olarak SPSS firmasının Clementine 11.1 paketi kullanılmıştır. Clementine, CRISP-DM endüstri standart süreç modeli çerçevesinde tasarlanmış bir veri madenciliği çalışma tezgahı (workbench) olarak tanımlanmaktadır (SPSS, 2007a: 1). İlgili yazılımla veri madenciliği algoritmalarının uygulanması ile birlikte veri seçme, veri birleştirme, veri dönüştürme gibi veri ön işleme aşamaları ve veri madenciliği modellerinin elde edilmesinden sonra model değerlendirme ve görselleştirme işlemleri gerçekleştirilebilmektedir. Bu çalışmada hastanenin Oracle 9i veritabanlarına erişilerek bu veritabanı içerisindeki tablolar veri kaynağı olarak kullanılmıştır.

Yoğunluk tahmini analizleri hastaneye yapılan toplam aylık başvuru sayıları dikkate alınarak gelecek aylarda gerçekleşecek olan başvuru yoğunluklarının tahmin edilmesi amacıyla gerçekleştirilmiştir. Veri setini hazırlamak amacıyla HASTA\_KABUL tablosunda yer alan 1 Ocak 2001 ve sonrasında gerçekleşmiş tüm kayıtlar ait oldukları yıl ve aya göre etiketlenmiştir. Yapay sinir ağları ve zaman serisi analizleri burada toplam hasta sayısı yoğunluğunu tahmin etmek üzere gerçekleştirilmiştir. Burada gelecek aylardaki toplam hasta sayısını tahmin etmek amacıyla 96 aylık veri kullanılarak (1) üstel düzgünleştirme yöntemleri ile geliştirilen zaman serisi modelleri, (2) ARIMA yöntemleri ile geliştirilen zaman serisi modelleri ve (3) yapay sinir ağları ile geliştirilen tahmin modelleri incelenmiştir. Her yöntemin modelleri önce kendi içinde kıyaslanmış, daha sonra da her yöntemin en iyi modelleri birbirleriyle kıyaslanmıştır.

### 4.1. Üstel Düzgünleştirme Modelleri

Bu aşamada Clementine Zaman Serileri Modelleme aracı kullanılarak; Winters Additive, Winters Multiplicative, Simple Seasonal, Damped Trend, Holt's Linear Trend ve Brown's Linear Trend üstel düzgünleştirme yöntemleriyle hasta sayısı yoğunluğu gelecek 9 ay için tahmin edilmiş ve birbirleriyle kıyaslanmıştır.

Tablo 1'de 2009 yılı Eylül ayına kadar farklı üstel düzgünleştirme yöntemleri ile yapılmış olan 9 aylık toplam hasta sayısı tahminleri yer almaktadır. Tabloda ayrıca bu aylar için gerçekleşen hasta sayıları da verilmiştir. Böylece her yöntemin tahmin rakamları ile gerçek değerler karşılaştırılmaktadır. Özellikle 9. aya (Eylül) gelindiğinde dahi bazı modellerin tahminlerinin gerçekleşen değerlere oldukça yakın olduğu görülmektedir.

Üstel düzgünleştirme yöntemleri ile geliştirilen modeller arasında hangi modelin en iyi kestirimci olduğunu anlamak için, Tablo 2'de geliştirilen modellerin Uyum İyiliği (Goodness of Fit) Kriterleri verilmiştir.

**Tablo 1. Üstel Düzgünleştirme Yöntemleri 2009 Yılı 9 Aylık Hasta Sayısı Tahminleri**

2009 Yılı (Aylar)	Gerçekleşen Toplam Hasta Sayısı	Üstel Düzgünleştirme Model Tahminleri (Toplam Hasta Sayısı)					
		Winters Additive	Winters Multiplicative	Simple Seasonal	Damped Trend	Holt's Linear Trend	Brown's Linear Trend
Ocak	64.615	61.737	65.002	63.305	57.432	57.495	56.414
Şubat	58.330	58.548	59.632	57.622	57.718	57.857	56.106
Mart	66.385	63.719	63.551	61.084	58.004	58.220	55.799
Nisan	64.946	62.003	62.012	59.036	58.289	58.582	55.492
Mayıs	62.632	63.937	62.886	59.463	58.574	58.944	55.185
Haziran	67.427	61.685	61.319	57.597	58.858	59.307	54.877
Temmuz	64.100	59.881	60.456	56.649	59.142	59.669	54.570
Ağustos	59.534	57.779	55.938	52.761	59.426	60.031	54.263
Eylül	59.310	59.826	56.620	53.760	59.709	60.393	53.955

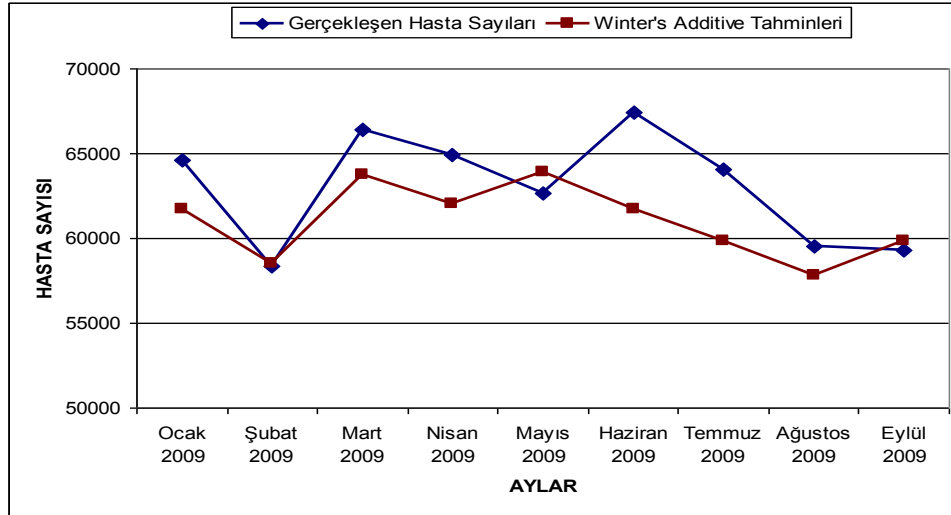
Burada modellerin uyum iyiliği kriterleri birbirleri ile karşılaştırmalı bir şekilde değerlendirilmektedir.  $R^2$ , yaygın olarak bilinen bir ölçüdür ve doğrusal modelin uyum iyiliği ölçütüdür, çoğunlukla determinasyon katsayısı olarak da isimlendirilmektedir. 0 ile 1 arasında değişmekte ve küçük değerler modelin veriye uyumunun iyi olmadığını göstermektedir. Durağan  $R^2$  ise modelin durağan kısmıyla temel modeli karşılaştıran bir ölçüdür. Bir eğilim (trend) veya mevsimsel bir örüntü olduğu durumda tercih edilmektedir. RMSE, hata karelerinin ortalamasının kareköküdür. Bağımlı serilerin model tarafından kestirilen seviyeden ne kadar farklı olduğunu ifade etmek için kullanılmaktadır. Küçük değerleri model kestirimlerinin daha iyi olduğunu göstermektedir. MAPE, ortalama mutlak yüzde hatayı göstermektedir ve serilerin birimlerinden bağımsızdır, dolayısıyla farklı serilerin karşılaştırılmasında da kullanılabilir. MAE, ortalama mutlak hatayı göstermektedir ve serilerin kendi birimleriyle ifade edilmektedir. MaxAPE, en yüksek mutlak yüzde hata ölçüsüdür. Tahmin edilen değerler arasında gerçekleşen en yüksek hatayı gösterir, yüzde olarak ifade edilir dolayısıyla birimden bağımsızdır. Tahminler arasında en kötü senaryo uygulamaları için kullanılabilen bir ölçüdür. MaxAE, en yüksek mutlak hatayı göstermektedir ve bağımlı seri ile aynı birimde ifade edilmektedir. Norm. BIC, normalize Bayesyan bilgi kriteri, modelin toplam uyumunun genel ölçüsüdür. Bu ölçü, aynı seriler söz konusu olduğunda farklı modeller arasında karşılaştırma yapmak için kullanılmaktadır ve düşük değerler daha iyi bir modeli göstermektedir (SPSS, 2007b: 675-676).

**Tablo 2. Üstel Düzgünleştirme Modellerinin Karşılaştırılması**

Model	Uyum İyiliği Kriterleri							
	Drğn. $R^2$	$R^2$	RMSE	MAPE	MAE	Max. APE	Max.AE	Norm. BIC
Winters Additive	0,517	0,924	3.150,91	6,11	2.299,91	43,56	10.585,57	16,25
Winters Multiplicative	0,387	0,913	3.373,92	5,98	2.355,25	42,27	10.271,70	16,39
Simple Seasonal	0,416	0,915	3.311,18	6,04	2.353,96	39,32	10.934,05	16,30
Damped Trend	0,277	0,880	3.948,08	8,03	2.998,15	41,33	11.169,11	16,70
Holt's Linear Trend	0,762	0,880	3.930,56	8,12	3.019,42	41,80	10.945,61	16,64
Brown's Linear Trend	0,738	0,868	4.095,31	8,38	3.167,58	40,29	10.288,93	16,68

Tablo 2'den  $R^2$ , RMSE (Root Mean Square Error, Hata Kareleri Ortalamasının Karekökü), MAE (Mean Absolute Error, Ortalama Mutlak Hata), ve Norm. BIC (Normalized Bayesian Information Criterion, Normalize Bayesyan Bilgi Kriteri) değerlerine göre en iyi kestirimci modelin Winters Additive modeli olduğu görülmektedir.

Şekil 5. 2009 Yılı ilk 9 Ayı için Hasta Sayısı Tahminleri ve Gerçekleşen Değerler (Winters Additive Modeli Tahminleri)



Şekil 5'te 2009 yılı ilk 9 aylık dönemi için Winters Additive modelinin tahmin değerleri ve bu dönemde gerçekleşen hasta sayıları grafik olarak karşılaştırılmıştır. Bu seri için en iyi üstel düzgünleştirme modeli olan Winters Additive tahminlerinin çoğu aylarda gerçekleşen değerlere çok yakın olduğu, en büyük sapmanın Haziran ayında 5.742 hasta ile gerçekleştiği görülmektedir. Bu da yaklaşık %8,5'lik bir hata anlamına gelmektedir. Ancak diğer aylarda ve özellikle 8. ve 9. aylarda elde edilen yakın tahminler modelin kestirimci gücünün yüksekliği konusunda bir fikir vermektedir.

#### 4.2. ARIMA Modelleri

Tablo 3'de karşılaştırma yapmak amacıyla incelenen ARIMA modelleri ve bunlara ait uyum iyiliği kriterleri verilmiştir.

Tablo 3. ARIMA Modellerinin Karşılaştırılması

Model	Uyum İyiliği Kriterleri							
	Drğn. R <sup>2</sup>	R <sup>2</sup>	RMSE	MAPE	MAE	Max APE	Norm. BIC	Ljung-Box Q
ARIMA (p,d,q)(P,D,Q) <sub>s</sub>								
(1,0,0)(0,0,0)	0,867	0,867	4.155,11	8,45	3.170,55	43,66	16,80	94,31
(1,0,0)(1,0,0) <sub>12</sub>	0,897	0,897	3.679,66	7,28	2.727,21	42,63	16,61	76,26
(0,0,1)(0,0,0)	0,840	0,840	4.568,61	9,01	3.463,16	42,69	16,99	92,50
(0,0,1)(0,0,1) <sub>12</sub>	0,866	0,866	4.196,14	8,12	3.160,51	42,02	16,87	71,60
(1,0,1)(0,0,0)	0,885	0,885	3.883,08	8,03	2.985,59	42,45	16,71	77,82
(1,0,1)(1,0,1) <sub>12</sub>	0,913	0,913	3.423,28	6,67	2.488,04	41,61	16,56	53,10
(0,1,1)(0,0,0)	0,276	0,879	3.968,00	8,05	3.010,91	41,14	16,71	79,85
(0,1,1)(0,1,1) <sub>12</sub>	0,306	0,876	3.878,13	6,81	2.746,34	40,88	16,73	86,66
(1,1,1)(0,0,0)	0,299	0,883	3.926,68	8,23	3.061,15	40,56	16,74	81,09
(1,1,1)(1,1,1) <sub>12</sub>	0,416	0,896	3.602,22	6,36	2.545,55	41,80	16,69	67,67
(2,0,0)(1,0,0) <sub>12</sub>	0,915	0,915	3.356,68	6,63	2.461,65	42,50	16,47	68,33
(2,0,0)(2,0,0) <sub>12</sub>	0,917	0,917	3.344,85	6,50	2.409,80	42,03	16,51	49,35
(0,0,2)(0,0,1) <sub>12</sub>	0,889	0,889	3.837,95	7,40	2.834,50	44,69	16,74	59,80
(0,0,2)(0,0,2) <sub>12</sub>	0,895	0,895	3.760,51	7,09	2.724,94	42,75	16,75	61,91
(2,0,1)(2,0,1) <sub>12</sub>	0,917	0,917	3.377,30	6,55	2.443,44	41,76	16,63	55,44
(1,0,2)(1,0,2) <sub>12</sub>	0,927	0,927	3.179,12	6,39	2.334,74	39,23	16,50	25,72
(2,0,2)(1,0,1) <sub>12</sub>	0,925	0,925	3.202,44	6,54	2.373,29	38,52	16,52	27,57
(2,0,2)(2,0,2) <sub>12</sub>	0,928	0,928	3.181,94	6,42	2.330,53	38,66	16,60	20,17
(3,0,0)(1,0,0) <sub>12</sub>	0,919	0,919	3.294,82	6,44	2.407,41	40,49	16,48	36,22
(3,1,0)(1,0,0) <sub>12</sub>	0,507	0,918	3.274,26	6,60	2.451,54	39,22	16,33	26,63
(3,0,1)(1,0,0) <sub>12</sub>	0,921	0,921	3.286,86	6,35	2.374,32	40,30	16,52	30,70
(1,0,3)(0,0,1) <sub>12</sub>	0,920	0,920	3.291,59	7,02	2.544,46	37,67	16,53	22,15
(0,1,3)(0,0,1) <sub>12</sub>	0,469	0,911	3.454,56	7,28	2.664,88	37,69	16,58	28,09
(0,0,3)(0,0,1) <sub>12</sub>	0,911	0,911	3.469,40	6,96	2.596,55	41,39	16,58	29,67
(3,0,3)(1,0,1) <sub>12</sub>	0,927	0,927	3.208,42	6,42	2.334,68	38,33	16,62	22,63

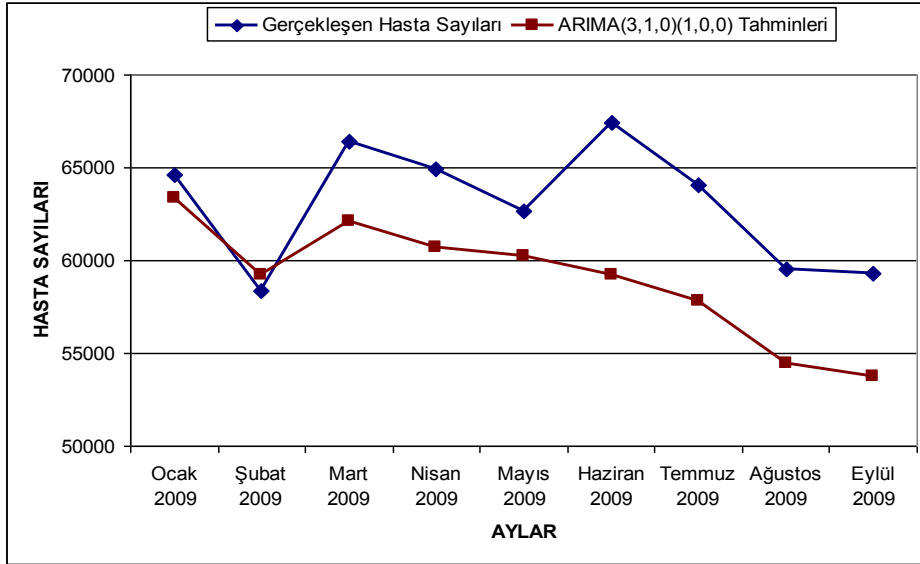


Bileşik Otoregresif Hareketli Ortalamalar (Auto Regressive Integrated Moving Averages, ARIMA) yöntemleri ile gelecek aylardaki hasta yoğunluğunun tahmin edilmesi analizlerinde ARIMA (p,d,q)(P,D,Q)<sub>s</sub> parametreleri kullanılarak mevsimsel ve mevsimsel olmayan AR(1), MA(1), IMA(1,1) ve ARMA(1,1) süreçleri incelenmiş mevsimsel modellerin mevsimsel olmayanlara göre daha kestirimci olduğu görülmüştür. Bundan sonra AR(2), MA(2), ARMA(2,1), ARMA(1,2), ARMA(2,2), AR(3), ARMA(3,1), ARMA(1,3), MA(3), ve ARMA(3,3) süreçleri incelenmiştir.

Burada da en kestirimci modeli seçmek için Normalize edilmiş Bayesyan Bilgi Kriteri (Normalized Bayesian Information Criterion, Norm.BIC) kullanılmıştır. Bu kriter Akaike Bilgi Kriteri'ne (Akaike Information Criterion, AIC) benzer bir kriter ve bu kriterle ilişkili olmakla birlikte, karşılaştırılan model parametrelerine koyduğu sınırlandırma daha güçlü olduğu için aşırı uyum (overfitting) problemi konusunda daha hassas bir ölçüdür. Bu ölçüye göre ARIMA modelleri arasında en iyi kestirimci model ARIMA(3,1,0)(1,0,0)<sub>12</sub> olarak görülmektedir. Bu model için Norm.BIC = 16,332 R<sup>2</sup> = 0,918 olmuştur.

Clementine programında yer alan Expert Modeller aracı tüm alternatifleri deneyerek ARIMA modelleri içinde en düşük Norm.BIC değerine sahip modeli çıkartmaktadır. Bu araç kullanıldığında da ARIMA(3,1,0)(1,0,0)<sub>12</sub> en kestirimci ARIMA modeli olarak bulunmuştur. Ancak Expert Modeller birden fazla alternatif model getirmemektedir. Bunun için farklı ARIMA süreçleri ile elde edilen modelleri kendi aralarında kıyaslamak için Tablo 3'de yer alan modeller tek tek oluşturularak denenmiş ve sonuçlarına kıyaslama amacıyla burada yer verilmiştir.

Şekil 6. 2009 Yılı ilk 9 Ayı için Hasta Sayısı Tahminleri ve Gerçekleşen Değerler (ARIMA(3,1,0)(1,0,0)<sub>12</sub> Modeli Tahminleri)



Şekil 6'da ARIMA(3,1,0)(1,0,0)<sub>12</sub> modelinin 2009 yılı ilk 9 ayına ait tahmin değerleri ve bu aylarda gerçekleşen toplam hasta yoğunlukları verilmiştir. Şekil 6 incelendiğinde ilk aylarda daha yakın tahminler görülürken süre arttıkça tahmin değerleri ve gerçekleşen değerler arasında daha fazla fark olduğu görülmektedir. Yine ARIMA modelinde de Üstel Düzgünleştirmede olduğu gibi en fazla tahmin hatasının Haziran ayında 8.230 hasta (%12,2) ile gerçekleştiği görülmektedir.

### 4.3. Yapay Sinir Ağları Modelleri

Yapay sinir ağlarının zaman serilerinde kullanımı giderek artmaktadır. Yapay sinir ağları ile yapılan tahminler bazen zaman serileri ile yapılan tahminlerden daha iyi sonuçlar verebilmektedir ancak bu her zaman olan bir durum değildir. Bu nedenle gelecek kestirimi konusunda yapay sinir ağlarının, zaman serileri modelleri ile kıyaslamaları yapılmaktadır. Bu çalışmada gelecekteki hasta yoğunluklarının yapay sinir ağları ile tahmin edilmesi uygulaması için Clementine'de bulunan yapay sinir ağları modelleme aracı kullanılmıştır.

Bu aşamada gerçekleştirilen analizde geri yayılım algoritması ile ağırlıkları düzelten ileri beslemeli yapay sinir ağı ve girdi değişkeni olarak da önceki zaman serisi yöntemlerinde olduğu gibi zaman endeksi (TimeIndex) değeri kullanılmıştır.

Yapay sinir ağı modellerinin eğitilmesi Exhaustive Prune, Prune, RBFN, Multiple ve Dynamic olmak üzere beş farklı yöntemle gerçekleştirilmiştir. Dynamic yönteminde bir başlangıç topolojisi oluşturulmakta ve modelin eğitimi sürecinde gizli birimler (katman veya düğüm) eklenerek veya çıkartılarak topoloji iyileştirilmektedir. Multiple yönteminde ise başlangıçta farklı topolojilere sahip birden fazla yapay sinir ağı üretilmekte ve ağlar paralel yöntemle eğitilmektedir. Bunun sonucunda en düşük ortalama hata karelerinin karekökü (RMSE) değerine sahip ağ son model olarak seçilmektedir. Prune yönteminde başlangıçta geniş bir yapay sinir ağı ile başlanmakta ve modelin eğitilmesi sürecinde gizli veya girdi katmanlarındaki en zayıf birimler elimine edilerek en iyi model bulunmaya çalışılmaktadır. RBFN (Radial Basis Function Network, Radyal Tabanlı Fonksiyon Ağları) yöntemi çok boyutlu uzayda hedef değişkenin değerlerine bağlı olarak eğri

uydurma yaklaşımıdır. Modelin eğitimi daha az zaman gerektirir ancak iyi sonuçların alınabilmesi için fazla miktarda veriye ihtiyaç duymaktadır. Exhaustive Prune yönteminde ise Prune yöntemine benzer bir yaklaşım söz konusudur ancak model eğitimi parametreleri mümkün modeller uzayının tamamının taranmasından emin olunacak şekilde seçilmektedir. Bu modelin eğitimi en yavaştır ancak genellikle en iyi sonuçları üretmesi beklenir. Bununla birlikte Exhaustive Prune yaklaşımında aşırı öğrenme probleminin ortaya çıkması da beklenebilir bir durumdur.

Tablo 4’de beş farklı yapay sinir ağı eğitim yöntemi ile elde edilen modellere ait bilgiler verilmiştir. Model topolojilerine bakıldığında Prune ve RBFN’de tek gizli katman yer aldığı diğerlerinin ise ikişer gizli katmana sahip olduğu görülmektedir. Ancak RBFN modelinde gizli katmanda 20 gizli birim (düğüm) kullanılırken, Prune yöntemiyle elde edilen modelin gizli katmanında sadece 2 nöron vardır. Modeller arasında en küçük yapıya sahip olan da budur. Yine Multiple yöntemle elde edilen model iki gizli katmanda toplam 6 nöron içerirken Dynamic ve Exhaustive Prune ile elde edilen modeller hem ikişer gizli katmana hem de daha fazla nöron sayısına sahiptir.

**Tablo 4. Eğitilen Yapay Sinir Ağı Modelleri ve Sonuçları**

Model Eğitim Yöntemleri	Exhaustive Prune	Prune	RBFN	Multiple	Dynamic
Model Topolojisi	1:9:4:1	1:2:1	1:20:1	1:4:2:1	1:12:6:1
En Düşük Hata (Min.Error)	-9.666	-11.221	-9.226	-11.932	-9.606
En Yüksek Hata (Max. Error)	7.039	13.805	6.424	16.269	7.786
Ortalama Hata (Mean Error)	-308,43	-149,88	-73,49	-115,72	133,18
Ortalama Mutlak Hata (MAE)	2.687,70	3.259,26	2.366,69	3.778,00	2.794,00
Standart Sapma	3.359,18	4.240,07	2.977,09	4.982,34	3.370,14
Doğrusal Korelasyon	0,955	0,930	0,965	0,903	0,954

Tablo 4’de yer alan sonuçlar incelendiğinde veriye en fazla uyum sağlayan modelin RBFN olduğu görülmektedir. RBFN modeline çok yakın sonuçları olan diğer iki yöntem de Exhaustive Prune ve Dynamic modelleri olmuştur.

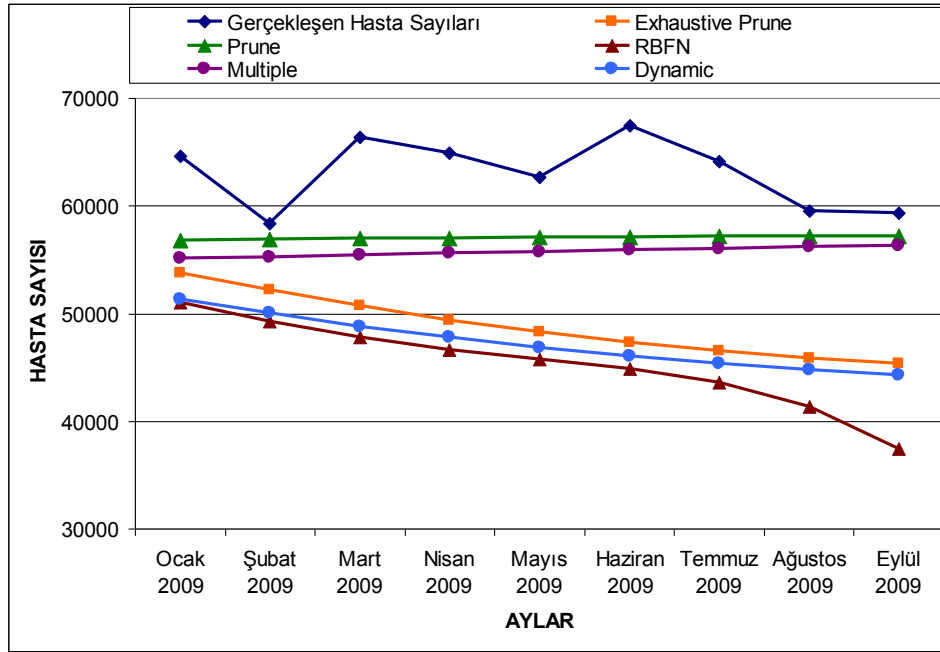
**Tablo 5. Yapay Sinir Ağı Modelleri 2009 Yılı Hasta Sayısı Tahminleri**

Aylar	Gerçekleşen Hasta Sayıları	Hasta Sayısı Tahminleri				
		Exhaustive Prune	Prune	RBFN	Multiple	Dynamic
Ocak 2009	64615	53744	56839	51033	55101	51323
Şubat 2009	58330	52156	56906	49295	55273	50004
Mart 2009	66385	50694	56966	47818	55438	48816
Nisan 2009	64946	49392	57018	46670	55598	47766
Mayıs 2009	62632	48263	57065	45783	55752	46852
Haziran 2009	67427	47306	57106	44910	55900	46063
Temmuz 2009	64100	46506	57142	43617	56043	45387
Ağustos 2009	59534	45847	57174	41333	56181	44810
Eylül 2009	59310	45309	57202	37437	56314	44317

Bu üç model ile Prune ve Multiple yöntemleri ile elde edilen modeller arasında veriye uyum açısından belirgin bir fark bulunmaktadır. Nispeten veriye daha düşük uyum sağlayan bu iki model arasında ise sonuçları itibarıyla Prune yöntemiyle elde edilen model daha iyi bir model olarak gözükmektedir. RBFN modelinin veriye oldukça yüksek uyum gösterdiği görülmektedir. Ancak yapay sinir ağları modellerinin iyi bir kestirimci olup olmadığının ölçülmesi için gelecek 9 aydaki tahmin değerleri ve gerçekleşen hasta sayıları dikkate alınmıştır. Bu değerler Tablo 5’de verilmiştir.

Şekil 7'de de modellerin tahmin değerleri ve gerçekleşen hasta sayılarına ait grafikler verilmiştir. Burada RBFN yöntemiyle eğitilen modelin geleceğe yönelik tahminlerinin diğer modellere göre daha kötü olduğu görülmektedir.

Şekil 7. Yapay Sinir Ağları Modellerinin Tahminleri ve Gerçekleşen Hasta Sayıları



Yapay sinir ağları modelleri arasındaki farklılık üstel düzgünleştirme veya ARIMA yöntemleriyle üretilen modellerin kendi aralarındaki farklılıklardan çok daha belirgin olmuştur. Şekil 7'deki tahmin değerleri grafikleri incelendiğinde gerçekleşen hasta sayısına en yakın tahmin değerlerinin Prune yöntemiyle elde edilen model ile ürettiği görülmektedir. Ondan sonra da en yakın tahmin sonuçlarını Multiple yöntemi modelinin ürettiği görülmektedir. Bu sonuçlara göre hasta sayısının tahmini için en iyi kestirimci yapay sinir ağı modelinin Prune yöntemiyle elde edilen ve 2 nöronlu 1 gizli katmanı bulunan model olduğu görülmektedir.

#### 4.4. Üstel Düzgünleştirme, ARIMA ve Yapay Sinir Ağları Model Sonuçlarının Karşılaştırılması

Üstel düzgünleştirme, ARIMA ve yapay sinir ağları modelleri hastanenin gelecek dokuz aylık dönemdeki hasta sayılarının tahmin edilmesi amacıyla kendi içinde karşılaştırılmış ve en iyi tahminleri yapan modeller belirlenmiştir. Burada ise her üç yöntemin en iyi modelinin hangisinin daha iyi kestirimci olduğu incelenecektir.

Tablo 6. Üstel Düzgünleştirme, ARIMA ve Yapay Sinir Ağı Modellerinin Uyum İyiliği Kriterleri

Model	R <sup>2</sup>	RMSE	MAPE	MAE	MaxAPE	MaxAE
Üstel Düzgünleştirme (Winters Additive)	0,924	3.150,91	6,11	2.299,91	43,56	10.585,57
ARIMA(3,1,0)(1,0,0) <sub>12</sub>	0,918	3.274,26	6,60	2.451,54	39,22	9.529,66
Yapay Sinir Ağları (Topoloji 1:2:1)	0,859	4.220,59	8,38	3.259,26	42,12	13.805

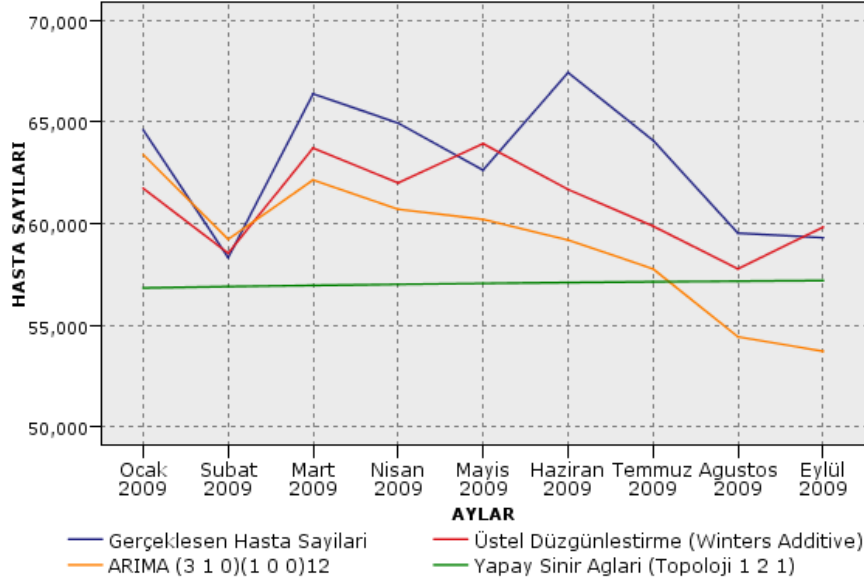
Tablo 6'da ilgili modellerin uyum iyiliği kriterleri özetlenmiştir. Bu kriterlere göre bakıldığında en büyük mutlak hata (MaxAE) ve en büyük mutlak yüzde hata (MaxAPE) kriterleri dışında Winters Additive üstel düzgünleştirme yöntemi modelinin en iyi değerlere sahip olduğu, bu kriterlere göre ise en düşük hata değerlerine ARIMA modelinin sahip olduğu görülmektedir. Ancak ortalama hatalara bakıldığında; ortalama mutlak yüzde hata (MAPE) ve ortalama mutlak hata (MAE) bakımından üstel düzgünleştirme modelinin hatasının daha düşük olduğu görülmektedir. Ortalama hatalar bakımından ise yapay sinir ağı modelinin hata değerleri diğerlerinden yüksektir. Hata kareleri ortalamalarının karekökü (RMSE) ve R<sup>2</sup> değerine göre ise en iyi model Winters Additive üstel düzgünleştirme modelidir. Bundan sonra sırasıyla

ARIMA(3,1,0)(1,0,0)<sub>12</sub> modeli ve Prune yöntemiyle elde edilen, bir gizli katmanı ve iki gizli katman nöronu bulunan yapay sinir ağı modeli gelmektedir.

Şekil 8’de Winters Additive üstel düzgünleştirme modeli, ARIMA(3,1,0)(1,0,0)<sub>12</sub> modeli ve yapay sinir ağı modeline ait gelecek dokuz aylık hasta sayısı tahminleri ve gerçekleşen hasta sayılarının grafikleri verilmiştir.

Şeki 8 incelendiğinde Winters Additive üstel düzgünleştirme modelinin tahminlerinin gerçekleşen hasta sayısı değerlerine diğerlerinden daha yakın olduğu görülmektedir. Ağustos ve Eylül aylarında ise yapay sinir ağı modelinin tahminlerinin ARIMA(3,1,0)(1,0,0)<sub>12</sub> modeli tahminlerinden daha iyi olduğu görülmektedir.

Şekil 8. Üstel Düzgünleştirme, ARIMA ve Yapay Sinir Ağı Modelleri Tahminleri ve Gerçekleşen Hasta Sayıları



## 5. SONUÇ

Veri madenciliği giderek günümüzün önemi ve yaygınlığı artan teknolojilerinden biri olmaktadır. Özellikle yığın halde verinin üretildiği alanlarda üretilen bilgi stratejik varlık olarak görülmektedir ve bu alanlarda veri madenciliğinin gelecekte çok daha yoğun olarak kullanılacağı görülebilmektedir. Yığın veri üretilen alanlardan önemli bir tanesi de sağlık hizmetleri alanıdır. Dolayısıyla sağlık alanında veri madenciliği yöntemlerinden faydalanılarak gerek tıbbi amaçlarla gerekse hastane yönetimlerine veya sağlık hizmetleri konusunda politika yapıcılara karar desteği sağlamak amacıyla veriden bilgi üretimi bu alandaki gelişmeye önemli ölçüde katkı sağlayabilecektir.

Çalışmada kestirimci analiz olarak gelecekteki hasta yoğunluklarının tahmin edilmesi amaçlanmış ve üç farklı veri madenciliği tekniği ve bunların da kendi içinde farklı modelleri üretilerek gelecekteki hasta yoğunluklarının tahmin edilmesi ve bu konuda en iyi modellerin belirlenmesi amaçlanmıştır. Üstel düzgünleştirme yöntemleri arasında en kestirimci model Winters Additive modeli olmuştur. ARIMA süreçleri içinde en kestirimci model ARIMA(3,1,0)(1,0,0)<sub>12</sub> modeli olmuştur. Yapay sinir ağları yöntemleri arasında ise en kestirimci model Prune yöntemiyle elde edilen model olmuştur. Yapay sinir ağları modelleri arasında veriye daha fazla uyum gösteren modeller olmasına rağmen bunlarda aşırı-öğrenme probleminin gerçekleştiği görülmüştür.

Her yöntemin en kestirimci modellerinin birbirleriyle kıyaslanması, uyum iyiliği kriterleri ve modellerin tahmin değerleri ile hastane veritabanından elde edilen gerçekleşen hasta sayısı değerlerinin karşılaştırılması yöntemleri kullanılarak gerçekleştirilmiştir. Her iki konuda da Winters Additive üstel düzgünleştirme modeli en kestirimci model olmuştur. Uyum iyiliği kriterleri bakımından ve ilk 7 aydaki tahminlerin gerçeğe yakınlığı bakımından ARIMA(3,1,0)(1,0,0)<sub>12</sub> modeli ikinci en iyi olsa da 8. ve 9. aylardaki tahmin değerleri kötüye gitmiş ve yapay sinir ağları modeli bu aylarda Winters Additive üstel düzgünleştirme modelinden sonra ikinci en iyi tahminleri gerçekleştirmiştir. Gerçekleşen sayıların tahminlere oldukça yakın olması bu tekniklerin hastanenin yoğunluk tahminleri için kullanılabilirliğini göstermektedir. Ayrıca veri madenciliği teknikleri kullanılarak sağlık sektörü veritabanları veya veri ambarlarından birçok amaç için faydalı bilgilerin elde edilmesi de mümkün görülmektedir.

**KAYNAKÇA**

- ARULAMPALAM, G. ve BOUZERDOUM, A., "A Generalized Feedforward Neural Network Architecture for Classification and Regression", *Neural Networks*, 16, (2003), 561-568.
- BERNARDOS, P. G. ve VOSNIAKOS, G. C., "Optimizing Feedforward Artificial Neural Network Architecture", *Engineering Applications of Artificial Intelligence*, 20, (2007), 365-382.
- BERRY, M. J. A. ve LINOFF, G. S., *Data Mining Techniques for Marketing, Sales, and Customer Relationship Management (Second Edition)*, Wiley Publishing Inc., Indianapolis, Indiana, 2004.
- BISHOP, C. M., *Neural Networks for Pattern Recognition*, Oxford University Press, Great Clarendon Street, Oxford, UK, 2005.
- BOWERMAN B. L. ve O'CONNELL, R. T., *Forecasting and Time Series: An Applied Approach*, Third Edition, Duxbury Thomson Learning, Pacific Grove, CA, USA, 1993.
- BOWERMAN, B. L., O'CONNELL, R. T. ve KOEHLER, A. B., *Forecasting, Time Series, and Regression: An Applied Approach*, Fourth Edition, Thomson Brooks/Cole, Belmont, CA, USA, 2005.
- BOX, G. E. P. ve JENKINS, G. M., *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1970.
- BROHMAN, M. K., "Knowledge Creation Opportunities in the Data Mining Process", *Proceedings of the 39th Hawaii International Conference on System Sciences*, Vol.8, (2006), 1-10.
- DELAVARI, N., BEIKZADEH, M. R. ve PHON-AMNUAISUK, S., "Application of Enhanced Analysis Model for Data Mining Processes in Higher Educational System", *IEEE ITHET 6th Annual International Conference*, Juan Dolio, Dominican Republic, (2005), F4B/1-6.
- DUNHAM, M. H., *Data Mining: Introductory and Advanced Topics*, Prentice-Hall, Upper Saddle River, NJ, USA, 2003.
- FAYYAD, U. M., "Data Mining and Knowledge Discovery: Making Sense Out of Data", *IEEE Intelligent Systems*, 11(5), (1996), 20-25.
- FAYYAD, U. M., Piatetsky-Shapiro, G. ve Smyth, P., "From Data Mining to Knowledge Discovery in Databases", *Artificial Intelligence Magazine*, Fall, (1996), 37-54.
- FU, Y., "Data Mining: Tasks, Techniques and Applications", *IEEE Potentials*, 16(4), (1997), 18-20.
- GIUDICI, P., *Applied Data Mining: Statistical Methods for Business and Industry*, John Wiley & Sons, West Sussex, England, 2003.
- GOH, S. L. ve MANDIC, D. P., "Recurrent Neural Networks with Trainable Amplitude of Activation Functions", *Neural Networks*, 16, (2003), 1095-1100.
- HAN, J. ve KAMBER, M., *Data Mining: Concepts and Techniques*, 2nd Edition, Morgan Kaufmann Publishers, San Francisco, CA, USA, 2006.
- HAND, D., MANNILA, H. ve SMYTH, P., *Principles of Data Mining*, The MIT Press, Cambridge, Massachusetts, USA, 2001.
- HILL, T., O'CONNOR, M. ve REMUS, W., "Neural Network Models for Time Series Forecasts", *Management Science*, 42, (1996), 1082-1092.
- HO, S. L., XIE, M. ve GOH, T. N., "A Comparative Study of Neural Network and Box-Jenkins ARIMA Modeling in Time Series Prediction", *Computers & Industrial Engineering*, 42, (2002), 371-375.
- IRMAK, S., *Veri Madenciliği Yöntemleri ile Sağlık Sektörü Veritabanlarında Bilgi Keşfi: Tanımlayıcı ve Kestirimci Model Uygulamaları*, Akdeniz Üniversitesi S.B.E. Yayınlanmamış Doktora Tezi, 2009.
- KADILAR, C., *SPSS Uygulamalı Zaman Serileri Analizine Giriş*, Bizim Büro Basımevi, Ankara, 2005.
- KANTARDZIC, M., *Data Mining: Concepts, Models, Methods, and Algorithms*, IEEE Press, Hoes Lane, Piscataway, NJ, USA, 2003.
- KENDALL, K. E. ve KENDALL, J. E., *Systems Analysis and Design*, 7/E, Prentice Hall, Upper Saddle River, NJ, USA, 2008.
- KIRKUP, L., *Data Analysis with Excel: An Introduction for Physical Scientists*, Cambridge University Press, Cambridge, UK, 2002.
- KOEHLER, A. B., SNYDER, R. D. ve ORD, J. K., "Forecasting Models and Prediction Intervals for the Multiplicative Holt-Winters Method", *International Journal of Forecasting*, 17, (2001), 269-286.
- KOZHADI, N., BOYD, M. S., KERMANSHAHI, B. ve KAASTRA, I., "A Comparison of Artificial Neural Network and Time Series Model for Forecasting Commodity Prices", *Neurocomputing*, 10, (1996), 169-181.
- LI, Q. ve KHOSLA, R., "Performance Optimization of Data Mining Applications Using a Multi-layered Multi-agent Data Mining Architecture", *CIMSA 2005 – IEEE International Conference on Computational Intelligence for Measurement Systems and Applications*, Giardini Naxos, Italy, July (2005), 227-231.
- MA, L. ve KHORASANI, K., "A New Strategy for Adaptively Constructing Multilayer Feedforward Neural Networks", *Neurocomputing*, 51, (2003), 361-385.
- MAIER, H. R. ve DANDY, G. C., "Neural Network Models for Forecasting Univariate Time Series", *Neural Networks World*, 6, (1996), 747-772.

- MARAKAS, G. M., *Modern Data Warehousing, Mining, and Visualization: Core Concepts*, Prentice Hall, Upper Saddle River, New Jersey, USA, 2003.
- MICROSOFT (2006), Microsoft Research, *Data Mining: Efficient Data Exploration and Modeling*, <http://research.microsoft.com/dmx/DataMining>, (12.05.2006).
- OLARU, C. ve WEHENKEL, L., “Data Mining”, *IEEE Computer Applications in Power*, 12(3), (1999), 19-25.
- ORHUNBILGE, N., *Zaman Serileri Analizi Tahmin ve Fiyat İndeksleri*, Avcıol Basım Yayın, İstanbul, 1999.
- PHAM, D. T., PACKIANATHER, M. S. ve CHARLES, E. Y. A., “A Novel Self-Organised Learning Model with Temporal Coding for Spiking Neural Networks”, (in Eds. PHAM, D. T., ELDUKHRI, E. E. ve SOROKA, A. J.) *Intelligent Production Machines and Systems*, Cardiff University, Manufacturing Engineering Centre, Cardiff, UK., (2006), 307-312.
- PRYBUTOK, V. R., YI, J. ve MITCHELL, D., “Comparison of Neural Network Models with ARIMA and Regression Models for Prediction of Houston’s Daily Maximum Ozone Concentrations”, *European Journal of Operational Research*, 122, (2000), 31-40.
- SHMUELI, G, PATEL, N. R. ve BRUCE, P. C., *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*, John Wiley & Sons, Hoboken, NJ, USA, 2007.
- SPSS (2007a), *Clementine11.1 User’s Guide*, Integral Solutions Limited, Chicago, IL., 2007.
- SPSS (2007b), *Clementine11.1 Node Reference*, Integral Solutions Limited, Chicago, IL, 2007.
- TAN, P.-N., STEINBACH, M. ve KUMAR, V., *Introduction to Data Mining*, Pearson, Addison-Wesley, Boston, MA, USA, 2006.
- TANG, Z. H. ve MACLENNAN, J., *Data Mining with SQL Server 2005*, Wiley Publishing Inc., Indianapolis, IN, USA, 2005.
- TSAI, C. Y. ve TSAI, M. H., “A Dynamic Web Service based Data Mining Process System”, *Proceedings of The Fifth International Conference on Computer and Information Technology (CIT’05)*, Washington, DC, USA, IEEE Computer Society, (2005), 1033-1039.
- TSENG, F. M., YU, H. C. ve TZENG, G. H., “Combining Neural Network Model with Seasonal Time Series ARIMA Model”, *Technological Forecasting & Social Change*, 69, (2002), 71-87.
- TSETSEKAS, C. A., FERTIS, A. G. ve VENIERIS, I. S., “Dynamic Application Profiles using Neural Networks for Adaptive Quality of Service Support in the Internet”, *Computer Communications*, 29, (2006), 2985-2995.
- WANG, L. ve FU, X., *Data Mining with Computational Intelligence*, Springer-Verlag Berlin Heidelberg, Germany, 2005.
- WINTERS, P. R., “Forecasting Sales by Exponentially Weighted Moving Averages”, *Management Science*, 6, (1960), 324-342.
- ZHANG, G. P., *Neural Networks in Business Forecasting*, Idea Group Publishing, Hershey, PA, 2004.
- ZOU, H. F., XIA, G. P., YANG, F. T. ve Wang, H. Y., “An Investigation and Comparison of Artificial Neural Network and Time Series Models for Chinese Food Grain Price Forecasting”, *Neurocomputing*, 70, (2007), 2913-2923.
- ZUBCOFF, J., PARDILLO, J. ve TRUJILLO, J., “A UML Profile for the Conceptual Modeling of Data-Mining with Time-Series in Data Warehouse”, *Information and Software Technology*, 51(6), (2009), 977-992.