

Manisa Pamuk Fiyatlarının Zaman Serisi Analizi Ve Yapay Sinir Ağı Teknikleri İle Tahminlenmesi Ve Tahmin Performanslarının Karşılaştırılması

Şengül CAN* Mustafa GERŞİL**

ÖZ

Tekstil sektörünün bitkisel hammaddesi olarak kullanılan pamuk gerek ülkemiz gerekse dünyada tarım, sanayi ve ticarete stratejik bir önem taşımaktadır. Ülkeler arası pamuk ticareti ise pamuk borsaları aracılığıyla gerçekleşmektedir. Ülkemizdeki birçok tarımsal ürünün değerini devlet belirler. Ancak pamuk borsalar tarafından fiyatı belirlenen birkaç üründen biridir. Bu çalışma kapsamında Manisa Tarım Borsasından yıllık pamuk fiyatları alınmıştır. Alınan veriler analize uygun olacak şekilde düzenlenmiş ve weka programında analiz edilmiştir. Zaman serisi ve yapay sinir ağı teknikleri kullanılarak 2017 yılına ait fiyatların değer tahminleri yapılmıştır. Yapılan tahminler MAE, MAPE ve RMSE değerleri üzerinden karşılaştırılarak hangi tekniğin daha başarılı tahmin performansı gösterdiğine karar verilmiştir. Elde edilen sonuçlar literatürdeki benzer çalışma sonuçlarıyla da karşılaştırılmıştır. Tüm bu sonuçlar doğrultusunda yapay sinir ağı tekniğinin daha başarılı sonuçlar elde ettiği görülmüştür.

Anahtar Kelimeler: Veri Madenciliği, Zaman Serisi, Yapay Sinir Ağı, Tahmin Performanslarının Karşılaştırılması

Jel Sınıflandırması: C45, C53

Estimating of Manisa Cotton Prices Using Time Series and Artificial Neural Network Techniques and Comparison of Their Estimating Performance

ABSTRACT

Cotton used as herba lraw material of textile industry is of strategic importance in our country and also in agriculture, industry and trade in the world. Cotton trade between countries takes place through cotton stock exchanges. The state determines the value of many agricultural products in our country. However, cotton is one of the few products whose prices are determined by the stock exchange. In this study, annual cotton prices were taken from Manisa Agricultural Market. Later, these data was first corrected, and analyzed in the weka program. Time series and artificial neural network techniques were used to estimate the value of prices for 2017. Based on the estimated values by MAE, MAPE and RMSE, it is decided which technique will yield better prediction performance. The results obtained were also compared with similar studies in the literature. All of these results show that the artificial neural network technique achieved more successful results.

Keywords: Data Mining, Time Series, Artificial Neural Network, Comparing Forecast Performance

Jel Classification: C45, C53

*Öğr. Gör. Manisa Celal Bayar Üniversitesi Araştırma Girişimcilik ve Yenilikçilik Koordinatörlüğü, sengul.can@cbu.edu.tr

**Doç. Dr. Manisa Celal Bayar Üniversitesi İ.İ.B.F. İşletme Bölümü, mustafa.gersil@cbu.edu.tr

(Makale Gönderim Tarihi: 06.09.2018 / Yayına Kabul Tarihi: 28.11.2018)

Doi Number: 10.18657/yonveek.457761

GİRİŞ

Günümüzde giderek yaygınlaşan veri madenciliği; iş dünyasındaki işletmelerin karlarını arttırabilmeleri amacıyla pek çok farklı alanda kullanılmaktadır. Kökleri yapay zekâ, makine öğrenimi, matematik, istatistik gibi pek çok farklı disipline dayanan veri madenciliği; karar destek sistemleri ve stratejik planlama çalışmalarında kullanılabilir (Irmak vd., 2012).

Tarım ve tarım ürünlerinin fiyatları da büyük miktarda veri saklanan ve veri madenciliği tekniklerinden elde edilen faydalı bilgiden yararlanabilecek alanlar içerisindedir. Örneğin dünya nüfusundaki hızlı artışın yanı sıra sanayileşme ve kalkınma toplumların yaşam standardını yükseltmiş ve pamuğa duyulan ihtiyacı da arttırmıştır. Tekstil sektörünün bitkisel hammaddesi olan pamuk farklı kullanım alanlarında gerek ülkemiz gerekse dünyada stratejik bir önem taşımaktadır (Aktaş, 2006).

Uluslararası Pamuk Danışmanları Komitesi (International Cotton Advisory Committee) kaynaklarına göre 2015-2016 üretim yılındaki hızlı düşüşe rağmen 21.48 milyon ton olan toplam dünya üretimi 2016-2017 üretim yılında %7 artmış ve 22.99 milyon ton seviyesine ulaşmıştır. 2017-2018 yılı için toplam üretim %11 artarak 25.43 milyon tonluk bir üretim hacmi öngörülmektedir (Manisa Ticaret Borsası Pamuk Raporu, 2017).

Büyük miktarlardaki pamuk üretiminin uluslararası arenadaki ticareti ise pamuk borsaları vasıtasıyla gerçekleştirilmektedir. Ticaret borsalarının kuruluş amaçları diğer tarım ürünlerinde de olduğu gibi fiyat belirlenmesi ve piyasa koşullarına uyarlanmasıdır. Özellikle borsalarda faaliyet gösteren tüccarların talepleri ve talep sonucu elde edilen veriyi üreticilerle düzenli olarak paylaşabilmeleri fiyat oluşumundaki etkinliğin sağlanmasında oldukça önemlidir. “Etkin Piyasa” kavramı; bilgilerin üreticiye tam olarak yansıtıldığı ve üreticinin de gelen talepler doğrultusunda istenilen kalite ve nitelikte üretim gerçekleştirebileceği yapılar olarak tanımlanmaktadır. Bu varsayım ile borsaların işlem maliyeti bulunmaması, elde edilen bilginin tüm paydaşlara maliyetsiz olarak verilmesi, mevcut ve gelecekteki fiyat dağılımı için cari bilgi etkisinin piyasadaki tüm katılımcılar tarafından kabul görmesi ile uzun vadede hiçbir ticari paydaşın ortalama üzerinde gelir elde etme ihtimali olmayacaktır. Böylece ürün fiyatları için piyasalarda etkin bir işleyiş olabilecektir (Mananyi ve Struthers, 1997).

Türkiye’de pek çok tarım ürünüde değişim değeri devlet tarafından belirlenen alım fiyatı ile geçmektedir. Ancak; pamuk, kuru meyve, fındık ve bakliyatla fiyat piyasa kuralları gereği borsalarda belirlenmektedir. Küresel ekonomide temel pamuk borsaları fiyatların belirlenmesi açısından önemli mesafe katetmiş olsa da fiyatların belirlenmesinde borsalar arası ilişkiler de önemlidir (Telatar vd., 2002).

Bu çalışma kapsamında önemli bir tarım ürünümüz olan pamuk fiyatlarının iki farklı veri madenciliği tekniği kullanılarak tahminlenmesi ve tahmin performanslarının karşılaştırılması amaçlanmaktadır. Çalışma dört bölüme ayrılmıştır. İlk bölüm Türkiye’de pamuk üretimi ile ilgili genel bilgiler verilmektedir. İkinci bölümde tahmin için kullanılan teknikler anlatılmaktadır.

Üçüncü bölümde verilerin tahmin edilerek tercih edilen yöntemlerin tahmin performansları karşılaştırılmıştır. Son bölümde ise çalışmanın sonuç ve değerlendirmeleri yer almaktadır.

I. TÜRKİYE'DE PAMUK ÜRETİMİ

Pamuk ekim alanları 2016/2017 yıllarında 415 hektar ile son beş yılın en düşük seviyesine gerilemiş olsa da bu sezonda 455 bin hektara çıkmıştır (Manisa Ticaret Borsası Pamuk Raporu, 2017).

Tablo 1: Türkiye Pamuk Üretimi

	2014/ 2015 Sezonu	2015/ 2016 Sezonu	2016/ 2017 Sezonu	2017/ 2018 Sezonu
Ekim Alanı(Hektar)	468.000	434.000	415.000	455.000
Kütlü Üretimi (Ton)	2.350.000	2.050.000	2.100.000	2.320.000
Kütlü Verimi (Kg/Dekar)	502	472	505	510
Lif Üretimi (Ton)	846.000	738.000	756.000	882.000
Lif Verimi (Kg/Dekar)	181	170	182	194
Çırcır Randımanı (%)	36	36	36	38

Kaynak: <https://www.calameo.com/books/0037407384aa93b9c1bfb>

Pamuk yaklaşık olarak 50 farklı sanayi dalı için hammadde üreten bir endüstri bitkisidir. Pamuk sektöründe bulunan arz açığı nedeniyle Türkiye dünyadaki önemli pamuk ithalatçısı ülkelerden biri konumundadır (Semerci ve Çelik, 2018).

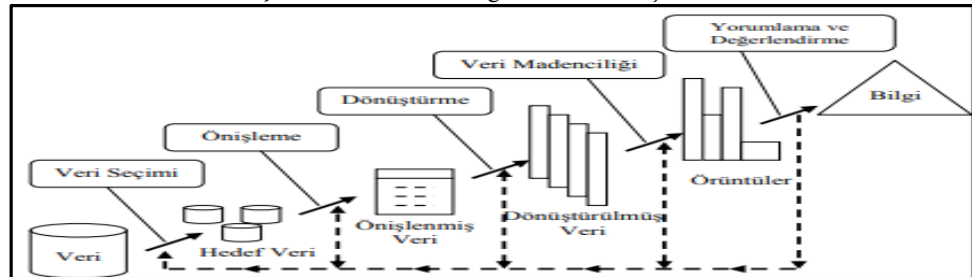
Türkiye'de üretim alanlarının %62'si Güneydoğu Anadolu bölgesinde, %20'si Akdeniz bölgesinde ve %18'i Marmara ve Ege bölgelerinde bulunmaktadır. Bölgelerin pamuk üretim miktarları ise sırasıyla; %58, %22 ve %18 oranındadır (TUİK, 2017).

II. VERİ MADENCİLİĞİ VE ÇALIŞMA İÇİN KULLANILAN TEKNİKLER

Veri madenciliği tekniğini daha derinlemesine anlayabilmek için gizli kalmış değerlerin yeryüzüne çıkarılması olarak tanımlanan madencilik kelimesini anlamak gerekmektedir. Bu kelimenin veri kelimesi ile ilişkilendirilmesi ise; büyük veri yığınları içerisinde gizli kalmış ve faydalı olabilecek bilginin çıkarılması şeklinde olmaktadır (Giudici, 2003).

Veri madenciliği sadece bir yöntem ve teknikler topluluğu değil aynı zamanda probleme özgü yöntem ve teknikleri barındıran ve sonuçları dolayısıyla kural, örüntü, eğilimleri modelleyip gösteren bir süreç olarak tanımlanmaktadır (Kantardzic, 2003).

Şekil 1: Veri Madenciliği Sürecindeki Aşamalar



Kaynak: Fayyad vd. 1996

Şekil 1’de veri madenciliği sürecinin literatürde yer alan en yaygın gösterimi bulunmaktadır.

A.Zaman Serisi Analizi

Araştırma konusu olan büyüklüğün zaman içindeki sıralanmış ölçümlerinin bir kümesi zaman serisi olarak isimlendirilir. Zaman serisi ile ilgili gözlem yapılmasının nedeni ise; gözlem yapılan dönemdeki temsil edilen gerçek bilginin anlaşılması ve zaman serisi değişkenlerinin gelecek değerlerinin doğru tahmin edilmesidir (Allen, 1964). Zaman serilerini oluşturan dört farklı unsur bulunmaktadır (Newbold, 2000):

- Genel Eğilim (Trend) Bileşeni: Gözlemi yapılan zaman serisinde uzun vadede görülen düşme veya yükselme süreçleri sonrasındaki kararlı durumdur. Zaman serileri uzun dönemde kararlı bir yükseliş veya alçalış eğilimine sahiptir.
- Mevsim Bileşeni: Gözlemi yapılan zaman serisindeki mevsimsel değişimi ifade eder. Zaman serilerinde kullanılan verilerin bazıları diğer dönemlere göre farklılık gösterebilir.
- Çevrim Bileşeni: Ekonomide görülen ancak mevsimler olmayan dönemlik değişimlerdir. Ekonomideki genel eğilimden bağımsız kısa vadeli küçülme veya büyümeler örnek olarak verilebilir.
- Düzensiz Bileşen: Belirli olmayan ve hata terimi ile ifade edilebilecek değişimlerdir.

Zaman serisi analiz tekniği ile gelecekte gerçekleşmesi olası olaylar tespit edilebilir. Özellikle borsa analizi gibi finansal sistemlerde stratejik öneme sahip zaman serileri analizi ile ileriye dönük tahminler yapılabilmektedir. Kullanılan zaman serisi analiz teknikleri ile veri madenciliğinde veri kümelerinin özellikleri çıkarılabilmektedir. Yani verilerin zamana bağlı hareketleri tespit edilerek gerçekleştirilen tahminlerin daha tutarlı ve doğru olması sağlanabilmektedir (Şeker, 2015).Zaman serilerinin kullanım amaçları:

- Aykırı (Outlier) Verileri Yakalama: Elde bulunan veriler zamana göre bir sıraya yerleştirildiğinde serinin uzağında bulunurlar, aykırı değerlerin yakalanması için serinin hareketinin yakalanması gerekmektedir (Barnett, 1994). Aykırı verileri yakalamak; elde bulunan veriler üzerinde zaman serisi analizi yaparak veri hareketlerinde zaman içerisinde farklılıkları yakalamaktır. Değeri yanlış girilmiş veriler, sistem içerisinde hata sonucu oluşmuş veriler veya belli bir olaya bağlı olarak elde edilen veriler olabilir. Aykırı verilerin yakalanması; bu verilerin oluşmasına neden olan olayın anlaşılması ve müdahale edilebilmesi açısından önemlidir (Şeker vd., 2013).
- Tahmin (Prediction): Zaman serisi analizi ile ileride gerçekleşmesi muhtemel olaylar belirlenebilir. Zaman serisi analizi ile ileriye dönük tahmin işlemleri de gerçekleştirilebilmektedir. Özellikle finansal veriler üzerinde ve stratejik yönetim karar aşamalarında oldukça önemlidir (Şeker vd., 2014).
- Eksik Verilerin Tamamlanması (Imputation): Zaman serisi analizi kullanılarak eksik verilerin yerine hangi değerlerin koyulacağı bulunabilmektedir (Hanoker, 2010).

- Hata Düzeltme (Data Scrubbing): Veri seti içerisinde bulunan aykırı değerlerin belirli teknikler ile diğer değerlere yaklaştırılması işlemleri için zaman serisi analizi kullanılabilir (Crosswhite, 2003).

Zaman Serisi Analiz Yöntemleri

- Hareketli Ortalama: Anlık, beklenmedik, hatalı veya aykırı verilerin genel veri seti içerisindeki zamana bağlı hareketleri üzerindeki etkilerini azaltarak düz bir çizgide olmalarını sağlamaktadır.

Basit Hareketli Ortalama:

$$HO(SMA) = \frac{P_m + P_{m-1} + \dots + P_{m-(n-1)}}{n}$$

Kümülatif Hareketli Ortalama:

$$KHO(CMA) = \frac{P_1 + P_2 + \dots + P_n}{n}$$

Ağırlıklı Hareketli Ortalama:

$$AHO(WMA) = \frac{nP_m + (n-1)P_{m-1} + \dots + 2P_{m-n+2} + P_{m-n+1}}{n + (n-1) + \dots + 2 + 1}$$

Üstel Hareketli Ortalama:

$$EMA_{Bugün(today)} = \alpha * (P_1 + (1-\alpha)P_2 + (1-\alpha)^2 P_3 + (1-\alpha)^3 P_4 + \dots$$

Basit Hareketli Orta Değer: Veri setinin orta değeri alınır.

Açıklanan bu zaman serisi analiz teknikleri veri madenciliğinde veri setinin niteliklerini çıkarmayı sağlamaktadır. Yani veri setinin zamana bağlı hareketleri bulunmaktadır. Bu şekilde gerçekleştirilen tahminler daha tutarlı ve doğru olmaktadır. Veri seti üzerinde yapılan zaman serisi analizi ile veriler arasında bulunan ilişkiler tespit edilmektedir (Şeker, 2015). Anlatılan zaman serisi analizlerinin yanısıra daha çok ekonometrik analizlerde kullanılan diğer modeller (Sevüktekin ve Çınar, 2014):

- Fark Denklemleri İle Zaman Serisi Analizi: Matematik ve zaman serisi analizlerinin temelini fark denklemleri oluşturmaktadır. Zaman serisi analizi olasılıklı bileşenler içeren fark denklemlerinin tahminleriyle de ilgilenmektedir. Stokastik fark denklemleri ise dinamik ekonometrik modellerde doğal bir şekilde oluşmaktadır. Yani tahmin edilen denklemler ekonomik verilerin yorumlarında ve hipotez testlerinde uygun şekilde kullanılabilir.
- Doğrusal Zaman Serisi Modelleri: Zaman serisi analizlerinde özellikle kısa dönem bilgilerinin elde edilmesi aşamasında farklı teknikler kullanılabilir. Örneğin tek değişkenli bir zaman serisinin kendi geçmiş değer ve hatalarına göre kurulan modelleri

içeren; otoregresif modeller (AR(q)), hareketli ortalamalar (MA(q)), otoregresif hareketli ortalamalar (ARMA(p,q)), homojen durağan dışı modeller (ARIMA(p,d,q)), box-jenkins yaklaşımı bulunmaktadır.

Otoregresif Model AR(q):

$$x_t = a + a_1 * x_{(t-1)} + a_2 * x_{(t-2)} + a_3 * x_{(t-3)} + \dots + \varepsilon$$

Hareketli Ortalama MA(q):

$$x_t = e_t - a_1 e_{t-1} - \dots, t = 1, 2, \dots n$$

ARMA(p,q) ve ARIMA(p,d,q):

$$Y_t = m + a_1 y_{t-1} + \dots + a_p y_{t-p} + u_t - b_1 u_{t-1} - \dots - b_q u_{t-q}$$

Box-Jenkins Yaklaşımı: Zaman serisinin sadece kendi geçmiş değerleri ve olasılıksal hata terimi ile açıklanması bu yöntemin temel vurgusudur. Box-jenkins yöntemi: Özdeşleşme (seriye ait p,d,q değerlerinin bulunması), Tahmin (verilerin belirlenen modele uygunlaştırılması), Tanısal Denetim (verilerin modele yeteri kadar uygun olup olmadığı incelenir ve gerekirse başa dönülerek yeni değerler belirlenir), Yordama (belirlenen model ile serinin değerlerinin tahminlenmesi).

- Vektör Otoregresif Modeller: Gerçekte pek çok zaman serisinin veri üretme süreci kendi geçmiş hareketlerinin yanı sıra diğer serilerin hareketlerine göre de değişmektedir. Bu durumda tek değişkenli analizler yerine Vektör Otoregresif VAR(p) gibi çok değişkenli analizler kullanılmaktadır.

B.Yapay Sinir Ağı Tekniği

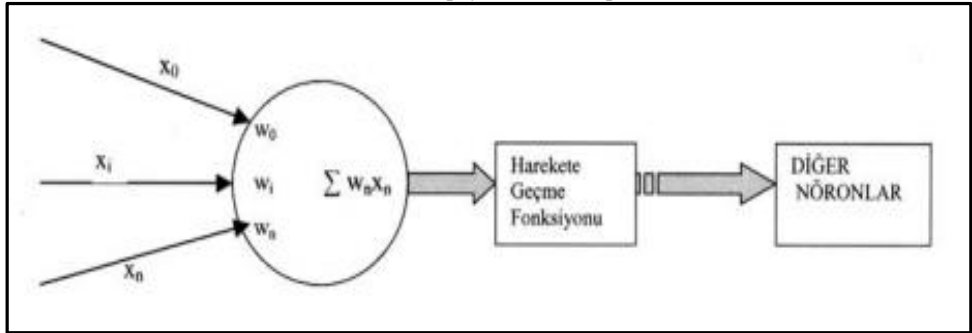
Yapay sinir ağları insan beynindeki temel öğrenme fonksiyonunu yerine getiren bir bilgisayar yazılımıdır. Örnekler kullanılarak öğrenme işlemi gerçekleştirilir. Bilinen hesaplama tekniklerinden farklı bir yöntem öneren yapay sinir ağı; bulunduğu ortama uyum sağlayabilen, adapte olabilen, eksik veri ile çalışabilen, belirsizlik durumunda karar verebilen, hata toleransı bulunan bir sistem olması nedeniyle hayatın hemen her alanında başarılı şekilde uygulanabilmektedir. Özellikle sınıflandırma, sinyal filtreleme, örüntü tanımlama, veri sıkıştırma ve optimizasyon konularında yapay sinir ağları güçlü teknikler içerisindedir (Öztemel, 2012).

Biyolojik sinir sisteminden etkilenerek geliştirilen yapay sinir ağları birbirleri arasında sinir hücreleriyle iletişim kurmaktadır. Girdi, ara ve çıktı katmanları bir araya gelerek bir ağ oluştururlar. Bilgiler öncelikle girdi katmanına iletilir. Daha sonra ara katmanda işlenerek çıktı katmanına iletilir. Bilgi işleme ile anlatılmak istenen ağa giriş yapan verinin ağdaki ağırlık değerleri kullanılarak çıktıya dönüştürülmesidir. Ağ içerisindeki girdilerden doğru çıktılar elde edebilmek için doğru ağırlıkların bulunması yani ağın eğitilmesi gerekmektedir. Eğitimi işleminin tamamlanmasının ardından test veri seti ağa gösterilir ve eğer test veri setine doğru

cevaplar alınırsa ağ eğitilmiş kabul edilir (Öztemel, 2012). Yapay sinir ağı modellerindeki temel varsayımlar (Fausett, 1994):

- Yapay sinir ağı sistemine gelen bilgi “Nöron” olarak isimlendirilen birimler aracılığıyla işlenir.
- Gelen sinyallerin iletimi nöronlar arasında yer alan bağlantılar aracılığıyla gerçekleştirilir.
- Her bir bağlantının belli ağırlığı vardır. Pek çok YSA’da bağlantı ağırlıkları iletilen sinyaller ile çarpılır.
- Nöronlar aracılığıyla gelen sinyallerin ağırlıklandırılmış toplamları net girdiyi oluştururken, nöron çıktı sinyalini belirlemek için gelen sinyallere genellikle doğrusal olmayan bir fonksiyon olan harekete geçme fonksiyonu uygulanır.

Şekil 2: Yapay Nöronun Yapısı



Kaynak: Diler, 2003.

Yapay sinir ağını oluşturan nöronlar yapısal olarak biyolojik nöronlar ile aynı niteliklere sahiptir. Çok sayıda nöronun bir araya gelmesiyle oluşan yapay sinir ağlarında her nöron bir diğeri ile bağlıdır. Tüm bu bağlantıların öğrenme süreci ile belirlenen ağırlıkları bulunmaktadır. Bu doğrultuda yapay sinir ağları aşağıdaki nitelikler ile karakterize edilmektedir (Diler, 2003):

- Yapı (Architecture): Genellikle tabakalar halinde yerleştirilen nöronların davranışlarını belirleyen ana faktörlerden; harekete geçme ve sinyal alıp-gönderen ağırlıklandırılmış bağlantı dokularıdır. Tabaka sayısına göre de ayrılabilen yapay sinir ağları literatürde sıklıkla ileri beslemeli ve geri beslemeli olarak ikiye ayrılmaktadır.
- Bağlantı ağırlıklarının belirlenmesi: Bağlantı ağırlıklarının belirlenmesi de yapının belirlenmesi kadar önemli ve yapay sinir ağlarını birbirlerinden ayıran önemli bir özelliktir. Buradaki temel iki öğrenme tekniği ise; şartlandırılmış öğrenme ve şartlandırılmamış öğrenmedir. Belirli bir girdi vektörüne karşılık belirli bir çıktı vektörünün olduğu durum şartlandırılmış öğrenme, girdi vektörüne karşılık çıktı olarak belirli bir hedefin olmadığı durum ise şartlandırılmamış öğrenme olarak tanımlanmaktadır.
- Harekete geçme fonksiyonu: Yapay nöronun temel bir işlevi de önce girdi sinyallerini toplamak daha sonra dönüştürerek bir çıktı oluşturmaktır. Nöronlar ağırlıklandırılarak topladığı sinyalleri bir fonksiyon yardımıyla dönüştürür. Dönüştürme işlemi sonrasında elde edilen değer eğer eşik değerden yüksek ise

nöron bir çıktı sinyali üretmektedir. Genel olarak doğrusal olmayan harekete geçme fonksiyonları, problem türüne göre değişiklik gösterebilir.

III.PAMUK ÜRETİM MİKTARLARININ WEKA PROGRAMINDA TAHMİNLENMESİ

2018 yılındaki değerleri tahmin edilecek veriler Manisa Ticaret Borsası'ndan alınmıştır. Manisa Ticaret Borsası'ndan alınan Şekil 2' de görülen örnek veriler içerisinde fiyat bilgisi bulunmayan günler ("Muamele Yoktur" ve "İsmin" yazan satırlar) veri içerisinde temizlenmiş ve analize uygun hale getirilmiştir.

Şekil 2:2017 Yılı Pamuk Verilerinden Örnekler

Tarih	Tür	EnAz	EnÇok
20.04.2017	ÇİĞİTLİ PAMUK	3,10 TL	3,10 TL
21.04.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,10 TL	3,10 TL
24.04.2017	ÇİĞİTLİ PAMUK	3,14 TL	3,14 TL
25.04.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,14 TL	3,14 TL
26.04.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,14 TL	3,14 TL
27.04.2017	ÇİĞİTLİ PAMUK ST-1	3,20 TL	3,20 TL
28.04.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,20 TL	3,20 TL
2.05.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,20 TL	3,20 TL
3.05.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,16 TL (İS	3,16 TL (İSMEN)
4.05.2017	ÇİĞİTLİ PAMUK	3,10 TL	3,10 TL
5.05.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,10 TL	3,10 TL
8.05.2017	ÇİĞİTLİ PAMUK ST-1	3,14 TL	3,14 TL
9.05.2017	ÇİĞİTLİ PAMUK ST-1	3,14 TL	3,14 TL
10.05.2017	ÇİĞİTLİ PAMUK	3,14 TL	3,14 TL
11.05.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,14 TL	3,14 TL
12.05.2017	ÇİĞİTLİ PAMUK ST-1 (MUAMELE YOKTUR)	3,14 TL	3,14 TL

Veri seti içerisinde dikkate alınmaması gereken satırlar temizlendikten sonra 135 günlük fiyat bilgisi elde edilmiştir. Temizlenmiş veri setinden örnekler Şekil 3'de gösterilmektedir.

Şekil 3:Temizlenmiş Veri Setinden Örnekler

Tarih	Tür	EnAz	EnÇok
2.01.2017	ÇİĞİTLİ PAMUK	2,36	2,36
3.01.2017	ÇİĞİTLİ PAMUK	2,44	2,44
4.01.2017	ÇİĞİTLİ PAMUK	2,48	2,48
5.01.2017	ÇİĞİTLİ PAMUK	2,52	2,52
6.01.2017	ÇİĞİTLİ PAMUK	2,52	2,52
9.01.2017	ÇİĞİTLİ PAMUK	2,52	2,52
10.01.2017	ÇİĞİTLİ PAMUK	2,68	2,68
11.01.2017	ÇİĞİTLİ PAMUK	2,72	2,72
12.01.2017	ÇİĞİTLİ PAMUK	2,72	2,72
13.01.2017	ÇİĞİTLİ PAMUK	2,72	2,72
16.01.2017	ÇİĞİTLİ PAMUK	2,64	2,64
17.01.2017	ÇİĞİTLİ PAMUK	2,64	2,64
18.01.2017	ÇİĞİTLİ PAMUK	2,66	2,66
19.01.2017	ÇİĞİTLİ PAMUK	2,66	2,66
20.01.2017	ÇİĞİTLİ PAMUK	2,66	2,66

2018 yılına ait veriler Weka programı kullanılarak tahmin edilecektir. Weka programı java yazılım dili kullanılarak Waikato Üniversitesi'nde geliştirilmiş bir veri analiz aracıdır ve aynı zamanda veri madenciliği alanında da yaygın olarak kullanılmaktadır. Weka programı üzerinde makine öğrenimi ve istatistik ile ilgili pek çok bilgi hazır olarak gelmektedir. Programın 3.7 versiyonu ve takip eden yeni versiyonlarında; desteklenen tahmin modellerinin geliştirilmesi, değerlendirilmesi ve görsel olarak sunulmasına imkân tanıyan özel bir zaman serisi analiz ortamı da sunulmaktadır. Weka zaman serisi analiz platformu, verileri standart bir veri madenciliği algoritmasının işleyebileceği bir formata dönüştürerek zaman serisini modelleyebilmek için makine öğrenimi/veri madenciliği metodolojisini benimsemektedir. Trend ve mevsimsellik gibi nitelikler algoritma tarafından otomatik olarak hesaplanır. Veri dönüşümü tamamlandıktan sonra weka programına ait algoritma eğitim aşamasını tamamlar. Zaman serisi analizi ve tahmininde bu teknik; Arma/Arıma gibi klasik istatistiksel tekniklerden çoğu zaman daha güçlü ve daha esnektir (Hall, 2014). Kümeleme, sınıflandırma ve ilişkilendirme yöntemlerini kullanan programın desteklediği dosya formatı ise .arff'dir (Veysel, 2016). Bu nedenle elimizdeki veriler Şekil 4'te görüldüğü gibi .arff dosya formatına uygun hale getirilmiştir.

Şekil 4: Verilerin .arff Formatına Dönüştürülmesi

```
@RELATION kpamuk

@attribute NUM NUMERIC
@attribute MIN NUMERIC
@attribute MAX NUMERIC

@DATA
1,2.36,2.36
2,2.44,2.44
3,2.48,2.48
4,2.52,2.52
5,2.52,2.52
6,2.52,2.52
```

Analiz için uygun hale getirilen veriler Weka programının forecast özelliği kullanılarak tahmin edilmiştir. Weka forecast özelliği programın gelişmiş özelliklerinden biridir. Manisa ticaret borsası pamuğun en az ve en çok değeri için tek fiyat bilgisi vermiştir. Bu nedenle zaman serisi ve yapay sinir ağı teknikleri ile tek bir değer için tahminleme işlemi gerçekleştirilmiştir.

Tahmin etmeye ilişkin iki önemli adım bulunmaktadır; ilki veriyi tahmin etmek için hazırlama ikincisi ise farklı tahmin edici modellerin karşılaştırılmasıdır. Modelleri karşılaştırma ölçütleri; doğruluk, hız, sağlamlık, ölçeklenebilirlik, yorumlanabilirliktir. MeanError (ME) – Ortalama Hata, MeanAbsoluteError (MAE) – Ortalama Mutlak Hata, MeanSquareError (MSE) – Ortalama Kare Hata, RootMeanSquareError (RMSE) – Kök Ortalama Kare Hatası, MeanAbsolutePercentageError (MAPE) – Ortalama Mutlak Hata Yüzdesi değerleri performans karşılaştırması için kullanılmaktadır (SeigalandMehrotra, 2012: 57). Bu çalışmada modellerin karşılaştırılmasında, RMSE, MAE ve MAPE değerleri kullanılacaktır. RMSE, hata kareleri ortalamasının kareköküdür. Bağımlı serinin, kullanılan model tarafından tahmin edilenden farkını ifade etmek için

kullanılmaktadır. Küçük değerlere sahip olması model tahmininin daha iyi olduğunu göstermektedir. MAPE, farklı serilerin karşılaştırılmasında da kullanılan bir değerdir ve ortalama mutlak yüzde hatayı temsil etmektedir. MAE, mutlak ortalama hatayı temsil etmektedir (Irmak vd. 2012).

C.Weka Programında Zaman Serisi Tekniği İle Tahminleme

Tahminlemeye uygun hale getirilen veriler üzerinden 15 günlük veri tahmini yapılmıştır. Periyot günlük olarak seçilmiş ve %95 güven aralığında verinin tamamı eğitim verisi olarak kullanılarak tahminleme işlemi gerçekleştirilmiştir. Gerçekleştirilen 15 günlük tahmin değerleri Tablo 2 ile gösterilmiştir.

Tablo 2: Zaman Serisi Tekniği İle Tahmin Değerleri

Tahmin Edilen Gün	Tahmin Değeri
1	2.6956
2	2.6933
3	2.6913
4	2.6959
5	2.6929
6	2.6906
7	2.6888
8	2.6876
9	2.776
10	2.7852
11	2.8727
12	2.8944
13	2.6744
14	2.6721
15	2.6701

D.Weka Programında Yapay Sinir Ağı Tekniği İle Tahminleme

Tahminlemeye uygun hale getirilen veriler üzerinden literatürde yaygın olarak kullanılan %80 eğitim %20 test oranında ağı eğitimi ve testi gerçekleştirilmiştir. Gerçekleştirilen tahmin sonuçlarının 15 günlük değerleri Tablo 3 ile gösterilmiştir.

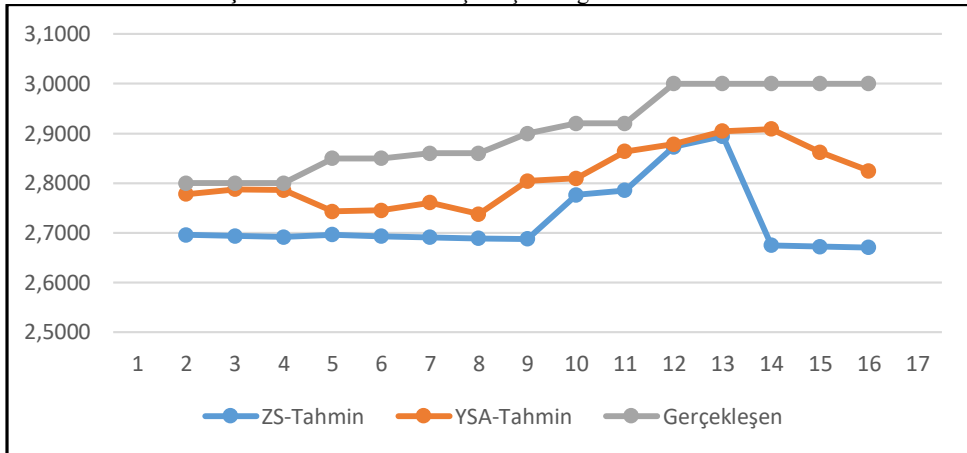
Tablo 3: Yapay Sinir Ağı Tekniği İle Tahmin Değerleri

Tahmin Edilen Gün	Tahmin Değeri
1	2.7776
2	2.7873
3	2.7861
4	2.743
5	2.7448
6	2.7608
7	2.7376
8	2.8042
9	2.8097
10	2.8638
11	2.8781
12	2.9047
13	2.9087
14	2.862
15	2.8246

Tablo 4: 2018 Yılında İlk 15 Gün Gerçekleşen Değerler

Gerçekleşen Gün	Gerçekleşen Değer
02.01.2018	2.80
03.01.2018	2.80
04.01.2018	2.80
05.01.2018	2.85
08.01.2018	2.85
09.01.2018	2.86
10.01.2018	2.86
11.01.2018	2.90
15.01.2018	2.92
16.01.2018	2.92
17.01.2018	3.00
18.01.2018	3.00
19.01.2018	3.00
22.01.2018	3.00
23.01.2018	3.00

Tahmin ve gerçekleşen değerler grafiksel olarak Şekil 5 ile gösterilmektedir.

Şekil 5: Tahmin Ve Gerçekleşen Değerlerin Grafiksel Gösterimi

Tahmin sonuçlarının karşılaştırılmasında kullanılan farklı performans ölçüm değerleri bulunmaktadır. MAE (MeanAbsoluteError- Mutlak Ortalama Hata) orijinal veriler ile aynı birimler halinde ölçülür ve genel olarak kök kareler ortalamasına göre büyüklük açısından benzerdir. MAE ile hatalar daha kolay ifade edilme yeteneğine sahiptir. RMSE (RootMeanSquareError - Kök Ortalama Kare Hatası) bir model tarafından öngörülen değerler ile gerçekte gözlemlenen değerler arasındaki farkın sıklıkla kullanılan bir ölçüsüdür. MAPE (MeanAbsolutePercentageError – Ortalama Mutlak Hata) trend tahmininde, bir tahmin yönteminin tahmin doğruluğunun bir ölçüsüdür. Dolayısıyla karşılaştırmalı modellerin performansında çoğunlukla RMSE, MAE veya MAPE' ye bağlı olduğu belirtilmektedir. Ayrıca, eğer iki modelin hata istatistikleri benzerse, daha basit ve / veya anlaşılması daha kolay olan tercih etmelidir. MAPE oranı 0' a ne kadar yakınsa tahmin değerleri de o kadar başarılı kabul edilir (Saigal ve Mehrotra, 2012). Tablo 5 ' de maksimum ve minimum değerlerin tahmini için elde edilen performans değerlendirme sonuçları bulunmaktadır.

Tablo 5: Model Performanslarının Karşılaştırılması

Model	MAE	RMSE	MAPE
Zaman Serisi Tekniği	0.3558	0.4654	13.2794
Yapay Sinir Ağı Tekniği	0.1052	0.1715	3.1117

SONUÇ

Günümüzde veri madenciliği önemi ve yaygınlığı giderek artan bir teknoloji olmaktadır. Özellikle büyük miktardaki veri yığınlarından üretilen stratejik bilgiler nedeniyle gelecekte de yaygın olarak kullanılabilceği öngörülmektedir. Tahmin edici veri madenciliği teknikleri ise bir model oluşturarak veri kümelerinin davranışlarını tahmin etmeye çalışır. Tahmin önemli modeller çıkartmak için veya gelecekteki trendleri tahmin etmek için kullanılabilen bir veri analiz biçimidir. Bu tür analizler hem mevcut veri setini daha iyi anlamamıza hem de veri setinin gelecekteki davranışını tahmin etmede sıklıkla kullanılmaktadır. Gelecekle ilgili sayısal veriler ise içerisinde bulunduğumuz teknoloji çağında sağlıktan eğitime, finanstan pazarlamaya kadar her alanda oldukça önemlidir.

Bu çalışma kapsamında pamuk fiyatlarının 2017 yılındaki değerleri Manisa Ticaret Borsasından alınarak veri madenciliği analizine uygun hale getirilmiştir. Veri seti içerisindeki eksik veriler temizlenmiş ve .arff dosya formatına uygun hale getirilmiştir. Temizlenen veri seti içerisinde yer alan 135 veri eğitim için kullanılmış ve 2018 yılına ait ilk 15 günlük fiyatlar WEKA yazılımının tahmin aracında iki farklı algoritma kullanılarak tahmin edilmiştir.

Tahmin sonucu elde edilen değerler ile gerçekleşen değerler grafiksel olarak karşılaştırılmış ayrıca istatistiksel hata oranlarının da karşılaştırılması gerçekleştirilmiştir. Literatürde yer alan Kaynar ve Taştan (2009), Asilkan ve Irmak (2009) ve Saigal ve Mehrotra (2012) çalışmalarında olduğu gibi performans karşılaştırması için MAE, RMSE ve MAPE hata oranları kullanılmıştır. Yapılan incelemeler sonrasında yapay sinir ağı tekniğinin İbrahim ve Rusli'nin (2007), Delen vd. (2005), Kayri vd. (2017) çalışmalarında olduğu gibi gerçeğe daha yakın değerler elde ettiği görülmektedir. Diğer tahmin edici veri madenciliği teknikleri ile kıyaslandığında yapay sinir ağı algoritmalarının tahmin performanslarının benzer çalışmalarda olduğu gibi daha iyi sonuçlar verdiği görülmüştür.

Bu incelemeler doğrultusunda elde edilen veri için fiyat tahmini yapılırken yapay sinir ağı tekniğinin daha başarılı sonuçlar verdiği görülmüştür. Günümüz teknolojik gelişmeleri ve küresel rekabet ortamında geleceğe yönelik bilgilerin ne kadar önemli olduğu gerçeği gözardı edilemez. Bu nedenle tahminleme yapılırken en başarılı performansı gösteren algoritma ile çalışmak ve gerçeğe en yakın sonuçlara ulaşmak oldukça önemlidir. Bu yönüyle gerçekleştirilen çalışmanın benzer çalışmalar için örnek teşkil edeceği düşünülmektedir. Ayrıca gelecek çalışmalar için çok katmanlı yapay sinir ağı tekniği de kullanılarak derinlemesine performans incelemesi yapılması da planlanmaktadır.

KAYNAKÇA

Aktaş, E. (2006). Çukurova Bölgesi'nde Pamuk Arz Duyarlılığının Tahmini Üzerine Bir Çalışma, <https://mpr.uab.uni-muenchen.de/8648/> (13.08.2018).

- Asilkan, Ö. ve Irmak, S. (2009). İkinci El Otomobillerin Gelecekteki Fiyatlarının Yapay Sinir Ağları İle Tahmin Edilmesi, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 4, Sayı 2, 375-391.
- Allen, R.G.D. (1964). *Statics for Economists*, Mc-Millan, UK, 1964.
- Barnett, V. And Lewis, T. (1994). *Outliers in statistical data*. J. Wiley & Sons
- Crosswhite, C. E. (2003). Method for determining optimal time series forecasting parameters.
- Delen, D., Walker, G. ve Kadam, A. (2004). Predicting Breast Cancer Survivability: A Comparison Of Three Data Mining Methods. *Artificial Intelligence in Medicine*, Volume 35, 113-127.
- Diler, A.İ. (2003). İMKB Ulusal-100 Endeksinin Yönünün Yapay Sinir Ağları Hata Geriye Yayma Yöntemi İle Tahmin Edilmesi. *İMKB Dergisi*, Cilt 7, Sayı 25-26.
- Fayyad, U. Piatetsky-Shapiro, G. ve Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases, *Artificial Intelligence Magazine*.
- Fausett, L. (1994). *Fundamentals of Neural Networks*, PrenticeHall, 1994.
- Giudici, P. (2003). *Applied Data Mining: Statistical Methods for Business and Industry*, John Wiley & Sons, West Sussex, England, 2003.
- Hall, M. (2014). *Time Series Analysis and Forecasting with Weka* (Erişim: 10.10.2018)
- Honaker, J. a. (2010). What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 561-581.
- Ibrahim, Z. ve Rusli, D. (2007). Predicting Students' Academic Performance: Comparing Artificial Neural Network, Decision Tree And Linear Regression, 21st Annual SAS Malaysia Forum, 5th September 2007.
- Irmak, S. Köksal, C. D. ve Asilkan, Ö. (2012). Hastanelerin Gelecekteki Hasta Yoğunluklarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi, *Uluslararası Alanya İşletme Fakültesi Dergisi*, Cilt 4, Sayı 1, 101-114.
- Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*, IEEE Press, Hoes Lane, Piscataway, NJ, USA, 2003.
- Kaynar, O. ve Taştan, S. (2009). Zaman Serisi Analizinde Mlp Yapay Sinir Ağları Ve Arıma Modelinin Karşılaştırılması, *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 33, 161-172.
- Kayri, M., Kayri, İ. ve Gencoğlu, M.T. (2017). The Performance Comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by Using Photovoltaic and Atmospheric Data. 2017 14th International Conference on Engineering of Modern Electric Systems (EMES).
- Manisa Ticaret Borsası Pamuk Raporu (2017). <http://manisatb.org.tr/199/pamuk-sektor-raporlari> (13.08.2018)
- Mananyi, A. ve Struthers, J. (1997). Cocoa Market Efficiency: A Cointegration Approach, *Journal of Economics Studies*, Vol. 24.
- Newbold, P. (2000). İşletme ve İktisat için İstatistik, Çev. Ümit Şenesen, Literatür Yayıncılık, İstanbul.
- Öztemel, E. (2012). *Yapay Sinir Ağları*, Papatya Yayıncılık, İstanbul.
- Saigal, S. ve Mehrotra, D. (2012). Performance Comparison Of Time Series Data Using Predictive Data Mining Techniques. *Advances in Information Mining*, Vol. 4 Num. 1.
- Semerci, A. ve Çelik, A.D. (2018). Hatay İlinde Pamuk Üretiminin Fonksiyonel Analizi. *Tekirdağ Ziraat Fakültesi Dergisi*, Cilt 15, Sayı 2.
- Sevüktekin, M. ve Çınar, M. (2014). *Ekonometrik Zaman Serileri Analizi*, Dora Yayınevi: Bursa.
- Şeker, S. E. (2015). *Zaman Serisi Analizi*, YBS Ansiklopedisi, Cilt 2, Sayı 4.
- Şeker, S. E., Mert, C., Al-Naami, K., Ozalp, N., ve Ayan, U. (2013). Correlation Between the Economy News and Stock Market in Turkey. *International Journal of Business Intelligence and Review (IJBIR)*, Cilt 4, Sayı 4, 1-21.
- Şeker, S. E., Cankir, B., ve Okur, M. E. (2014). Strategic Competition of Internet Interfaces for XU30 Quoted Companies. *International Journal of Computer and Communication Engineering*. 3 (6).
- Telatar, E., Türkmen, Ş. Ve Teoman, Ö. (2002). Pamuk Borsalarında Oluşan Fiyatların Etkinliği, *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 17, Sayı 2.

- Tüik, (2017). Bitkisel Üretim Veri Tabanı([https://biruni.tuik.gov.tr/ bitkiselapp/bitkisel.zul](https://biruni.tuik.gov.tr/bitkiselapp/bitkisel.zul); 01.08.2018).
- Veysel, O. (2016). Weka Nedir? ARFF Dosya Yapısı Nasıldır? (<http://onurveysel.blogspot.com/2016/09/weka-nedir-arff-dosya-yapisi-nasildir.html>; 02.08.2018)

SUMMARY

Data mining is becoming increasingly popular today; many different areas are used in order to increase the profits of the business world. Data mining; decision support systems and strategic planning studies. With the time series analysis technique, possible future events can be detected. Especially in financial systems such as stock market analysis, forecasts can be made with time series analysis with strategic priorities. With the time series analysis techniques used, the characteristics of data clusters can be extracted in data mining. In other words, by determining the movements of the data related to the time, the predictions can be made more consistent and accurate. Artificial neural network technique can be successfully applied in almost every area of life because it is adaptable to adapt to the medium in which it is located, adapts, works with incomplete data, decides in case of uncertainty, and has fault tolerance. Especially in classification, signal filtering, pattern identification, data compression and optimization, artificial neural networks are powerful techniques.

The prices of agriculture and agricultural products are also in areas where large amounts of data are stored and useful information from data mining techniques can be exploited. For example, in addition to the rapid increase in the world population, industrialization and development have increased the standard of living of societies and increased the need for cotton. Cotton, which is the raw material of the textile sector, has a strategic importance both in our country and in the world in different usage areas.

The exchange value of many agricultural products in Turkey willingly goes with the purchase price set by the government. In this study, it is aimed to compare the estimation and forecasting performance of cotton, which is an important agricultural product, by using two different data mining techniques. The study consists of four parts. The first chapter provides general information about cotton production in Turkey. In the second part, the techniques used for estimation are explained. In the third section, the predictive performance of the preferred methods is compared. In the last part, the results and evaluations of the study are included.

In this study annual cotton prices were taken from Manisa Agricultural Market. Later, these data were first corrected, and analyzed in the Weka program. Time series and artificial neural network techniques were used to estimate the value of prices for 2017. Estimated values and actual values are compared graphically and statistical error rates are also compared. MAE, RMSE and MAPE error ratios were used for performance comparison as in Kaynar and Taştan (2009), Asilkan and Irmak (2009) and Saigal and Mehrotra (2012) studies in the literature. Based on the estimated values by MAE, MAPE and RMSE, it is decided which technique will

yield better prediction performance. The results obtained were also compared with similar studies in the literature.

It has been seen that the artificial neural network technique gives more successful results when estimating the price for the data in the direction of these examinations. In addition, in-depth performance studies are planned using multi-layer artificial neural network technique for future studies.