

Sosyal Bilimler Dergisi • Cilt 8–Sayı 15 • Haziran 2018

**PASLANMAZ ÇELİK SEKTÖRÜ SATIŞ TAHMİNİNDE VERİ MADENCİLİĞİ
YÖNTEMLERİNİN KARŞILAŞTIRILMASI**

Comparison of Data Mining Methods in Stainless Steel Sector Sales Forecasting

Orhan ECEMİŞ

Öğr. Gör. Dr. Gaziantep Üniversitesi Oğuzeli Meslek Yüksekokulu Bilgisayar Teknolojileri
Bölümü, ocemis@gantep.edu.tr
Orcid: 0000-0002-8270-0857

Sezgin IRMAK

Doç. Dr. Akdeniz Üniversitesi Uygulamalı Bilimler Fakültesi, Yönetim Bilişim Sistemleri
Bölümü, sezgin@akdeniz.edu.tr
Orcid: 0000-0003-2188-7344

Makale Bilgisi / Article Information

Makale Türü / Article types
Araştırma Makalesi / Research Article

Geliş Tarihi / Received
15 Şubat 2018 / 15 February 2018

Kabul Tarihi / Accepted
20 Şubat 2018 / 20 February 2018

Yayın Tarihi / Published
Haziran / June 2018

Yayın Sezonu / Pub Date Season
Haziran / June

Atıf/Cite as

Ecemiş, O, Irmak, S, (2018), Paslanmaz Çelik Sektörü Satış Tahmininde Veri Madenciliği
Yöntemlerinin Karşılaştırılması/Comparison of Data Mining Methods In Stainless Steel Sector Sales
Forecasting, Sosyal Bilimler Dergisi, 8 (1), **148-169**

Bu makale en az iki hakem tarafından incelenmiştir.

Rights reserved.
For Permissions
sbedergi@kilis.edu.tr

PASLANMAZ ÇELİK SEKTÖRÜ SATIŞ TAHMİNİNDE VERİ MADENCİLİĞİ YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Comparison of Data Mining Methods In Stainless Steel Sector Sales Forecasting

ÖZ

Firmaların üst yönetimi ve tüm departmanları, planlama ve karar alma sürecinde, satış tahminine yönelik verilere ihtiyaç duymaktadırlar. Bu çalışmada paslanmaz çelik sektöründe faaliyet gösteren bir distribütör firmanın satış yaptığı sektörlere göre satış tahminleri gerçekleştirilmiştir. Bunun için firmanın veri tabanından Ocak 2008 ile Mart 2016 arasındaki günlük satış verileri elde edilmiştir. Ham veri setinde bulunan satış hareketleriyle müşteri bilgileri eşleştirilerek sektörlere ait aylık satış rakamları tespit edilmiştir. Veri madenciliği yöntemleriyle (Veri Önleme, Destek Vektör Regresyonu ve Yapay Sinir Ağları) toplam satış ve sektörlere göre satış değerlerinin tahminleri gerçekleştirilmiştir. Modelleme sonucunda Destek Vektör Regresyon yönteminin nispeten daha başarılı olduğu görülmüştür.

Anahtar Kelimeler: Veri Madenciliği, Satış Tahmini, Destek Vektör Regresyonu

ABSTRACT

The senior management and all departments of the companies need data to be able to forecast the sales in the planning and decision-making processes. In this study, sales forecasts of a stainless steel distributor company were made according to the sectors which stainless steel were sold. In this context, daily sales data from January 2008 to March 2016 were obtained from the database of the company. The monthly sales data for the sectors were gathered by matching the customer firms' information and the sales records from the raw data. Total sales and sectoral sales values were estimated by using data mining methods (Support Vector Regression and Artificial Neural Networks). As a result of the modeling, Support Vector Regression method is found to be more successful than the other methods.

Keywords: Data Mining, SalesForecasting, Support Vector Regression

GİRİŞ

Günümüzde ekonomi, meteoroloji, tıp, mühendislik alanlarında, alanlara özgü bilimsel tahmin yöntemleri yoğun olarak kullanılmaktadır. Bu amaçla, tahmin doğruluğunu artırıcı bilimsel çalışmalar yapılmakta ve yeni bilimsel yöntemler araştırılmaktadır. İşletmeler için tahmin çalışması kuruluş aşamasıyla başlamaktadır. Yeni bir kuruluş yeri ihtiyacı zamana ve olaylara bağlı olarak farklılık gösterebilir. Kapasitenin, maliyetlerin ve talebin durumu değişkenlik gösterebilir. Ürün pazarlarının değişmesi durumunda, tüketicilere daha iyi hizmet sunabilmek için tesis yerini değiştirmek gerekebilir (Demirdöğen ve Bilgili,2004). Kuruluş yeri seçimi, yatırım kararı alma, yeni bir ürün üretme, pazarlama stratejilerinin belirlenmesi, personel temini, satış hedeflerinin belirlenmesi gibi kritik süreçlerde uzmanlar tarafından analiz ve tahmin çalışması yapılmaktadır. Bu açıdan tahmin, işletmelerin faaliyetlerinin planlanmasında, hedeflerin belirlenmesinde ve geleceğini yönlendirecek kararların alınmasında bir yol gösterici olduğu ifade edilebilir (Eroğlu, 1996).

Satış tahmini, işletmelerde tüketici talebini anlamaya ve öngörmeye çalışan, yöneticilerin alacakları kararlara destek sağlayan analiz süreci olarak düşünülebilir. Etkin bir talep tahmini, bir şirket için rekabet avantajı sağlar ve pazarlama fonksiyonlarının daha işlevsel kullanılması imkânını sunar. İşletmelerin büyümelerinde, pazar payını korumasında, maliyet yönetiminde, tüketici taleplerinin karşılanması ve müşteri tatmini için alacakları stratejik kararlarda uygun tahmin modellerini kullanması çok önemlidir.

Satış tahminine yönelik çalışmalar, bir asır öncesine dayanan analitik çalışmalarla başlamış, matematik, istatistik ve ekonometri alanındaki gelişmelerle temelleri sağlamlaşmış geçtiğimiz son on yılda bilişim sektöründeki kazanımlar ve veri madenciliği yöntemleriyle ivme kazanmıştır (Chase, 2013:1). Satış tahminleri işletmelerin karar destek sistemlerinde belirgin bir rol oynamaktadır. Etkin satış tahminlerini önceden elde etmek karar vericinin üretim ve malzeme maliyetlerini hesaplamalarına, hatta satış fiyatını belirlemesine yardımcı olabilir. Bu, daha düşük bir envanter seviyesinde çalışmaya ve tam zamanında hedefe ulaşmaya katkı sağlaması beklenmektedir (Kuo ve Xue, 2012).

Paslanmaz Çelik sektöründe fiyatı etkileyen birçok faktör bulunmaktadır. Dünya Borsalarında yer alan ham çelik, nikel, demir, krom madenleri gibi emtia fiyatları paslanmaz çelik üreticilerinin ülkelerinin ekonomik göstergeleri ve tutumları genel olarak fiyatları etkilemektedir.

Emtia fiyat dalgalanmalarına ek olarak, ithal talepten dolayı döviz kuru satış hareketlerini etkilemektedir.(Uçkun, 2010)

Türkiye paslanmaz çelik talebinde büyük oranda dışa bağımlı olduğundan dolayı firmaların satış tahminlerindeki gerçekleşmesi muhtemel büyük sapmalar, stokta çalışma ve bekleme maliyeti gibi ekstra maliyetler oluşturabilir. Maliyet yönetiminde problemlerin yaşanması da firmanın finansal yapısını kırılgan hale getirebilir. Sektörde faaliyet gösteren ithalatçı firmaların, karşılaşılabilecekleri finansman sorununu çözebilmeleri ve etkin kararlar alabilmeleri için satış tahminlerine ihtiyaç duymaktadırlar.

Veri madenciliği, bilgi teknolojilerinin gelişiminin bir sonucu olarak yaygın olarak ve veri tabanları ile büyük veri setlerinde kullanılabilir hale gelmiştir. Veri tabanı ve veri yönetimi endüstrisi, veri toplama ve veri tabanı oluşturma, veri yönetimi (veri depolama, geri alma ve veri tabanı kayıt işleme dahil) ve gelişmiş veri analizi (veri ambarı ve veri madenciliği dahil olmak üzere) gibi kritik işlevlerin geliştirilmesiyle büyümektedir (Han ve Kamber, 2012:3-4).Veri madenciliği yöntemleri Tanımlayıcı ve Tahmin Edici modeller olarak ikiye ayrılabilir. Tanımlayıcı modeller, veri seti hakkında açıklayıcı özet bilgiler sunmaktadır. Tahmin Edici modeller ise belirli hedef doğrultusunda çıktı (bağımlı) değişkenin bilinmeyen değerini tahmin etmeye çalışmaktadır (Hand, 2001:195).Veri Madenciliği uygulamaları, ağırlıklı olarak karar alma ve planlama süreçlerinde yoğunlaşmaktadır. Buna talep, satış ve maliyet tahminlerinin yapılması, pazarlama stratejilerinin belirlenmesi gibi örnekler verilebilir.

Uluslararası Paslanmaz Çelik Forumu (International Stainless Steel Forum - ISSF) tarafından yayımlanan verilere göre, 2012 yılında dünya paslanmaz ham çelik üretimi 2011 yılına göre %5,2 oranında artarak 35,4 milyon ton seviyesine ulaşmıştır. 2002-2012 yılları arasında toplam üretim 15 milyon ton artarak %70'lik bir artış göstermiştir.

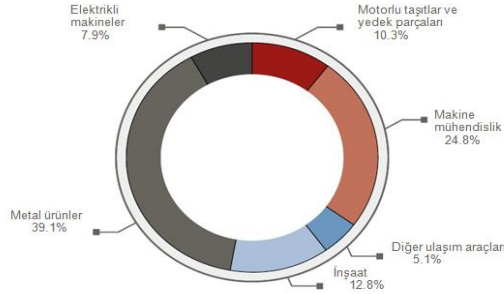
Şekil 1 Ülke/Bölge Bazlı Dünya Paslanmaz Çelik Üretimi



Kaynak: ISSF, 2015

Şekil 1'de 2015 yılı verilerine göre dünya paslanmaz çelik üretiminde Asya Ülkeleri (Çin hariç) %39'luk, Avrupa Birliği ülkeleri %34'lük, Çin ise %13'lük üretim payına sahiptir. 2015 yılı verilerine göre ise, Çin %52'lik üretim ile en büyük paslanmaz çelik üreticisi konumuna geçmiştir. 2015 yılı itibariyle paslanmaz çelik üretiminin %75'lik bölümü Çin ve diğer Asya ülkelerinde gerçekleşmiştir.

Şekil 2 Dünya Geneli Sektörlere Göre Paslanmaz Çelik Talebi



Kaynak: ISSF, 2015

Sektörlere göre dünya geneli paslanmaz çelik talebine ilişkin 2013 yılı grafiği Şekil 2'de yer almaktadır. Talep miktarındaki ilk üç sektörün sıralaması, %39,1 Metal ürünler, %24 Makine, %12,8 Yapı-İnşaat sektörü olarak gerçekleşmiştir.

Bu çalışmada Türkiye’de sektörlere göre paslanmaz çelik satış tahminleri, veri madenciliği teknikleriyle gerçekleştirilmeye çalışılmıştır. Firma veri tabanından satış hareketleri ve müşteri bilgileri tabloları seçilmiş, müşterilerin faaliyet gösterdikleri sektörler belirlenerek müşteri ve satış hareketleri tabloları birleştirilmiş, weka dosya biçimi olan “arff” formatına dönüştürülmüştür. Destek Vektör Regresyon ve Yapay Sinir Ağları yöntemleri ile tahminler üretilmiştir. Sektörlere göre yöntemlerin ürettiği sonuçlara yer verilmiş ve yine sektörlere göre yöntemlerin performansları kıyaslanmıştır.

2. LİTERATÜR ÖZETİ

Literatürde farklı veri madenciliği yöntemleri ile farklı alanlarda yapılmış birçok tahmin çalışması mevcuttur. Son yıllarda bu yöntemler arasında, Destek Vektör Regresyon (DVR) ve Yapay Sinir Ağları (YSA) yöntemleri öne çıkmaktadır. Paslanmaz çelik satış tahmini çalışmaları kısıtlı/erişilememişinden dolayı DVR-YSA yöntemleriyle yapılan tahmin çalışmaları aşağıda yer almaktadır.

Kim (2003), DVM ve YSA yöntemleriyle finansal zaman serileri kullanılarak borsa fiyat indeksini tahmini gerçekleştirilmiştir. Uygulama sonucunda DVM yönteminin YSA'ya göre tahmin çalışmasında daha başarılı olduğu görülmüştür.

Tolun, (2008), DVM-YSA Türkiye’de 1995-2001 yılları arasında faaliyet gösteren özel sermayeli ticaret bankalarının (37 banka) 64 adet değişkenle başarısızlık tahmini için oluşturulan modeller oluşturulmuştur. Destek Vektör Makineleri yönteminin banka başarısızlığının tahmin edilmesinde tutarlılık ve tahmin gücü açısından tatmin edici sonuçlar verdiği görülmüştür.

Khashei ve Bijari (2010), ARIMA-YSA modelini kullanan yapay sinir ağı tasarlamışlardır. ARIMA yöntemiyle model için gerekli veriler üretilmiş, YSA yöntemiyle temel veri üretme süreci ve öngörü gerçekleştirilmiştir. Model, The Wolf’s Sunspot, The Canadian Lynx series, Döviz Kuru (British pound /US dollar) veri setlerine uygulanmış ve geliştirilen modelin geleneksel YSA modellerine göre başarılı bulunmuştur.

Jiang ve He (2012), Finansal zaman serileri tahmini için yapmış oldukları çalışmada Gri İlişkisel Derecelendirme ve Lokal Destek Vektör Makinelerini birleştirerek finansal zaman serileri tahmini gerçekleştirmişlerdir. Çalışma neticesinde LG-DVM'nin tahmin doğruluğunu ve hızını artıracığı görülmüştür.

Marvuglia ve Messineo, (2012) Radyal Tabanlı Dayalı Yapay Sinir Ağı yöntemiyle, hava durumu verileri, klima olup olmama gibi değişkenlerle saat ölçeğinde hane halkı Elektrik tüketim tahminini gerçekleştirmiştir.

Hicham ve Mohamed 2012 Bulanık Küme, YSA Bulanık kümeleme ve Uyarlanmış Öğrenme Oranı (FCBPN) ile Geri-yayılm (BP) Yapay Sinir Ağları üzerine kurulu olan hibrit satış tahmin modeli öne sürmüşlerdir. Bulanık kümeleme yöntemiyle kümelerin daha büyük olduğu ve dolayısıyla önerilen tahmin modelinin endüstriyel tahmin için başarılı olabileceği sonucuna ulaşmışlardır.

Yang ve Li (2016), Otomobil satışlarını tahmin çalışmasında Mevsimsel İndeks ve RBF sinir ağı yöntemlerinin birleşimine dayalı tahmin modeli önerilmiştir. Önerilen birleşim modeli, Mevsimsel İndeks ve RBF modellerine göre daha başarılı bulunmuştur.

Megahed vd. (2016) YSA-ARIMA modelleriyle dergi satışlarına yönelik yapılan çalışmada satırlara ait zaman serileri ve dergi içeriğinin popülerliği ile ilgili veriler kullanılmıştır. Tahmin modeli Geri Yayılm Ağlar ve ARIMA birlikte kullanılarak oluşturulmuştur. Önerilen tahmin modelinin, geleneksel satış tahmin tekniklerine göre daha başarılı olduğu sonucuna ulaşılmıştır.

Vhatkar ve Dias (2016), YSA Kısa raf ömrüne sahip, hızlı tüketim ürünleri olan ağız bakım ürünlerinin (Jel Esaslı-380 mg. Beyazlatıcı diş macunu Jel-500gm) satış tahmini Geriye Yayılm algoritmasıyla gerçekleştirmişlerdir. Geriye Yayılm YSA'nın bu tip ürün satış tahmininde başarılı bir şekilde uygulanabilirliği sonucuna ulaşılmıştır.

3. YÖNTEM

Veri Madenciliği, yapay öğrenme yöntemlerinin büyük veri setlerinde uygulanması olarak ifade edilmiştir (Alpaydın,2011:2-3).Yapay Öğrenme, Denetimli ve Denetimsiz olarak iki farklı türde gerçekleşmektedir. Denetimli öğrenme, var olan örnekleri kullanarak yeni model veya ileriye yönelik olması muhtemel durumların tahmininde kullanılan öğrenme yöntemidir. Denetimsiz öğrenme, sadece genel etiketsiz veri kümesi olduğu durumlarda başvurulmuş öğrenme yöntemidir.

Gelişen veri tabanı yönetim sistemleriyle verinin ifade edilmesi ve kayıt edilmesinde önemli kolaylıklar sağlanmış, bunun sonucu olarak büyük veri yığınları oluşmuştur. Veri tabanı yönetim sistemlerinde sorgulama ve raporlama yöntemlerinden ziyade kaliteli bilgiye ulaşılması ihtiyacı yeni yöntemlerin keşfedilmesini sağlamıştır.

Veri madenciliği yöntemleri, veri setindeki anlamlı kalıpları ve kuralları keşfetmek için büyük miktarda verinin otomatik veya yarı otomatik yollarla araştırılması ve analizi için geliştirilmiştir. Veri madenciliği yöntemleri varsayımlardan ziyade keşif odaklı kullanılmaktadır (Chien ve Chen, 2008). Veri madenciliği genel olarak büyük oradaki veri setlerindeki çeşitli bilinmeyen ya da gizli kalmış kuralların keşfedilmesine ve analizine imkân tanıyan yöntem ve teknikler bütünüdür. Diğer bir ifadeyle veriden bilgi çıkarma sanatıdır (Tuffery, 2011).Veri madenciliği, sınıflandırma, kümeleme, tahmin, ilişki analizi, örüntü tanıma gibi problemlerin çözümünde var olan/geliştirilen algoritmaların, büyük veri setlerinde ön işleme-dönüştürme-modelleme aşamalarından geçerek hızlı ve güvenilir şekilde bilgiye ulaşılmasını sağlayan süreç olarak ifade edilebilir.

Destek vektör makineleri(DVM),sınıflandırma problemlerinin çözümü için geliştirilmiştir. Destek Vektör Makinelerinin, karmaşık sınıflandırma problemlerinde (el yazısı tanıma, zaman serisi analizi, kanser teşhisi) başarılı olduğu söylenebilir (Lantz,2013:225; Akpınar,2014:268, Cortes ve Vapnik, 1995). DVM'lerin en önemli avantajı, basit bir geometrik yorumu olması ve seyrek bir çözüm sunmasıdır. Yapay sinir ağlarının aksine, DVM'lerin hesaplama karmaşıklığı, giriş alanının boyutuna bağlı değildir. YSA'lar ampirik risk minimizasyonunu kullanırken, DVM'ler yapısal risk

minimizasyonunu kullanmaktadırlar. DVM'lerin bazı dezavantajları olarak çekirdek parametrelerinin seçimi, ayırık verilerin işlenmesi, karesel programlamanın algoritmik karmaşıklığı ve yüksek bellek gereksinimi olarak ifade edilebilir (Olson ve Delen 2016:139). Yüksek bellek gereksinimi sınıflandırma probleminde SMO, regresyon probleminde SMOreg algoritmasıyla çözülmüştür (Platt, J. C. 1999).

Destek Vektör Makineleri, diğer sınıflandırma yöntemlerine göre eğitim süresinin daha fazla olmasına rağmen, yüksek güvenilirlik düzeyi, aşırı öğrenmeye dayanıklı olması ve doğrusal olmayan sınıflandırma problemlerinde başarılı bir şekilde uygulanması ile öne çıkmaktadır.

Destek Vektör Regresyonda DVM'ye benzer biçimde kenar payı, artık değişkenler tanımlanmaktadır (Alpaydın,2011:279-280). Destek Vektör Regresyon (DVR) yöntemi, Destek Vektör Makinelerinin regresyon problemine uyarlanması olarak ifade edilebilir.

$$f(x) = \mathbf{w}^T x + w_0$$

Doğrusal regresyonda hata, gözlenen değer ile beklenen değerlerin farkının karesidir.

$$e_e(r^t, f(x^t)) = [r^t - f(x^t)]^2$$

Destek Vektör Regresyonda ϵ duyarsız kayıp işlevi kullanılmaktadır.

$$e_\epsilon(r^t, f(x^t)) = \begin{cases} 0 & \text{eğer } |r^t - f(x^t)| < \epsilon \\ |r^t - f(x^t)| - \epsilon & \text{değilse} \end{cases}$$

ϵ duyarsız kayıp işleviyle, ϵ değerine kadar olan hatalar göz ardı edilmekte, daha büyük hataların etkisi regresyondan farklı olarak karesel değil, doğrusal olmaktadır. Bunun sonucu olarak gürültüye dayanıklı, gürbüz bir regresyon modeli elde edilmektedir. (Alpaydın,2011:279-280)

DVR aşağıdaki ifadesi yer alan karesel programlamayla çözülmektedir.

$$\min \frac{1}{2} \|w\|^2 + C \sum_t (\xi^t_+ + \xi^t_-)$$

C hataların uyarlanması ile regresyon fonksiyonunun düzlüğü arasındaki dengeyi sağlayan önceden belirlenen bir düzenleme bir parametresidir. $(\xi^t_+ + \xi^t_-)$ örneklerin ϵ -tüpüne girip girmediğini gösteren artık değişkenlerdir.

Artık değişkenlerimiz artı ve eksi yönde tanımlanır ve kısıtlar:

$$r^t - (w^T x + w_0) \leq \epsilon + \xi^t_+$$

$$(w^T x + w_0) - r^t \leq \epsilon + \xi^t_-$$

$$\xi^t_+, \xi^t_- \geq 0$$

Langrange terimleri:

$$\begin{aligned} L_p &= \frac{1}{2} \|w\|^2 + C \sum_t \xi^t_+ + \xi^t_- \\ &- \sum_t \alpha^t_+ [\epsilon + \xi^t_+ - r^t + (w^T x + w_0)] \\ &- \sum_t \alpha^t_- [\epsilon + \xi^t_- + (w^T x + w_0) - r^t] + \\ &- \sum_t \alpha^t_- [\epsilon + \xi^t_- + (w^T x + w_0) - r^t] \\ &- \sum_t (\mu^t_+ \xi^t_+, \mu^t_- \xi^t_-) \end{aligned}$$

Kısmi türevlerin alınarak:

$$\frac{\partial L_p}{\partial w} = w - \sum_t (\alpha_+^t - \alpha_-^t)x^t = 0 \rightarrow w = \sum_t (\alpha_+^t - \alpha_-^t)x^t$$

$$\frac{\partial L_p}{\partial w_0} = \sum_t (\alpha_+^t - \alpha_-^t)x^t = 0$$

$$\frac{\partial L_p}{\partial \xi_+^t} = C - \alpha_+^t - \mu_+^t = 0$$

$$\frac{\partial L_p}{\partial \xi_-^t} = C - \alpha_+^t - \mu_+^t = 0$$

Duali:

$$L_d = -\frac{1}{2} \sum_t \sum_s (\alpha_+^t - \alpha_-^t)(\alpha_+^s - \alpha_-^s)(x^t)^T x^s$$

$$- \epsilon \sum_t (\alpha_+^t + \alpha_-^t) - \sum_t r^t (\alpha_+^t - \alpha_-^t)$$

Kısıtlar:

$$0 \leq \alpha_+^t \leq C, 0 \leq \alpha_-^t \leq C, \sum_t (\alpha_+^t - \alpha_-^t) = 0$$

İstenilen duyarlılıkta tahmin edilen örnekler ϵ alanında yer almaktadır. İki tür destek vektörü oluşmaktadır. (Alpaydın,2014:370-371)

- İlk türde ϵ alanın sınırında oluşan destek vektörleri için $\alpha_+^t > 0$ yada $\alpha_-^t > 0$ koşulları sağlanır w_0 değeri hesaplanır.
- İkinci tür ϵ alanın dışında yer aldığından dolayı başarılı tahmin yapılamamıştır.

Destek vektörlerinin ağırlıklı toplamını kullanarak ifade edilebilir.

$$f(x) = \mathbf{w}^T x + w_0 = \sum_t (\alpha_+^t - \alpha_-^t) (x^t)^T X + w_0$$

Yapay sinir ağları (YSA) sınıflandırma ve regresyon problemlerinde kullanılan parametrik olmayan bir tahmin yöntemidir (Alpaydın,2011:197). YSA'nın temel birimi olan nöronlar, ele alınan problemin özelliğine göre birbirlerine bağlı olarak büyük bir ağ yapısını oluşturabilirler. Böylece karmaşık ilişkilerin modellenmesine imkân sağlamaktadırlar. YSA yapısında giriş, gizli, çıkış katmanları olmak üzere en az üç katman bulunmaktadır. Yapay sinir ağlarının eğitiminde genellikle çevrimiçi öğrenme yöntemi kullanılmaktadır. Çevrimiçi öğrenmede, öğrenme kümesinin tamamı yerine örnekler ardışık olarak kullanılmaktadır. Her örnekte, ağ parametrelerinin güncellenmekte ve zaman içinde veriye uyarlanmaktadır.

Yapay Sinir Ağları nöronlarında birleştirme ve aktivasyon fonksiyonları olmak üzere iki farklı fonksiyon görev yapmaktadır. Birleştirme fonksiyonu, bir hücreye gelen net girdiyi hesaplayan bir fonksiyondur. Herhangi bir nöronunun toplam girdisi, diğer nöronlardan gelen değerlerin ağırlıklı toplamı ile eşik değerinin toplamına eşittir (Kamruzzman, 2006, 3). Aktivasyon fonksiyonunun nörona gelen net girdiyi işlenmesinin ardından üretilecek çıktıyı belirleyen fonksiyondur. Hücrenin gerçekleştireceği işleve göre lineer, sigmoid, hiperbolik tanjant vb. çeşitli tipte aktivasyon fonksiyonları kullanılabilir.

Sınıflandırma, tahmin ve karmaşık problemlerin çözümünde kullanılan Yapay Sinir Ağları, Geriye Yayım Modeli veya "Hata Yayma Modeli olarak da bilinmektedir. Geri yayılım eğitim algoritması, yapay sinir ağı modellenmesinde yaygın olarak kullanılan bir algoritmadır. Geri yayılma algoritması, ağın istenen ve fiili çıkışı arasındaki ortalama karesi hatayı en aza indirmek için ağ ağırlıklarını değiştiren en küçük ortalama karesel hata ölçümünün genelleştirilmesidir (Chiang vd.,2016). YSA yapısında girdi katmanı, gizli katman ve çıktı katmanı bulunmaktadır. Ele alınan probleme göre gizli katman sayısı ayarlanabilir (Zhang vd., 1998:37-38).

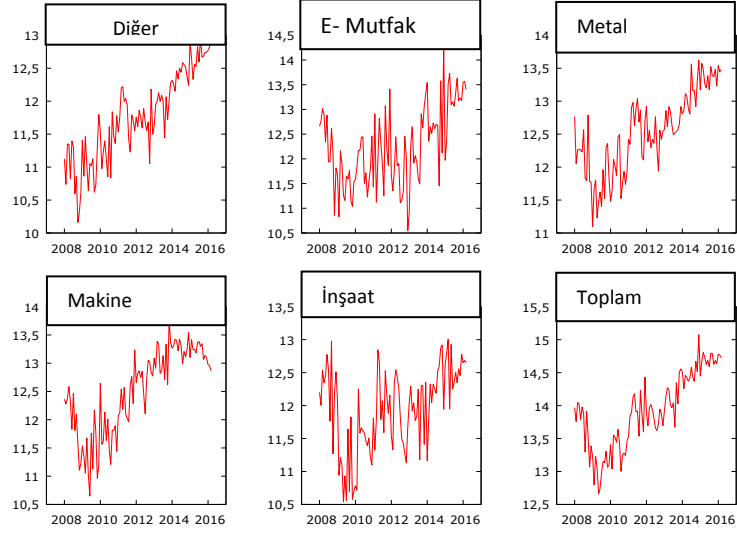
4. UYGULAMA

Verinin modellenmesi aşamasında açık kaynak kodlu WEKA yazılımından yararlanılmıştır. WEKA, makine öğrenimi algoritmalarını içeren Waikato Üniversitesi bilim insanları tarafından geliştirilmiştir. Genel Kamu Lisansına sahip, açık kaynak kodlu bir veri madenciliği programıdır (Written ve Frank, 2005).

Weka'da, 3.7.3 sürümünden itibaren, tahmin modellerinin geliştirilmesine, değerlendirilmesine ve görselleştirilmesine olanak veren özel bir zaman serisi analiz paketi bulunmaktadır. Bu pakette, verilerin standart önermeli öğrenme algoritmalarının işlenebileceği bir forma dönüştürerek zaman serilerini modellemek için bir makine öğrenme / veri madenciliği yöntemleri uygulanabilmektedir. Gecikme (Lag) değerleri, trendleri ve mevsimselliği modellemesi için çeşitli diğer alanlar da otomatik olarak hesaplanır. Veriler dönüştürüldükten sonra, bir modeli öğrenmek için, klasik istatistiksel tekniklerin (ARMA ve ARIMA) çoğu zaman esnek olan Weka'nın regresyon algoritmalarından herhangi biri uygulanabilir (Bouckaert vd 2013; wiki.pentaho.com/).

Türkiye genelinde paslanmaz çelik distribütörü olan bir firmaya ait veri tabanından 2008-2016 yılları arasındaki yüz binden fazla kayıt içeren Satış Hareketleri ve Müşteriler tablolarından alınan veriler eşleştirilerek, satış yapılan müşteri firmalar düzeyinde günlük satış verileri elde edilmiştir. Günlük satış verileri aylık verilere dönüştürülerek sektörlere göre aylık satış rakamları elde edilmiştir. Firmanın satış yaptığı sektörler, ağırlıklı olarak paslanmaz çeliğin yoğun olarak kullanıldığı Metal ve Makine sektörleri ile Endüstriyel Mutfak ve Yapı-İnşaat sektörlerini kapsamaktadır. Toplam satışa göre ve yoğun satış yapılan sektörlere göre oldukça az satış yapılan (%1-%2) sektörler "diğer sektör" olarak gruplandırılmıştır. Satış verilerine ait zaman serisi Şekil 3'te yer almaktadır.

Şekil 3 Aylık satış verilerine ait zaman serileri



Uygulamada veri seti, tek değişkenli, sektörlere göre satış verilerinden oluşmaktadır ve eğitim ve test aşaması olarak %70-%30, %80-%20,%90-%10 oranında ayrılmıştır. Veri setinin eğitim ve test aşaması olarak belirtilen oranlarda ayrılması ayrıca her denemede ön işleme (normalizasyon ve standardizasyon), YSA için öğrenme oranı, momentum, katman sayıları, aktivasyon fonksiyonu, DVR için çekirdek ve parametre seçimi gerçekleştirilmiştir.

Parametre seçiminin yanı sıra, zaman serilerinin varyansları, model ve gecikmeli değişkenler oluşturulmadan önce, değişkenlerin logaritmaları alınarak düzeltilmektedir. Weka'da, gecikme uzunluğu, zaman serilerinin periyodik (saat, gün, hafta, ay) özelliklerine göre sistem / kullanıcı tarafından en küçük, en büyük değerlerin atanmasıyla belirlenmektedir. Gecikme değerlerinin kaç dönem öncesi başlaması ve ardışık gecikme değerlerinin ortalamaları alınarak da belirlenebilmektedir. Gecikme değerlerinin ortalaması ve ardışık uzunlukların belirlenmesiyle değişken sayısını azaltılmakta böylece veriler üzerinde geniş bir pencere elde edilmektedir.

Satış verilerine ait tanımlayıcı istatistik değerler aşağıdaki Tablo 1'de yer almaktadır.

Tablo 1 Tanımlayıcı İstatistik değerler

Kriterler	Diğer	E.Mutfak	Metal	Makine	İnşaat	Toplam
Çarpıklık	0,922929	2,114362	0,839848	0,429178	0,601803	0,741406
Basıklık	-0,01072	6,676414	-0,16408	-0,72151	-0,37799	-0,2268
Ortalama	156963,7	285076,7	331605,3	340399	183373,3	1297418
Standart Sapma	100049,7	248829,7	194987,1	209305	102476	707095,2

Weka yazılımı kullanılarak sektörlere göre satış değerlerinin ve toplam satış tutarının tahmin edilmesi Destek Vektör Regresyonu (DVR) ve Yapay Sinir Ağları (YSA) yöntemleriyle yapılmış ve sonuçlar birbirleriyle karşılaştırılmıştır. Yöntemler, eğitim-test seti oranları ve kullandıkları parametrelerin oluşturduğu MAPE kriterine göre değerlendirilmiştir. DVR’de (Çekirdek fonksiyonu, C katsayısı, Öğrenme oranı, Gecikme, Ardışık ortalama) göre tekrarlı bir şekilde denenmiştir. DVR’de elde edilen en başarılı modellere ait parametre değerleri Tablo 2’de yer almaktadır.

DVR ve YSA yönteminde en başarılı sonuçlar veri setinin yaklaşık son bir yılını test seti olarak ifade edebileceğimiz %90-%10 olarak ayrılması elde edilmiştir. DVR parametreleri aşağıdaki Tablo 2’de yer almaktadır.

Tablo 2 DVR parametreleri

Parametre	Toplam	Metal	Makine	E- Mutfak	İnşaat	Diğer
C	0.85	2	0,9	1	1	1
Çekirdek	Normalized PolyKernel	PolyKernel	PolyKernel	PolyKernel	Normalized PolyKernel	PolyKernel
Ön işleme	Normalize	Standardize	Standardize.	Standardize	Normalize	Standardize
Gecikme/Ardışık Ortalama	1/2	2/2	2/2	2/2	1/2	2/2

Çok katmanlı algılayıcı yönteminde, gizli katmanda bulunan düğüm sayısı, öğrenme katsayısı ve momentum parametreleri modelin başarısında etkilidir. YSA parametreleri aşağıdaki Tablo 3’de yer almaktadır. Aktivasyon fonksiyonu olarak sigmoid fonksiyon seçilmiştir.

Tablo 3 YSA parametreleri

Parametre	Toplam	Metal	Makine	E-Mutfak	İnşaat	Diğer
Öğrenme Oranı	0,1	0,2	0,1	0,2	0,2	0,1
Momentum	0,1	0,2	0,1	0,2	0,2	0,1
Node	14	7	6	7	7	4
Gecikme/Ardışık Ortalama	1/2	2/2	2/2	2/2	1/2	2/2

Tablo 4'de her iki yöntem için bütün sektörler için satış tahmin modellerinin eğitim aşaması model değerlendirme sonuçları sunulmuştur.

Tablo 4 Eğitim Aşaması Sonuçları

Sektörler	Yöntemler ve Model Değerlendirme Kriterleri					
	MAPE		MAE		RMSE	
	DVR	YSA	DVR	YSA	DVR	YSA
Toplam	16,84	14,32	205885,04	177022,62	324168,73	261651,79
Metal	16,84	24,01	61948,02	68290,82	93692,58	89230,69
Diğer	23,24	21,66	27680,86	26908,36	38036,28	36367,86
Makine	24,20	24,85	68662,57	72152,01	103464,61	100109,46
Yapı-İnşaat	33,01	24,41	49358,65	39952,51	72507,13	60748,31
E-Mutfak	47,41	59,56	114066,76	120169,17	215876,53	201133,65

Eğitim aşaması sonuçlarına göre DVR yönteminde MAPE değerleri %16,84 ve %47,41 arasındadır. YSA yönteminde ise MAPE değerleri %14,3 ve %59,56 arasında yer almaktadır. Tablo 5'te eğitim aşamasında elde edilen Destek Vektör Regresyonu ve Yapay Sinir Ağları modelleri ile bütün sektörlerin satış değerlerine ilişkin gerçekleştirilen test aşaması sonuçları yer almaktadır. Genel olarak tahmin çalışmalarında eğitim aşamasında hesaplanan MAPE değerleri test aşamasından daha başarılı çıkmaktadır. Uygulamada yöntemlere ait parametreler sabit tutma, artırma ve azaltma gibi işlemlerle defalarca denenmiş ve en iyi değerler sunulmuştur. İlk aşamada, genel kanıya uygun olarak, elde edilen bulgularda eğitim aşaması, test aşamasından daha başarılı çıkmıştır. Gecikme parametresi ve ardışık gecikme ortalama değerlerinin ayarlanmasıyla test aşaması, eğitim aşamasından daha başarılı olmuştur

Tablo 5 Test Aşaması Sonuçları

Sektörler	Yöntemler ve Model Değerlendirme Kriterleri					
	MAPE		MAE		RMSE	
	DVR	YSA	DVR	YSA	DVR	YSA
Toplam	6,21	5,53	158845,08	136785,9	237224,75	163459
E-Mutfak	8,75	7,85	54124,09	45043,60	69948,08	65684,20
Makine	8,25	13,8	50015,39	81088,26	72092,83	105907,45
Diğer	8,37	9,85	31474,84	34822,55	46067,43	38588,91
Yapı-İnşaat	8,47	9,14	21745,46	25046,11	27230,61	31768,9
Metal	9,9	9,19	68688,17	59580,54	97935,42	73436,61

Tablo 5'e göre YSA ve DVR yöntemleri firmanın Toplam Satış miktarını sırasıyla %5,53 ve %6,21 MAPE değerleriyle tahmin etmiştir. Bu değerler ayrıca yöntemlerin en başarılı oldukları değerlerdir.

- Endüstriyel Mutfak sektörü satışları YSA ile %7,85, DVR ile %8,75 MAPE değerleriyle tahmin edilmiştir.
- Makine Sektör Satışları DVR ile %8,25, YSA ile %13,8 MAPE değeri ile tahmin edilmiştir.
- Diğer Sektör Satışları DVR ile %8,37, YSA ile %9,85 MAPE değeri ile tahmin edilmiştir.
- Yapı-İnşaat Sektör Satışları DVR ile %8,47, YSA ile %9,14 MAPE değeri ile tahmin edilmiştir.
- Metal Sektör Satışları YSA ile %9,19, DVR ile %9,9 MAPE değeri ile tahmin edilmiştir.

YSA yöntemi Toplam, Endüstriyel Mutfak ve Metal Sektörlerini DVR yöntemine göre daha iyi tahmin etmiştir. DVR yöntemi Makine, Diğer ve Yapı İnşaat Sektörlerini YSA'dan daha iyi tahmin etmiştir. YSA yöntemi en başarılı Toplam Satış ve en başarısız Makine Sektörü tahmin değerlerini üretmiştir. DVR yönteminde üretilen tahminlerde, MAPE değeri %10 altında kalmasından dolayı, yöntemin çeşitli veri setlerine uyumu açısından YSA yöntemine göre daha başarılı olduğu söylenebilir.

SONUÇ

ISSF verilerine göre ülkemiz, dünya genelinde paslanmaz çelik sektöründe 2012 rakamlarına göre 5. büyük ithalatçıdır. Ülkemizde sanayi sektörü ve hayat standartları geliştiği oranda ikamesi olmayan paslanmaz

çelik ihtiyacının artması beklenmektedir. Paslanmaz çelik yurt içi talebin yanı sıra ülkemizin ihracatında önemli yeri olan Makine, Metal ve Yapı-İnşaat ve Endüstriyel Mutfak gibi alanlarda kullanılmaktadır. Bu nedenle ithalatçı firmaların gerek kendi sürdürülebilir büyümeleri gerekse ülke ekonomisine destek sağlamaları açısından talep tahminleri etkin bir şekilde yapmalıdırlar.

DVR, DVM'lerin regresyon problemine uyarlanmasıdır. DVM, istatistiksel öğrenme teorisine dayanmaktadır. DVM'lerin, geleneksel sinir ağları tarafından benimsenen ampirik risk yerine, daha üstün olan yapısal risk minimizasyonunu benimsemesi, genelleme kabiliyetinin benzersiz bir çözüm sağlamadaki avantajları nedeniyle, araştırmacıların dikkatini çekmiş ve doku sınıflandırması, görüntü tanıma, veri madenciliği ve biyoformatik, zaman serisi gibi birçok uygulamada uygulanmıştır (Lu,2009).

Yapay Sinir Ağları; gelişimi McCulloch-Pits modeli ile başlayarak pek çok önemli kavramın ortaya çıkmasını sağlamış, duraklama ve yeniden doğuş dönemleri geçirmiş araştırmacılar tarafından birçok alanda başarı ile uygulanmıştır(Akpınar,2014).

Bu çalışmada veri madenciliği sürecine uygun olarak oldukça büyük ham veri setinden ilk önce günlük satış hareketleri elde edilerek, Makine, Metal, Endüstriyel Mutfak, Yapı-İnşaat sektörlerine ait aylık satış rakamları belirlenmiştir. Bu sektörlerin dışında kalan aynı zamanda toplam satışa katkısı az olan sektörler için aylık satış rakamları Diğer Sektör adı altında kaydedilmiştir. Veri Madenciliği tahmin yöntemlerinden Yapay Sinir Ağları, Destek Vektör Regresyon yöntemleriyle toplam ve sektörlerin 1 aylık kısa dönem satış tahminleri gerçekleştirilmiş ve sonuçlar karşılaştırılmıştır. DVR sonuç değerleri, % 6,21 (Toplam Satış) ve %9,9 (Metal) aralığında olup başarılı olduğu söylenebilir. YSA sonuç değerleri ise %5,51 (Toplam Satış)-%13,8 (Makine) aralığında gerçekleşmiştir.

Tahmin problemlerinde son yıllarda yapılan çalışmalarda veri madenciliği yöntemleri uygulanmış ve başarılı sonuçlar elde edilmiştir. Veri Madenciliği sürecine uygun yürütülen çalışmalar, ham veri setinden elde edilen veriler ışığında, sınıflandırma, kümeleme, karar ağacı uygulamaları ve tahmin çalışmalarında çok önemli bulgulara ulaşılmasını sağlamaktadır. Bu çalışmadaki veri setine göre YSA yöntemi ile en iyi (Toplam Satış,%5,53) ve en kötü tahmin (Makine,%13,8) değerleri elde edilmiştir. YSA yöntemi Makine sektörü haricinde kalan sektörler için başarılı olduğu söylenebilir. DVR, en başarılı sonucu bulamasa da, YSA'da var olan ampirik risk minimizasyonu yerine yapısal risk minimizasyonu prensibini benimsenmesinden dolayı, veri setine daha iyi uyum sağlayarak bütün sektörler için başarılı sonuçlar verdiği, dolayısıyla daha tutarlı tahminler

ürettiği söylenebilir. Uygulamada mevcut lisanslı ticari yazılımlar yerine kullanılan WEKA yazılımı, açık kaynak kodlu, genel kamu lisansına sahip ve kolay uygulanabilir olmasından dolayı tercih edilmiştir.

KAYNAKÇA

Alpaydın, E. (2011),Yapay Öğrenme, İstanbul, Boğaziçi Üniversitesi Yayınevi.

Alpaydın, E. (2014). *Introduction to Machine Learning*. MIT press.

Akpınar, H. (2014).DATA Veri Madenciliği – Veri Analizi, İstanbul, Papatya Yayınevi.

Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., ve Scuse, D. (2013). WEKA Manual for Version 3-7-8, 2013.

Chase, C. W. (2013). *Demand-Driven Forecasting: A Structured Approach To Forecasting*. John Wiley & Sons.

Chien, C. F. ve Chen, L. F.(2008). Data MiningtoImprove Personnel Selectionand Enhance Human Capital: a Case Studyin High-Technology Industry, Expert Systems with Applications, 34: 280-290.

Chiang, W. C., Enke, D., Wu, T. ve Wang, R. (2016). An Adaptive Stock Index Trading Decision Support System. Expert Systems with Applications, 59, 195-207.

Cortes, C. ve Vapnik, V. (1995). Support-Vector Networks. *Machine learning*, 20 (3), 273-297.

Demirdöğen, O. ve Bilgili, B. (2004). Organize Sanayi Bölgeleri İçin Yer Seçimi Kararlarını Etkileyen Faktörler: Erzurum Örneği. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 4(2).

Eroğlu, A. (1996). Pazarlama Planlaması Sürecinde Satış Tahminleri Ve Çoklu Regresyon ve Korelasyon Analizinin Etkinliği. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 1(1).

- Frank** E., Wang Y., Inglis S., Holmes, G. ve Witten I. H. (1998). Using Model Trees for Classification. *Mach. Learn.* 32(1): 63-76
- Han**, J. ve Kamber, M. (2012). *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, USA.
- Hand**, D., Mannila, H. ve Smyth, P. (2001). *Principles of Data Mining*, The MIT Press, Cambridge, Massachusetts, USA.
- Hicham**, A. ve Mohamed, B. (2012, November). A Model for Sales Forecasting Based on Fuzzy Clustering and Back-Propagation Neural Networks With Adaptive Learning Rate. In *Complex Systems (ICCS), 2012 International Conference on*, IEEE, 1-5.
- Jiang**, H. ve He, W. (2012). Grey Relational Grade in Local Support Vector Regression for Financial Time Series Prediction Expert Systems with Applications, 39(3): 2256-2262.
- Kamruzzaman**, J. (Ed.). (2006). *Artificial Neural Networks In Finance and Manufacturing*. IGI Global.
- Khashei**, M., ve Bijari, M. (2010). "An Artificial Neural Network (P,D,Q) Model for Time Series Forecasting", *Science Direct Journal of Expert System with Applications*, 37: 479-489.
- Kim**, K. J. (2003). Financial Time Series Forecasting Using Support Vector Machines *Neurocomputing*, 55(1): 307-319.
- Kuo**, R.J. ve Xue, K.C. (1998) "An Intelligent Sales Forecasting System Through Integration of Artificial Neural Network and Fuzzy Neural Network", *Computers In Industry* 37: 1-15.
- Lantz**, B. (2013). *Machine Learning With R*. Packt Publishing Ltd.
- Lu**, C. J., Lee, T. S., ve Chiu, C. C. (2009). Financial Time Series Forecasting Using Independent Component Analysis And Support Vector Regression. *Decision Support Systems*, 47(2), 115-125.

- Marvuglia**, A. ve Messineo, A.(2012) “Using Recurrent Artificial Neural Networks to Forecast House hold Electricity Consumption”, Journal of Energy Procedia,14: 45-55.
- Megahed**, A.,Yin, P.ve Nezhad, H. R. M. (2016, June). An Optimization Approach to Services Sales Forecasting in a Multi-Staged Sales Pipeline. In Services Computing (SCC), 2016 IEEE International Conference on: 713-719, IEEE.
- Olson**, D. L., ve Delen, D. (2008). *Advanced datamining techniques*. Springer Science& Business Media.
- Platt**, J. C. (1999). 12 Fast Training Of Support Vector Machines Using Sequential Minimal Optimization. *Advances in Kernel Methods*, 185-208.
- Silahtaroglu**, G. (2013) "Veri Madenciliği" Papatya Yayınları, İstanbul.
- Tolun**, S. (2008).Destek Vektör Makineleri: Banka Başarısızlığın Tahmini Üzerine Bir Uygulama, İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 2008
- Tuffery**, S., (2011) *Data Mining and Statistics For Decision Making*, 1st ed., Wiley, USA
- Uçkun**, N. (2010). Paslanmaz Çelik Sektöründe Fiyat Riskinden Korunmak İçin Krom Nikel Ham Çelik ve EURO/USD Paritesinin Fiyatlara Etkilerinin İncelenmesi. *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, 11(1).
- Written**, I. H. ve Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques Second Edition*, Morgan Kaufmann Publishers, San Francisco, CA, USA, 2005.
- Vhatkar**, S.ve Dias, J. (2016). Oral-Care Goods Sales Forecasting Using Artificial Neural Network Model. *Procedia Computer Science*, 79: 238-243.

Yang, L., ve **Li, B.** (2016). The combination forecasting model of auto sales based on seasonal index and RBF neural network. *Decision Support Systems*, 9(1).

Zhang, G.,**Patuwo, G. E.** ve **Hu, M. Y.** (1998). “Forecasting with Artificial Neural Networks: The State of The Arts”, *International Journal of Forecasting*, 4(1): 35.

Crisp-Dm, SAS Enterprise Miner,
<http://www.sas.com/offices/europe/uk/technologies/analytics/atamin ing/miner/semma.htm>, (erişim tarihi: 01.01.2015)

The International Stainless Steel Forum
(ISSF)<http://www.worldstainless.org/>(erişim tarihleri:01.01.2014-06.06.2016

<https://wiki.pentaho.com/>