



İKİNCİ EL OTOMOBİLLERİN GÜNCEL PAZAR FİYATLARININ VERİ MADENCİLİĞİ YÖNTEMLERİYLE MODELLENMESİ



Yrd. Doç. Dr. Özcan ASILKAN¹

ÖZET

Bilindiği gibi, ikinci el otomobiller tüketicilerin en çok ilgi gösterdiği dayanıklı ürünlerdendir. Çok sayıda teknik ve opsiyonel özelliğe sahip olan otomobillerin ikinci el pazarındaki fiyatlarının matematiksel bir modelini oluşturmak oldukça güçtür. Bu çalışmada iki farklı veri madenciliği yöntemi kullanılarak ikinci el otomobillerin pazardaki güncel fiyatları modellenmeye (tahmin edilmeye) çalışılmıştır. Girdi verisi olarak Avrupa merkezli çok sayıda ikinci el otomobil web sitesinden elde edilmiş olan ilanlar kullanılmıştır. Regresyon Analizi ve Yapay Sinir Ağları uygulaması ile modeller oluşturulduktan sonra bu modeller karşılaştırılmıştır. Elde edilen sonuçlar, veri madenciliğinin ikinci el otomobillerin güncel pazar fiyatlarını modellemede başarılı olarak kullanılabileceğini ortaya koymuştur.

Anahtar Kelimeler: Veri Madenciliği, İkinci El Otomobil, Fiyat Modelleme, Tahmin

MODELING THE CURRENT MARKET PRICES OF THE SECOND-HAND CARS BY USING DATA MINING METHODS

ABSTRACT

As is known, second-hand cars are among the durable goods of which the customers are highly interested in. It's very difficult to form a mathematical model of second-hand cars which include many technical and optional attributes. This study has tried to model (forecast) the current market prices of the second-hand automobiles using two different data mining techniques. The advertisement data gathered from Europe-based second-hand automobile web sites have been used as the input of the study. After the models had been formed with Regression Analysis and Artificial Neural Networks, they also have been compared. As a conclusion, it has been shown that data mining can be used successfully to model the current market prices of the second-hand cars.

Keywords: Data Mining, Second-Hand Car, Price Modeling, Forecast

¹ Epoka Üniversitesi, MMF, Bilgisayar Mühendisliği Bölümü, Tiran-Arnavutluk, osilkan@epoka.edu.al



1. GİRİŞ

Günümüzde otomotiv endüstrisi, neredeyse tüm sanayileşmiş ülke ekonomilerinde lokomotif konumunda bulunmaktadır. Diğer yandan, ikinci el otomobil satış hacmi, hemen her ülkede yeni otomobil satış hacminden çok daha büyük hale gelmiştir.

İkinci el otomobil pazarındaki en önemli unsur, hiç şüphesiz ikinci el otomobillerin fiyatlarındaki değişkenliklerdir. İkinci el otomobillerin fiyatlarının yeni otomobil fiyatlarına göre çok daha değişken olması, tahmin edilmelerini de zorlaştırmaktadır. Bu nedenle, ikinci el otomobillerin fiyatlarını tahmin edebilmek için her şeyden önce çok kapsamlı ve nitelikli verilerin toplanması gerekmektedir.

Veri madenciliği teknikleri son yıllarda gizli örüntüleri ortaya çıkarmak, bilinenlerden yola çıkarak bilinmeyenleri tahmin etmek amacıyla hemen her alanda yaygın olarak kullanılmaktadır. Özellikle doğrusal olmayan zaman serilerinde Yapay Sinir Ağları (YSA)'nın geleneksel yöntemlerden daha üstün başarı göstermesi bu yöntemin tercih edilmesinde önemli bir faktör olmuştur (Zhang vd., 1998: 35-62). Literatürde bu konuda değişik alanlarda yapılmış çok sayıda çalışmaya rastlamak mümkündür. Ancak ne yazık ki otomobil sektöründeki çalışma sayısı çok kısıtlıdır. Bu sektörde rastlanmış olan en kayda değer çalışmada İşeri ve Karlık (2008), YSA'nı kullanarak yeni otomobillerin özellikleri ile fiyatları arasındaki ilişkiyi ortaya çıkarıp fiyatlarını tahmin etmeye çalışmışlardır. Asilkan ve Irmak (2009) da Zaman Serileri ve YSA'nı kullanarak, ikinci el otomobillerin gelecekteki fiyatlarını tahmin etmeye çalışmışlardır. Ancak ikinci el otomobillerin pazardaki güncel fiyatlarını modellemeye (tahmin etmeye) yönelik kayda değer bir çalışmaya rastlanamamıştır. Bu çalışmanın, belirtilen konudaki yazına katkı sağlaması beklenmektedir.

Çalışmada, veri madenciliği yöntemleri arasında yer alan Regresyon Analizi (RA) ve YSA kullanılarak, ikinci el otomobillerin teknik, opsiyonel ve zamanla değişen özellikleri ile pazar fiyatları arasındaki ilişki tahmin edilmeye, diğer bir deyişle fiyat modelleri oluşturulmaya çalışılmıştır. Çalışmanın sonunda, her iki yöntemden elde edilen sonuçlar karşılaştırılmış ve YSA'nın RA'ne göre daha iyi model ürettiği gözlemlenmiştir.

2. İKİNCİ EL OTOMOBİL PAZARI

Önceden de belirtildiği gibi, dünya otomotiv endüstrisindeki toplam motorlu araç üretiminin yaklaşık % 70'ini otomobiller oluşturmaktadır. 2006 yılında Avrupa' da üretilen tüm motorlu araçların % 87' sini de yine otomobiller oluşturmuştur (Web-ACEA, 2007). Bu verilerden hareketle, otomobil sektörü için "otomotiv endüstrisinin temel yapı taşı" nitelemesi yapılabilir. Otomobil üretimi, güçlü yan sanayisi ile birlikte diğer taşıtların üretimine de destek olmaktadır (Onat, 2007; Web-Kobifinans, 2008). Günümüzde, ikinci el otomobil pazarının potansiyelini fark eden neredeyse tüm büyük otomobil üreticileri, şimdiden bu pazara girmişler ya da girme hazırlığı yapmaktadırlar. Bu da ticari açıdan ikinci el otomobil pazarının cazibesini ortaya koymaktadır.

İkinci el otomobil satış hacminin pek çok ülkede yeni otomobil satış hacminden çok daha büyük hale gelmiş olması, ikinci el araç pazarlarının artan önemini ortaya koymaktadır.



Örneğin ABD ekonomisinde en büyük perakende sektörünü oluşturan ikinci el otomobil pazarının hacmi, yeni otomobil pazar hacminin iki katından daha fazladır ve sürekli olarak büyümeye devam etmektedir (Lee, 2006: 120). 2005 yılında ABD’ de satılan ikinci el otomobil sayısı 44 milyon, satış değeri ise 370 milyar dolardır (Web-Wikipedia, 2008).

Avrupa ikinci el otomobil pazarı 2003-2006 yılları arasında yavaş fakat istikrarlı bir büyüme göstermiştir. Bu gidişatın sonraki beş yılda da devam edeceği beklenmektedir. Pazar değeri, 2003-2006 yılları arasındaki yaklaşık % 2 civarındaki bileşik yıllık büyüme oranıyla 2006 sonunda 263 milyar dolar değerine ulaşmıştır. 2007 – 2012 arasında % 11,3 büyüyerek 298,5 milyar dolar değerine ulaşacağı tahmin edilmektedir (Datamonitor, 2007).

Gelişmiş ülkelerde internet, ikinci el otomobil satışlarında çok önemli pazar haline gelmiştir. Neredeyse her vatandaşın hızlı ve ucuz internet erişimi olması nedeniyle ikinci el otomobil satıcıları ve alıcıları internet ortamını aktif olarak kullanmaktadır. Avrupa’ daki yüzlerce ikinci el otomobil e-ticaret sitesinin içerisinde en popüler olanları şunlardır: Autoscout24.de (Almanya), Mobiliti.de (Almanya), Paruvendu.fr (Fransa), Viaturasonline.pt (Portekiz), Segundamano.es (İspanya), DerStandard.at (Avusturya), Autoscout24.ch (İsviçre).

3. REGRESYON ANALİZİ

RA bir bağımlı (açıklanan) değişken ile en az bir bağımsız (açıklayıcı) değişken arasındaki ilişkiyi matematiksel bir denkleme dönüştürmek için kullanılan bir istatistiksel analiz yöntemidir. Ortaya çıkarılan matematiksel denklem, “regresyon denklemi” olarak adlandırılır. RA’nde amaç her bağımsız değişkenin bağımlı değişkendeki değişime katkısının hesaplanması, yani bağımsız değişkenlerin değerinden hareketle bağımlı değişkenin değerinin tahmin edilmesidir. Böylece ilgilenilen değişken üzerine yapılacak plânlarda hangi faktörlerin ne derecede etkisinin olduğunu önceden bilme imkânı elde edilmiş olur (Xu, 2003).

RA, kullanılan bağımsız değişken sayısına göre;

- Basit RA (Tek bağımsız değişken),
- Çoklu RA (Birden çok bağımsız değişken),

fonksiyon tipine göre;

- Doğrusal RA,
- Doğrusal olmayan (eğrisel) RA,

verilerin kaynağına göre;

- Anakütle verileriyle RA,
- Örneklem verileriyle RA



olmak üzere üç ana grupta toplanabilir (Orhunbilge, 2002, s.5-12).

Çalışmanın uygulama bölümünde çok değişkenli doğrusal regresyon yöntemi kullanıldığı için aşağıda sadece basit ve çoklu doğrusal regresyon yöntemleri kısaca açıklanmıştır

3.1. Basit Doğrusal Regresyon

Aralarında doğrusal ilişki bulunan bir bağımlı ve bir bağımsız değişken arasındaki ilişkiyi çözümleyen RA türüdür. Basit doğrusal regresyon modeli aşağıdaki denklemde görülmektedir.

$$Y = b_0 + b_1 X + e$$

Bu denklemdeki;

Y : Bağımlı değişkeni,

X : Bağımsız değişkeni,

b_0 : Doğrunun $X=0$ olduğunda y eksenini kestiği noktayı,

b_1 : Doğrunun eğimini, diğer adıyla regresyon katsayısını,

e : Hata değerini, diğer adıyla artıkları (ya da kalıntıları)

göstermektedir. Regresyon katsayısı b_1 , bağımsız değişkendeki bir birimlik değişiminin bağımlı değişkende ne miktarda bir değişim oluşturduğunu gösterir.

3.2. Çoklu Doğrusal Regresyon

Çoklu RA, birden çok sayıdaki bağımsız değişkenden yola çıkarak bağımlı değişkenin tahmin edilmesi ve hangi bağımsız değişkenlerin bağımlı değişken üzerinde ne kadarlık etkiye sahip olduğunun bulunması için kullanılır (Xu, 2003).

Çoklu doğrusal regresyon modeli aşağıdaki denklemde görülmektedir. Bu denklemdeki X_n , n nolu bağımsız değişkeni ve bu değişkenin regresyon katsayısını göstermektedir.

$$Y = b_0 + b_1 X_1 + b_1 X_1 + \dots + b_n X_n + e$$

Regresyon denklemlerindeki regresyon katsayıları en küçük kareler yöntemi kullanılarak hesaplanır. Oluşturulan regresyon modelinin veriyi ne kadar iyi açıkladığının ölçülmesi için 0-1 aralığında yer alan R , R^2 ve düzeltilmiş R^2 değerleri kullanılmaktadır. Bu değerler aşağıda kısaca açıklanmıştır (Web-Sakarya; Xu, 2003):

R : Bağımlı değişkenle bağımsız değişken arasındaki korelasyonu temsil eden bu değer yüksek olması bağımsız değişkenle bağımlı değişken arasında sıkı bir ilişkinin



olduğunu veya bağımsız değişkenin bağımlı değişkendeki değişimin önemli bir kısmını açıkladığını göstermektedir.

R^2 : Açıklayıcılık katsayısı olarak da bilinen bu değer, bağımlı değişkendeki değişimin yüzde kaçının bağımsız değişken tarafından açıklandığını ifade etmektedir.

Düzeltilmiş R^2 : Regresyon modeline ilave edilen her yeni bağımsız değişken bağımlı değişkeni açıklama yeteneği olsun veya olmasın R^2 değerinin büyümesine sebep olarak sanki modelin açıklayıcılığının iyileştiği izlenimi vermektedir. Bu sebeple modele görece katkısı az olan veya katkısı olmayan değişkenlerin sebep olduğu R^2 değerindeki yapay artışın dikkate alınarak modelin açıklayıcılığının belirlenmesi gerekmektedir. Düzeltilmiş R^2 değeri bu düzeltme sonucunda ortaya çıkan bir değer olup, regresyon modelinin gerçek açıklama gücünü göstermektedir. Böylece, daha gerçekçi bir regresyon modelinin oluşturulması ve R^2 değerinin hesaplanması mümkün olmaktadır. Genellikle R^2 değerinden biraz daha küçük değere sahip olan düzeltilmiş R^2 , model uygunluğu ölçümünde göreceli olarak daha iyi bir performansa sahiptir ve birçok istatistik yazılımı tarafından tercih edilmektedir.

F Testi: Regresyon modelinin anlamlı olup olmadığını incelemek için ANOVA (Analysis of Variance) testi uygulanmaktadır. ANOVA testi sonucunda ortaya çıkan F değerine karşılık gelen anlamlılık seviyesi, oluşturulan modelin uygun olup olmadığının kararında yardımcı olmaktadır. F test değerinin anlamlı olması (yani 0,05'ten küçük olması) söz konusu modelin bağımlı değişkeni açıklamada önemli katkı sağladığına işaret etmektedir.

4. YAPAY SINIR AĞLARI

YSA, karmaşık yapıda içsel ilişkilere sahip problemlere çözüm bulabilmek için geliştirilmiş bir yapay zekâ tekniğidir (Çakar vd., 1996: 77). Yapay zekâ teknikleri, karmaşık ilişkilerin modellenmesindeki üstünlükleri ve veriler hakkında ön kabullenmelere ihtiyaç duymamaları sebebiyle son yıllarda çok büyük popüleriteye sahip olmuştur.

Haykin (1999), YSA için şu tanımı kullanmaktadır: “Bir sinir ağı, basit işlem birimlerinden oluşan, deneysel bilgileri biriktirmeye yönelik doğal bir eğilimi olan ve bunların kullanılmasını sağlayan yoğun bir şekilde paralel dağıtılmış bir işlemcidir. Bu işlemci iki şekilde beyin ile benzerlik göstermektedir: 1. Bilgi, ağ tarafından bir öğrenme süreciyle çevreden elde edilir. 2. Elde edilen bilgileri biriktirmek için sinaptik ağırlıklar olarak da bilinen nöronlar arası bağlantı güçleri kullanılır.”

İnsan beyninden esinlenerek ortaya çıkarılan YSA, deneme yolu ile öğrenme ve genelleştirme yapabilmektedir.

4.1. Yapısı

Bir yapay sinir ağı, kendisini oluşturan yapay sinir hücrelerinin birbirleriyle çeşitli katmanlar içerisinde paralel bağlantılar kurarak bir araya gelmelerinden oluşmaktadır. Yapay sinir hücreleri, biyolojik sinir sisteminin doğal yapısının matematiksel olarak modellenmesiyle oluşturulmaya çalışılmaktadır. Bu şekilde modellenen bir yapay sinir hücresi Şekil 1’ de görülmektedir (Aydın, 2005). Yapay bir sinir hücresinde girdiler (X_1 ,



X_2, \dots, X_n) kendilerine ait ağırlıklarla çarpılarak toplandıktan sonra aktivasyon fonksiyonuna gönderilirler. Aktivasyon fonksiyonuna gelen sinyal, burada işlenerek çıktıya dönüştürülür (Öztemel, 2003).

Yapay sinir hücreleri, katmanlar içerisinde paralel şekilde bir araya gelerek YSA' nı meydana getirmektedir. YSA' nda üç ana katman bulunmaktadır: girdi katmanı, gizli katman ve çıktı katmanı. Girdilerden oluşan girdi katmanı ile çıktılardan oluşan çıktı katmanı tektir. Gizli katman sayısı ise YSA' nın yapısına göre bir veya birden çok olabilmektedir. Girdi katmanından alınan girişler, girdi katmanı ve gizli katman arasında bulunan bağlantı ağırlıkları ile çarpılıp gizli katmana iletilmektedir. Gizli katmandaki nöronlara gelen girişler toplanarak aynı şekilde gizli katman ile çıktı katmanı arasındaki bağlantı ağırlıkları ile çarpılarak çıktı katmanına iletilir. Çıktı katmanındaki nöronlar da, kendisine gelen bu girişleri toplayarak uygun bir çıkış üretirler (Efe ve Kaynak, 2000). Bağlantıların ağırlık değerleri öğrenme sürecinde belirlenmektedir. Şekil 2' de bir yapay sinir ağını oluşturan katmanlar görülmektedir.

YSA, hücrelerin bağlanma biçimlerine göre “ileri beslemeli” ve “geri beslemeli” olmak üzere iki mimari yapı altında sınıflandırılabilir (Slaughter, 2003):

İleri beslemeli ağlar: Verilerin sadece girdi birimlerinden çıktı birimlerine ileri doğru aktığı ağ yapısıdır. Bu yapıda nöronlar katmanlar şeklinde düzenlenir. Bir katmandaki nöronların çıkışları bir sonraki katmana ağırlıklar üzerinden giriş olarak verilir. Aynı katmandaki nöronlar arasında veya bir önceki katmana bağlantı, yani geri besleme çevrimi yoktur. Uygulamalarda genellikle bu tip ağların tercih edildiği görülmektedir.

Geri beslemeli ağlar: Veri akışının sadece ileriye doğru değil geriye doğru da olabileceği ağ yapısıdır. Bu yapıda en az bir tane geri besleme çevrimi bulunur. Geri besleme, aynı katmandaki hücreler arasında olabileceği gibi farklı katmanlardaki nöronlar arasında da olabilir.

4.2. Süreçleri

YSA iki temel süreci içermektedir; İşlem ve eğitim-öğrenim süreci (Hosein, 2002):

4.2.1. İşlem Süreci

İşlem sürecinde toplama ve aktivasyon adlarında iki fonksiyon yer almaktadır.

Toplama (Birleştirme) Fonksiyonu, yapay sinir hücresine gelen girdilerin kendilerine ait ağırlıklarla çarpıldıktan sonra birleştirilmesi işlemini gerçekleştiren fonksiyondur. Bu fonksiyon, adından da anlaşılacağı gibi, genelde toplama işlemini kullanmakla birlikte farklı işlemleri de kullanabilir. Hatta araştırmacının kendi kurduğu işlemi de kullanması mümkündür. Toplama fonksiyonunda kullanılan işlem, genellikle seçilen ağ mimarisine de bağlıdır. En sık kullanılan fonksiyonlar şunlardır; Toplama, çarpım, maksimum, minimum, çoğunluk, kümülatif toplam (Öztemel, 2003).



Aktivasyon Fonksiyonu, yapay sinir hücresinin çıktısının büyüklüğünü sınırlandıran fonksiyondur. Bazı kaynaklarda transfer, eşik veya sıkıştırma fonksiyonu olarak da isimlendirilmektedir (Mandic ve Chambers, 2001). Bir ağdaki tüm hücrelerin aktivasyon fonksiyonu birbirinden farklı olabilir. Aktivasyon fonksiyonunda doğrusal fonksiyonlar genelde tercih edilmez. Zaman serileri için “Sigmoid”, ikili (binary) değişkenler için “Adım” fonksiyonu önerilmektedir (Tebelkis, 1995). En çok kullanılan aktivasyon fonksiyonları şunlardır: Doğrusal, adım, eşik değer, hiperbolik tanjant, sigmoid (Demuth ve Beale, 2000; Şen, 2004):

4.2.2. Eğitim ve Öğrenim Süreci

YSA’ nın en önemli özelliklerinden birisi, ilgili probleme ait örnekler verildiğinde girdiler ile çıktılar arasındaki ilişkiyi öğrenebilmesidir. YSA’ nın diğer bir süreci olan eğitim ve öğrenim sürecinin temel amacı girdi ve çıktı arasında bir eşleştirme yapmaktır. Danışmanlı öğrenme olarak adlandırılan bu yöntem ile ağ yeterince eğitildiğinde daha önceden görmediği girdilere karşı uygun çıktılar üretebilmektedir. Bu özellik YSA’ nın genelleme kapasitesi olarak adlandırılır ve çoğunlukla ağı yapısına ve eğitim için uygulanan yineleme sayısına bağlıdır (Alkan, 2001).

YSA’ nın veri yapısındaki ilişkiyi öğrenmesi, probleme ait örnekler yardımı ile ağ ağırlıklarının en uygun değerlerinin belirlenmesine dayanır. Herhangi bir ağırlığın yeni değeri, eski değeri ile ağırlık değerlerinin değişim miktarının (ΔW) toplamı olarak belirlenir ($W_{\text{yeni}} = W_{\text{eski}} + \Delta W$). Belli bir kurala göre hesaplanan bu değişim miktarını (ΔW) belirlemek için tanımlanmış kurallara “öğrenme algoritmaları” denir.

YSA’ nın eğitimi için birçok değişik optimizasyon yöntemi bulunmakla birlikte optimum sonucu garanti eden genel bir algoritma bulunmamaktadır. YSA’ nın eğitiminde hatayı en aza indirmek için genellikle, hata fonksiyonunun yönünü bulmaya ve çıktı değerini azaltmaya çalışan dereceli azaltma (gradient descent) tabanlı algoritmalar kullanılır. Dereceli azaltma algoritmaları içerisinde en bilinen ve uygulamalarda en çok tercih edilen algoritma “Geri Yayılım Algoritması (GYA)” dır (Auclair, 2004).

4.3. Geri Yayılım Algoritması

Danışmanlı öğrenen, ileri beslemeli Geri Yayılım Ağları (Back Propagation Networks), adını hatayı yayma biçiminden alır. Hata azaltma işlemini hatayı tüm ağa yayarak gerçekleştirdiği için bu ağa “Hatayı Geriye Yayma – Error Back Propagation” da denmektedir. Bu algoritmada elde edilen çıktı ile olması gereken çıktı arasındaki fark, yani hata tüm ağırlıklara yansıtılarak dereceli olarak minimum düzeye indirilmeye çalışılır (Aydoğmuş ve Çöteli, 2005). Geri yayılım algoritmasında eğitime rastgele bir ağırlık kümesi ile başlanır, birçok uygulamada ağı başarısı ağa atanan ilk ağırlık değerlerinin uygun seçilmesine bağlıdır (Nabiyev, 2003).

GYA, öğrenmesi ve uygulaması kolay bir modeldir. Günümüzdeki uygulamaların yaklaşık % 80’inde kullanıldığı tahmin edilen bu ağ, özellikle mühendislik alanında neredeyse her türlü problemin çözümü için kullanılabilir kadar güçlüdür (Kaastra ve Boyd,



1996). GYA, finansal zaman serilerinin gelecekteki davranışlarını tahmin etmede de sıklıkla tercih edilmektedir.

5. UYGULAMA

İkinci el otomobil pazarları ile yeni otomobil pazarları arasında çok yakın bir ilişki olduğu bilinmektedir. Her ne kadar ikinci el otomobil pazarları yeni otomobil pazarlarına belirli bir oranda bağımlı olsa da, yine de kendilerine özgü birtakım özellikleri mevcuttur. Genel olarak, tedarikleri esnek olmayan ve pazar dışı faktörlerden etkilenen ikinci el otomobillerin talep edilmeleri en az yeni araçlardaki kadar esnektir. Bu nedenle de, ikinci el otomobillerin fiyatları yeni otomobil fiyatları ile karşılaştırıldığında, çok daha değişken oldukları görülmekte, buna bağlı olarak da tahmin edilmeleri zorlaşmaktadır (Lee, 2006: 120-121). Bu tahminleri yapabilmek için her şeyden önce kapsamlı ve nitelikli bir veri toplama işlemi gerçekleştirilmelidir. İnternet, bu konuda veri toplamayı kolaylaştırmış olsa da her gün milyonlarca ikinci el otomobilin sergilendiği yüzlerce web sitesinin sayfalarındaki ham verileri elde etmenin ve yapısal hale getirmenin zorluğu, bu konuda yapılabilecek çalışmalarındaki en önemli sıkıntıyı oluşturmaktadır. Bu çalışmanın uygulamasında kullanılan verilerin internetten toplanarak yapısal hale getirilmiş olması, çalışmanın önemini artırmaktadır.

Bu çalışmadaki uygulamanın aşamaları ve sonuçları aşağıda özet olarak anlatılmıştır. Uygulamada oluşturulan modellerin doğruluğunu ölçmek için literatürde yaygın olarak kullanılan üç istatistiğe başvurulmuştur: Ortalama Mutlak Hata (OMH) (Mean Absolute Error, MAE), Ortalama Mutlak Yüzde Hata (OMYH) (Mean Absolute Percentage Error, MAPE), Ortalama Hata Kareleri Kökü (OHKK) (Root Mean Squared Error, RMSE).

5.1. Verilerin Toplanması

Çalışmada Avrupa’ da ikinci el otomobil ilanı yayınlayan yaklaşık 40 adet web sitesinden sistematik olarak toplanmış olan veriler kullanılmıştır. Bu sitelerde yer alan veriler şu 8 ülkeyi içermektedir: Almanya, Fransa, İtalya, Portekiz, İspanya, Avusturya, Hollanda ve İsviçre. 2007 yılının Aralık ayında elde edilmiş olan verilerin ülkelere göre dağılımı Tablo 1’ de görülmektedir.

5.2. Verinin Hazırlanması

Çalışmadaki en fazla zaman ve enerjinin harcandığı bu aşamada; veri toplama, birleştirme, temizleme ve dönüştürme işlemleri neticesinde verilerin analiz ve modelleme çalışmasına uygun hale getirilmesi hedeflenmiştir. Bu amaçla öncelikle web sayfalarından düz metin olarak toplanmış olan veriler yapısal sorgulama dili (SQL, Structured Query Language) kullanılarak veritabanlarındaki uygun tablolara aktarılmıştır. Bu aşamada; örneğin herhangi bir ilanda rastlanan bir sözcük, bir araç markası ile eşleşiyor ise bu sözcük ilgili tablonun o ilana yönelik satırında BRAND (marka) alanına aktarılmış, şayet sayısal bir değer



olup yanında para birimi bulunuyorsa (örneğin Euro sözcüğü ya da € simgesi), PRICE (fiyat) alanına aktarılmıştır. Bu işlemler, otomobillerin değerlendirmeye alınan alanan tüm özellikleri benzer yöntemlerle uygulanmıştır.

Verilerin tablolara aktarılması sonrasında elde edilen kayıtlarda gözlemlenen sorunlar yine SQL kodları yardımıyla giderilmeye çalışılmıştır (Örneğin bir aracın üretim yılının 1900 olarak girilmesi gibi mantıksal hatalar). Sorunların giderilmesi sırasında, Avrupa’ da üretilen neredeyse tüm araçların detaylı teknik bilgilerini içeren bir “model bankası” referans olarak kullanılmıştır. Sorunların çözümleri fazla miktarda teknik bilgi gerektirdiği için burada anlatımına gerek duyulmamış, sadece sık rastlanan sorunlardan bir kısmı aşağıda ismen listelenmiştir:

- Tekrar eden kayıtlar (aynı ilanın farklı sitelerde yer alması nedeniyle)
- Yazım hataları
- Mantıksal hatalar
- Eksik bilgi
- Birim farkları

5.3. Veri Kümesi ve Model Bankası

Veriler SQL komutlarıyla veri madenciliğine uygun hale getirildikten sonra 1.253.000 kayıtlık veri kümesi elde edilmiştir. Bu büyüklükte veri kümesini işlemede, eldeki donanım ve/veya yazılımların yetersiz kalması üzerine, her ülkeden rastgele 1000 adet kayıt alınarak toplamda 8000 kayıttan oluşan bir veri kümesi kullanılmıştır.

Daha önce de belirtildiği gibi, çalışmada bir “model bankası” referans amacıyla kullanılmıştır. 48 adet otomobil markasına ait toplam 20.000 otomobil modelinden oluşan bu referans modeli, otomobil üreticisi firmalar tarafından belirli periyotlarla güncellenerek yayınlanan bilgilerin toplamından elde edilmiş olup bu üreticilere ait tüm otomobil modellerini ve ayrıntılı özelliklerini içermektedir. İlanlardan elde edilen özellikler arasında bir modeli benzersiz olarak tanımlayan 6 temel özellik (Marka, Temel Model, Detay Model, Kasa Tipi, Kapı Sayısı, Motor Gücü) kullanılarak ilanın model bankasındaki karşılığı tespit edilmekte ve ilanlarda bildirilen özellikler ile birleştirilerek veri madenciliği analizleri için oluşturulan veri ambarına aktarılmaktadır. Bu veri ambarında, her otomobile ait 100’ ün üzerinde özellik tutulmaktadır. Ancak bu özelliklerin büyük çoğunluğu otomobil fiyatı ile doğrudan ilişkili olmayıp bilgi amaçlı saklanmaktadır (Örneğin araç modellerinin hangi tarihte üretime başlandığı bilgisi gibi). Bu nedenle, otomobil fiyatları ile doğrudan ilgisi olacağı düşünülen özellikler sezgisel olarak belirlenip analizlerin yapıldığı tablolara aktarılmıştır. 17 adet temel, 32 adet opsiyonel özellikten oluşan bu yapı Tablo 2’ de gösterilmiştir.

5.4. Regresyon Analizi Sonuçları



<http://www.akademikbakis.org>

Optimum regresyon modelinin oluşturulması için öncelikle modele anlamlı katkısı olan değişkenlerin belirlenmesine çalışılmıştır. Bunun için SPSS ile gerçekleştirilen ilk RA sonucu oluşan modelin özeti Tablo 3’ de görülmektedir. Bu tablo, otomobil fiyatlarının % 80’ e yakın bölümünün bağımsız değişkenler tarafından açıklanabildiğini göstermektedir. Bu yüksek oran, ikinci el otomobil fiyatlarının tahminine yönelik olarak gerçekleştirilen veri toplama ve hazırlama işlemlerinin başarısını göstermektedir. Aynı zamanda, tüm veriler içerisinde fiyatlarla doğrudan ilişkisi olduğu düşünülen değişkenlerin sezgisel olarak belirlenmesinin de oldukça isabetli yapılmış olduğu söylenebilir.

Bu regresyon modelinin anlamlı olup olmadığını anlamak için incelenen ANOVA testi sonuçları Tablo 4’ de yer almaktadır. Bu testin sonucunda ortaya çıkan F değerine karşılık gelen p anlamlılık seviyesinin 0,05’ten çok küçük olması nedeniyle, söz konusu regresyon modelinin açıklayıcılığının istatistiksel açıdan önemli olduğunu ifade edebiliriz.

Daha sonra, mevcut modelin değişkenlerinin modele olan katkılarını belirlemek ve anlamlı katkısı olmayan değişkenleri modelden çıkarmak amacıyla, RA’nden elde edilmiş olan katsayılar tablosu incelenmiştir. Tabloda, fiyatı ters yönde etkilemesi beklenen yaş (age), kat edilen mesafe (mileage) gibi önemli değişkenlerin negatif işaretli çıkması, Hair vd.’nin (1998) ifade ettiği gibi, regresyon modelinin anlamlılığının diğer bir göstergesi olmuştur. Katsayılar tablosunun incelenmesinden sonra, şu değişkenlerin modele anlamlı katkısının olmadığı anlaşılmıştır: Cylinders_No (Silindir Sayısı), Color (Renk), Zip_Code (Posta Kodu), Heated_Seats (Isıtmalı Koltuklar), Alloy_Rims (Alüminyum Jantlar), Electric_Windows (Elektrikli Pencereler), Power_Steering (Hidrolik Direksiyon), Alarm_Device (Alarm Cihazı), Cd_Changer (CD Değiştirici), Cooling_Box (Buzdolabı), Night_Vision (Gece Görüşü). Daha duyarlı bir regresyon denklemi elde edebilmek amacıyla, bu değişkenler elimine edilerek yeniden RA gerçekleştirilmiştir. Bu analiz sonucu elde edilen model özeti; otomobil fiyatlarının % 80’ inin bağımsız değişkenler tarafından açıklanabildiğini gösterirken, ANOVA testi sonuçları da yine modelin istatistiksel geçerliliğini ortaya koymuştur. Yeni modeldeki değişkenlere ait katsayılar tablosu kullanılarak oluşturulan regresyon denklemi, diğer bir deyişle fiyat modeli aşağıda görülmektedir:

$$\begin{aligned} \text{FİYAT (PRICE)} &= \\ 14987.813 * 1 (\text{sabit}) &+ \\ -0.041 * \text{Model_Name_Detail_No} &+ \\ 143.990 * \text{Body_Type_No} &+ \\ -519.552 * \text{Doors_No} &+ \\ 205.472 * \text{Engine_Power} &+ \\ -1.315 * \text{Cylinder_Capacity} &+ \\ 2807.328 * \text{Fuel_Type_No} &+ \\ -0.034 * \text{Mileage} &+ \\ -93.128 * \text{Age} &+ \\ 108.567 * \text{Country} &+ \\ 882.573 * \text{Automatic_Gears} &+ \\ -627.482 * \text{Leather_Seats} &+ \end{aligned}$$



<http://www.akademikbakis.org>

1209.559	* Navigation	+
-1741.538	* Air_Conditioner	+
575.335	* Xenon_Lights	+
668.277	* Sunroof	+
-542.407	* Board_Computer	+
2399.647	* Parking_System	+
-633.713	* Cruise_Control	+
-808.711	* Elbow_Rest	+
1513.273	* Electric_Seats	+
-476.520	* Metallic_Lacquering	+
-697.958	* ESP	+
1122.684	* Central_Door_Locking	+
-4059.457	* ABS	+
-11362.311	* Airbags	+
-713.992	* Immobiliser	+
627.288	* Electric_Outside_Mirrors	+
1788.594	* Multifunctional_Steering_Wheel	+
-886.279	* Fog_Lamp	+
1307.844	* Cd_Player	+
703.422	* Rain_Sensor	+
3058.928	* TV	+

5.5. Yapay Sinir Ağları Uygulama Sonuçları

YSA uygulamasında SPSS ve Clementine yazılımları kullanılarak optimum parametre tespitleri yapılmaya çalışılmıştır. Tasarlanan Geri Yayılımlı YSA' nda kullanılan parametreler şu şekilde belirlenmiştir:

- *Eğitim ve test* kümeleri için sırasıyla %70-%30, %80-%20 ve %90-%10 oranları denenmiş ve en başarılı sonuçların %80-%20 oranlarında elde edildiği gözlenmiştir.
- *Girdi nöronu sayısı*: Neden-sonuç ilişkisine dayanan tahmin problemlerinde girdi nöronu sayısı, girdi vektöründeki değişken sayısına eşit olduğu için bu parametre kolaylıkla belirlenmiştir. Veri kümesindeki analizlerde kullanılan değişken sayısı 43 olduğu için bu değer girdi nöronu sayısı olarak alınmıştır.
- *Gizli katman sayısı*: Yazındaki çalışmalarda genellikle tek gizli katmanın ya da en fazla iki katmanın kullanılarak başarılı sonuçlar elde edildiği görülmüştür. Bu sayının artırılmasının, hesaplama zamanını da artırmasının yanında ağın öğrenme yerine ezberlemesine neden olduğu bilinmektedir. Bu nedenle, bu uygulamada da verilerin çokluğu göz önüne alınarak iki gizli katman kullanılmasına karar verilmiştir.
- *Gizli nöron sayısı*: Yine yazındaki çalışmalarda gizli nöron sayısının tayininde çok



belirgin bir formülün bulunmadığı, bu sayının sezgisel olarak deneme-yanılma yöntemiyle belirlendiği görülmüştür. Bu bilgilere istinaden bu çalışmada da deneme-yanılma yöntemiyle optimum sayının bulunmasına çalışılmıştır.

- *Çıktı nöronu sayısı:* Çıktı nöronu sayısı neden sonuç ilişkisine dayalı problemlerde çıktı sayısına eşit olarak alındığı için bu parametre kolaylıkla belirlenerek 1 olarak alınmıştır.
- *Aktivasyon Fonksiyonu ve Öğrenme Algoritması:* Yazındaki yaygın kullanımlar dikkate alınarak, aktivasyon fonksiyonu olarak Sigmoid fonksiyonu, öğrenme algoritması olarak geri yayılım (back-propagation) algoritması kullanılmıştır. Geri yayılım algoritmalarında ağırlıkların değişim büyüklüğünü belirleyen öğrenme oranı (katsayısı) ile eğitim sürecini hızlandırmayı sağlayan momentum katsayısının belirlenmesi için yazında yine belirli bir formül bulunmadığı ve deneme-yanılma yönteminin kullanıldığı görülmüştür. Bu çalışmada da benzer yöntemle optimum parametrelerin bulunması yoluna gidilmiştir.

Bu değerlendirilmelerin sonucunda ortaya çıkan YSA yapısı Tablo 5’ de özetlenmiştir.

5.6. Modellerin Değerlendirilmesi ve Karşılaştırması

Modellerin tahmin doğruluğunu saptayabilmek ve bir başarı kıyaslaması yapabilmek için RA’nden elde edilen değerlere ait doğruluk ölçüleri, ile YSA uygulaması sonucu elde edilen değerlere ait doğruluk ölçüleri karşılaştırmalı olarak Tablo 6’ da görüldüğü gibi incelenmiştir.

Bilindiği gibi, tahmin hatası ile tahmin doğruluğu ters orantılıdır, yani tahmin hatası küçüldükçe modelin doğruluk derecesi artmaktadır. Bu nedenle, farklı tahmin modellerini karşılaştırırken tüm istatistiklerde en küçük hata değerine sahip model en iyi model olarak seçilmektedir. Bir modelin doğruluğunu test etmek için OMH, OMYH ve OHKK değerlerini ayrı ayrı incelemek faydalı olacaktır, ancak hataları yüzde olarak ifade etmesi nedeni ile tek başına da bir anlam taşıyan OMYH istatistiği referans alınıp modellerin doğruluk seviyelerini belirlemede ve karşılaştırmada kullanılabilir. OMYH değeri % 10’un altında olan modeller “çok iyi”, % 10 ile % 20 arasında olan modeller “iyi”, % 20 ile % 50 arasında olan modeller “kabul edilebilir” ve % 50’nin üzerinde olan modeller ise “yanlış ve hatalı” olarak kabul edilmektedir (Akgül, 2003; Lewis, 1982).

Tablo 6 incelendiğinde, çalışmada uygulanan her iki yöntemde de OMYH değeri % 50’ nin altında olduğu için iki yöntemin de “kabul edilebilir” model ürettiği görülmüştür. Diğer taraftan, ikisi yöntem arasında bir kıyaslama yapıldığında, YSA’ nın RA’nden daha başarılı sonuçlar verdiği ve genel olarak “iyiye yakın” bir model olduğu söylenebilir.

6. SONUÇ



Bu çalışmada, en sık kullanılan veri madenciliği yöntemleri arasında yer alan RA ve YSA kullanılarak ikinci el otomobillerin güncel pazar fiyatlarının modellenebileceği, özellikle de YSA' nın iyi modeller üretebileceği sonucuna varılmıştır.

İleriki çalışmalara ışık tutması beklentisiyle uygulamada yaşanan bazı kısıtlamalar ve öneriler şu şekilde özetlenebilir:

Bilindiği gibi YSA, çok sayıda parametre içermekte ve çok iyi sonuç için çok sayıda veriye ihtiyaç duymaktadır. Çalışma kapsamında her ne kadar çok sayıda veri elde edilmiş olsa da, eldeki bilgisayar donanımlarının yetersiz seviyede olması nedeniyle çok küçük bir bölümü analizlere dahil edilebilmiştir. Bu nedenle, daha güçlü donanım kullanılarak daha fazla veri ile aynı uygulama gerçekleştirildiğinde çok daha başarılı sonuçlar alınacağı rahatlıkla ileri sürülebilir.

Girdi olarak kullanılan fiyat verileri gerçek satış fiyatlarından değil, web ilanlarından elde edilen arz fiyatlarından oluşmaktadır. Bu da oluşturulan modellerin başarı seviyesini kısmen olumsuz etkilemiştir. Ancak, kapsamlı gerçek satış fiyatlarını edinmek pratikte mümkün görünmediği için mevcut verilerle elde edilen sonuçların da oldukça faydalı olduğu ifade edilebilir.

Ayrıca, ikinci el otomobillerin bir ülkenin farklı bölge ya da şehirlerinde farklı fiyatlara sahip olduğu bilinen bir gerçektir. Çalışmada; internette yer alan ilanların çoğunda şehir adının değil yerel alan kodunun (Zip_code) bulunduğu gözlemlenmiştir. Bu kodlar; mahalle, semt, cadde gibi yerel alanları gösterdiği için doğal olarak analizlerde önemli bir etkisi görülebilmiştir. Örneğin aynı şehirdeki farklı mahalle ya da semtlerde bulunan otomobillerin fiyat ilişkileri ortak bir şehir kodu altında incelenememiştir. Verilerin toplandığı ülkelerdeki alan kodu farklılıkları nedeniyle bu kodlardan, benzersiz (unique) şehir ya da bölge kodlarının kolayca elde edilemeyeceği ve çözümü için ek bir çalışma yapılması gerektiği anlaşıldığı için, çalışma kapsamında zaman kısıtı nedeniyle bu ayrıştırma işlemine girilmemiştir. İleriki çalışmalarda bu işlemin gerçekleştirilmesi, modelin açıklayıcılığını artıracak gibi şehirlerin ya da bölgelerin fiyatlar üzerindeki etkisini görme fırsatı da sunacaktır.

Bunlara ilaveten, ikinci el otomobil fiyatlarının modellenmesine yönelik veri madenciliği çalışmalarında tahmin doğruluğunu artırmak amacıyla farklı yöntemlerin birlikte kullanıldığı melez (hybrid) yöntemlerin kullanılması önerilebilir. Örneğin güncel pazar fiyatlarının tahmin edilmesine yönelik uygulamalarda RA ile modele anlamlı katkısı olan değişkenler belirlenip YSA' nın girdisi olarak sadece bu değişkenler kullanılarak YSA ile çok daha başarılı modeller oluşturulabilir.

Sonuç olarak, YSA' nın kara kutu (black box) olarak algılanmalarına rağmen, tasarımlarının kolay olması, probleme hızlıca adapte olabilmeleri ve az sayıda veriye rağmen başarılı sonuçlar üretebilmeleri nedeniyle pek çok alanda olduğu gibi ikinci el otomobil sektöründe de kullanılmaları tavsiye edilebilir.



KAYNAKÇA

1. ALKAN A., Predictive Data Mining with Neural Networks and Genetic Algorithms, Ph.D. Thesis, İTÜ, İstanbul, 2001, 51.
2. ASILKAN Ö., IRMAK S., “İkinci El Otomobillerin Gelecekteki Fiyatlarının Yapay Sinir Ağları ile Tahmin Edilmesi”, Süleyman Demirel University Journal of Faculty of Economics and Administrative Sciences, 2009, Vol. 2.
3. AUCLAIR A., Feed-Forward Neural Networks Applied To The Estimation of Magnetic Distributions, M.S. Thesis, Mcgill University, Montreal, Canada, 2004, 128.
4. AYDIN Ö., Yapay Sinir Ağlarını Kullanarak Bir Ses Tanıma Sistemi Geliştirilmesi, Yüksek Lisans Tezi, Trakya Üniversitesi, Edirne, 2005, 74.
5. AYDOĞMUŞ Z., ÇÖTELİ R., “Yapay Sinir Ağları Yardımıyla İzolatör Yüzeyinde Potansiyel Tahmini”, F.Ü. Fen ve Müh.Bil.Dergisi, 17/2, 2005, 239-246.
6. BOX G., JENKINS M., Time Series Analysis Forecasting and Control, Holden Day Inc., California, 1976.
7. ÇAKAR T., TÜRKER A. K., TORAMAN A., “İmalat Sistemlerinin Tasarlanmasında Yapay Sinirsel Ağların Kullanılması”, Birinci Ulusal Zeki İmalat Sistemleri Sempozyumu Zis’96 Bildirisi, Sakarya, Mayıs 1996, 10.
8. DATAMONITOR, “Used Cars in Europe”, Industry Profile, Reference Code: 0201-0750, December 2007.
9. DEMUTH H., BEALE M., Neural Network Toolbox for Use with Matlab, MathWorks Inc., 2000.
10. EFE Ö., KAYNAK O., Yapay Sinir Ağları ve Uygulamaları, Boğaziçi Üniversitesi Yayınları, Basım No: 696, İstanbul, 2000, 148.
11. FIRAT M., GÜNGÖR M., “Askı Madde Konsantrasyonu ve Miktarının Yapay Sinir Ağları ile Belirlenmesi”, İMO Teknik Dergi, 2004, 3267-3282.
12. HAIR J.F., ANDERSON R.E., TAHTAM R.L., BLACK W.C., Multivariate Data Analysis, Pearson Education, New Jersey, 1998.
13. HAYKIN S., Neural Networks: A Comprehensive Foundation, Prentice Hall International Inc., 1999.
14. HOSEIN P., “Stock Price Prediction By Artificial Neural Networks: A Study of Tehran’s Stock Exchange”, Management Studies Quarterly Journal of Management & Accounting School, No.31&32, 2002.
15. İŞERİ A., KARLIK B., “An Artificial Neural Networks Approach on Automobile



- Pricing”, Expert Systems with Applications, January 2008.
16. KAASTRA I., BOYD M., “Designing A Neural Network for Forecasting Financial and Econometric Time Series”, Neurocomputing, Vol. 10, 1996, 215-236.
 17. LEE J., Empirical Analysis of Wholesale Used Car Auctions, A Dissertation in University of California, 2006, 120-121.
 18. MANDIC D. P., CHAMBERS J. A., Recurrent Neural Networks for Prediction - Learning Algorithms Architectures and Stability, John Wiley & Sons Ltd, 2001.
 19. NABİYEV V.V.,Yapay Zeka, Seçkin Yayıncılık, Ankara, 2003.
 20. ONAT M. G., Otomotiv Sektöründe Oranlar Yöntemi Aracılığı ile Finansal Analiz, Yüksek Lisans Tezi, Marmara Üniversitesi, İstanbul, 2007
 21. ORHUNBİLGE N., Uygulamalı Regresyon ve Korelasyon Analizi, İstanbul Üniversitesi İşletme Fakültesi Yayını, No: 281, İstanbul, 2002, 5-12.
 22. ÖZTEMEL E., Yapay Sinir Ağları, Papatya Yayıncılık, Ağustos 2003.
 23. SLAUGHTER G. E. F., Artificial Neural Network for Temporal Impedance Recognition of Neurotoxins, M.S. Thesis, Virginia Common Wealth University, Virginia, 2003, 115.
 24. ŞEN Z., Yapay Sinir Ağları İlkeleri, Su Vakfı, İstanbul, 2004.
 25. TEBELKIS J., Speech Recognition Using Neural Networks, Ph.D. Thesis, Carnegie Mellon University, Pennsylvania, Mayıs 1995.
 26. WEB-ACEA, Association Des Constructeurs Europeens D’automobiles, Economic Report, Acea’s Position On Motor Vehicle Distribution in The European Union, 2001, www.acea.be, (15/09/2007).
 27. WEB-KOBİFİNANS, www.kobifinans.com.tr/tr/sector/011001/12455, (01.06.2008).
 28. WEB-SAKARYA, www.sakarya.edu.tr/~altunr/site/ders/pa, (01.08.2008).
 29. WEB-WIKIPEDIA, en.wikipedia.org/wiki/Used_car, (11.06.2008).
 30. XU Y., Using Data Mining In Educational Research: A Comparison of Bayesian Network with Multiple Regression in Prediction, Department of Educational Psychology, The University of Arizona, Arizona, 2003.
 31. ZHANG G., PATUWO B.E., HU M.Y., “Forecasting with Artificial Neural Networks: The State of The Art”, International Journal of Forecasting, Vol. 14, No.1, 1998, 35-62.



EKLER: TABLO VE ŞEKİLLER LİSTESİ

Tablo 1. Veri kümesinin ülkelere göre dağılımı

Ülke	İlan Sayısı
Almanya	822446
Fransa	221208
İspanya	47228
İtalya	26886
Avusturya	20647
Portekiz	13194
İsviçre	7989
Hollanda	5917

Tablo 2. Örneklemin veri yapısı

No	Orijinal Alan Adı	Açıklama	Veri Türü	Min.-Maks.
1	Brand_No	Marka Numarası	Sayısal	1-69
2	Brand	Marka	Metin	
3	Model_Name_Rough_No	Temel Model Numarası	Sayısal	1-772
4	Model_Name_Rough	Temel Model	Metin	
5	Model_Name_Detail_No	Detay Model Numarası	Sayısal	1-15685
6	Model_Name_Detail	Detay Model	Metin	
7	Body_Type_No	Kasa Tipi	Kategorik	0-14
8	Doors_No	Kapı Sayısı	Kategorik	2-5
9	Engine_Power	Motor Gücü	Sayısal	29-450
10	Cylinder_Capacity	Silindir Kapasitesi	Sayısal	599-7011
11	Cylinders_No	Silindir Sayısı	Sayısal	2-12
12	Fuel_Type_No	Yakıt Tipi	Sayısal	1-9
13	Mileage	Katedilen Mesafe	Sayısal	1-350.000
14	Age	Yaş (Ay Olarak)	Sayısal	1-176



No	Orijinal Alan Adı	Açıklama	Veri Türü	Min.-Maks.
15	Color	Renk	Kategorik	0-10
16	Country	Ülke	Kategorik	1-8
17	Zip_Code	Posta Kodu	Kategorik	0-99990
18	Automatic Gears	Otomatik Vites	Kategorik	0-1
19	Leather Seats	Deri Koltuk	Kategorik	0-1
20	Navigation	Navigasyon Sistemi	Kategorik	0-1
21	Air Conditioner	Klima	Kategorik	0-1
22	Xenon Lights	Xenon Önfarlar	Kategorik	0-1
23	Sunroof	Sunroof (Üst Pencere)	Kategorik	0-1
24	Heated Seats	Isıtılmalı Koltuklar	Kategorik	0-1
25	Board Computer	Yol Bilgisayarı	Kategorik	0-1
26	Parking System	Park Sensörü	Kategorik	0-1
27	Cruise Control	Hız Kontrol Düzeni	Kategorik	0-1
28	Elbow Rest	Kolçak	Kategorik	0-1
29	Alloy Rims	Alüminyum Jantlar	Kategorik	0-1
30	Electric Seats	Elektrikli Koltuk	Kategorik	0-1
31	Electric Windows	Elektrikli Pencere	Kategorik	0-1
32	Metallic Lacquering	Metalik Boya	Kategorik	0-1
33	ESP	Esp	Kategorik	0-1
34	Power Steering	Hidrolik Direksiyon	Kategorik	0-1
35	Central Door Locking	Merkezi Kilit Sistemi	Kategorik	0-1
36	ABS	ABS Fren sistemi	Kategorik	0-1
37	Airbags	Hava Yastıkları	Kategorik	0-1
38	Alarm Device	Alarm Cihazı	Kategorik	0-1
39	Cd Changer	CD Değiştirici	Kategorik	0-1
40	Immobiliser	Immobiliser	Kategorik	0-1
41	Electric Outside Mirrors	Elektrikli Yan Aynalar	Kategorik	0-1
42	Cooling Box	Buzdolabı	Kategorik	0-1
43	Multifunctional Steering Wheel	Çok Fonksiyonlu Direksiyon	Kategorik	0-1
44	Fog Lamp	Sis Lambası	Kategorik	0-1
45	Night Vision	Gece Görüşü	Kategorik	0-1
46	Cd Player	CD Çalar	Kategorik	0-1
47	Rain Sensor	Yağmur Algılayıcısı	Kategorik	0-1
48	TV	Televizyon	Kategorik	0-1
49	Price	Fiyat	Sayısal	190-266,350

Tablo 3. Regresyon model özeti



Model	R	R ²	Düzeltilmiş R ²	Standart Hata
1	0,894	0,799	0,798	5760,504

Tablo 4: Regresyon modeli ANOVA sonuçları

Model	Kareler Toplamı	df	Ortalama Kare	F	p	
1	Regresyon	1,048E12	43	2,437E10	734,283	0,000
	Kalıntı	2,640E11	7956	3,318E7		
	Toplam	1,312E12	7999			

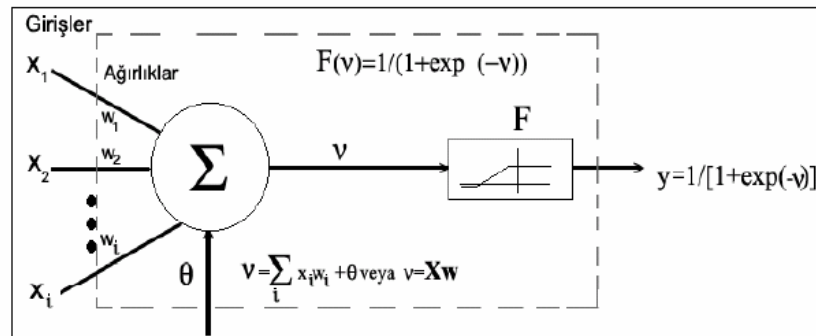
Tablo 5. YSA modeli

Girdi nöronu sayısı	Gizli katman sayısı	Gizli nöron sayısı	Çıktı nöronu sayısı:	Aktivasyon Fonksiyonu
43	2	2198	1	Sigmoid

Tablo 6. Modellerin doğruluk ölçüleri

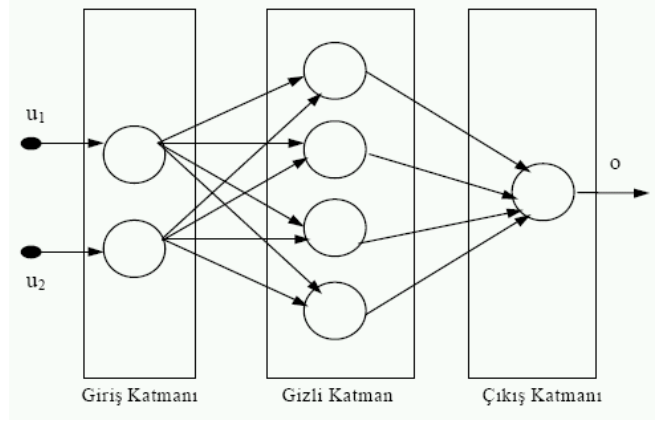
Model	OMH	OMYH	OHKK
Regresyon Analizi	3826,016	0,462	6222,049
Yapay Sinir Ağları	2785,218	0,336	4893,494

Şekil 1: Yapay sinir hücresi



(Kaynak: Aydın, 2005)

Şekil 2: İleri beslemeli sinir ağı yapısı



(Kaynak: Fırat ve Güngör, 2004)