

## FIYAT TAHMİNLEMESİNDE MAKİNE ÖĞRENMESİ TEKNİKLERİ VE DOĞRUSAL REGRESYON YÖNTEMLERİNİN KIYASLANMASI; TÜRKİYE'DE SATILAN İKİNCİ EL ARAÇ FİYATLARININ TAHMİNLENMESİNE YÖNELİK BİR VAKA ÇALIŞMASI

<sup>1</sup>Ersin NAMLI, <sup>2</sup>Ramazan ÜNLÜ, <sup>3</sup>Ecem GÜL

<sup>1,3</sup> *Istanbul Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği Bölümü, İstanbul, TÜRKİYE*  
<sup>2</sup> *Gümüşhane Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Gümüşhane, TÜRKİYE*

<sup>1</sup>[enamli@istanbul.edu.tr](mailto:enamli@istanbul.edu.tr), <sup>2</sup>[ramazanunlu@gumushane.edu.tr](mailto:ramazanunlu@gumushane.edu.tr), <sup>3</sup>[ecemmgul@gmail.com](mailto:ecemmgul@gmail.com)

(Geliş/Received: 12.12.2018; Kabul/Accepted in Revised Form: 23.05.2019)

**Öz:** İkinci el araç alım satım piyasası dünyada olduğu gibi Türkiye'de de çok hareketli ve kayda değer büyüklükte bir pazardır. Satıcılar araçları için alabilecekleri maksimum fiyatı ararken alıcılar olabildiği kadar düşük fiyata maksimum kalitede bir araç almak için uğraşırlar. Ancak söz konusu araçların alım satımı esnasında çeşitli problemler olabilmekte ve belirli bir araç için bayi düzeyinde dahi standart bir fiyatlandırma politikası uygulanmamaktadır. Bu çalışmada bu problemi çözmek adına ikinci el bir araç için fiyatlandırma politikası oluşturup oluşturulamayacağı araştırılmıştır, bu kapsamda gerçek veriler toplanarak istatistiksel ve yapay zekâ tabanlı yöntemlerle tahmin modelleri oluşturulmuştur. Yapay zekânın alt dallarından bir tanesi olarak düşünebileceğimiz makine öğrenmesi teknikleri doğal dil işleme, metin madenciliği, görüntü işleme gibi çok kompleks problemlerin yanı sıra regresyon problemlerinde de doğrusal regresyon gibi klasik yöntemlere göre daha başarılı sonuçlar vermektedir. Bu noktadan hareketle bu çalışma da ikinci el araç satışlarındaki fiyatlandırma sisteminin standardize edilebilmesi için Yapay Sinir Ağları, Destek Vektör Makineleri ve Doğrusal Regresyon yöntemleri uygulanmış ve kullanılan metotlar sıkça tercih edilen çeşitli değerlendirme açısından kıyaslanmıştır. Çalışmanın bulgularına göre makine öğrenmesi teknikleriyle ikinci el araç alım satımında bir fiyatlandırma standardizasyonu yapabilmek mümkündür ve söz konusu makine öğrenme teknikleri doğrusal regresyon gibi klasik yöntemlere göre daha başarılı sonuçlar vermektedir.

**Anahtar Kelimeler:** Yapay Sinir Ağları, Destek Vektör Makineleri, Doğrusal Regresyon, İkinci El Araç Fiyatları

### A Comparative Study of Machine Learning and Linear Regression in Prediction of Pricing: A Case Study of Used Cars Price Prediction in Turkey

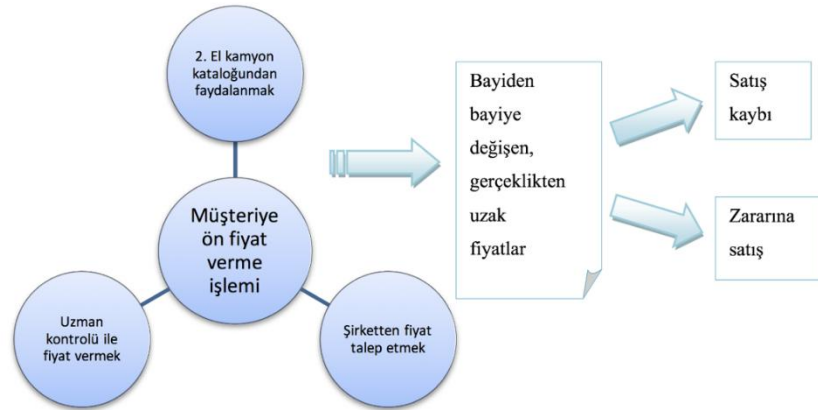
**Abstract:** Used cars market is very buoyant in the World as well as in the Turkey and has a significant size. While sellers are looking for the maximum price for their vehicles, buyers are willing to buy a good quality vehicle with a minimum price. However, during the trading process various problems might occurred and even dealer might not have standard pricing policy for a specific vehicle. In this study, it is focused on whether a pricing policy for a used car can be created or not, collected real data, and applied methods based on statistics and artificial intelligence. As a subbranch of artificial intelligence, In addition to Machine learning's success in complex problems such as natural language processing, text mining, image processing, it can also outperforms classical methods such as linear regression to solve regression problems. From this point, during this study Artificial Neural Networks, Support Vector Machines and Linear Regression methods are used and compared with the aim of standardizing used cars pricing. Based on the findings, it is possible to standardize used cars pricing with machine learning methods and chosen methods gives better results than classic linear regression method.

**Keywords:** Artificial neural networks, Support vector machines, Linear regression, Used car prices.

## GİRİŞ (INTRODUCTION)

Son yıllarda makine öğrenmesi birçok alanda çok başarılı sonuçlar vermiştir. Metin işleme (Aggarwal & Zhai, 2012; Berry & Castellanos, 2004; Tan, 1999), doğal dil işleme (Jurafsky, 2000; C. D. Manning ve diğ., 1999; C. Manning ve diğ., 2014), resim işleme (Chan ve diğ., 2015; LeCun, Bengio, & Hinton, 2015; Wan ve diğ., 2014; Wang & Yeung, 2013), zaman serileri tahminlemesi (Ahmed ve diğ., 2010; Cao & Tay, 2003), sınıflandırma (Kotsiantis ve diğ., 2007; Nasrabadi, 2007) ve gruplandırma çalışmaları (Hartigan & Wong, 1979; Jain & Dubes, 1988; Steinbach ve diğ., 2000) örnek olarak verilebilir. Klasik programlamanın aksine makine öğrenmesi verilerden öğrenme, ve öğrenilen bilginin genelleştirilebilmesine olanak sağlar. Makine öğrenmesi en bilinen şekliyle denetimli öğrenme (sınıflandırma problemleri) ve denetimsiz öğrenme (gruplandırma problemleri) olarak ikiye ayrılabilir. Denetimli öğrenmede mevcut verinin doğru çıktısı bilinirken, denetimsiz öğrenmede doğru olan çıktılar mevcut değildir. Bu problemlere ek olarak makine öğrenmesi regresyon problemlerinde de kullanılabilir. Doğrusal regresyon gibi klasik yöntemlere kıyasla makine öğrenmesi yöntemleri çok daha kompleks ve büyük verilerden anlamlı çıktılar bulabilir. Bu noktadan hareketle çalışmamızda otomotiv sektöründe hizmet veren bir şirketin, kamyon ürün grubunun ikinci el araç alımlarında yaşanan bir sorun ele alınmıştır. Bu sorun genel hatlarıyla Şekil 1’de görselleştirilmiştir. Söz konusu probleme yanıt bulabilmek amacıyla gerekli veriler yine aynı kaynaktan 2013 yılı içerisinde gerçekleşen durumlar için çekilmiş, düzenlenmiş ve Python platformunda scikit-learn kütüphanesi kullanılarak regresyon analizi yapılmış ve sonuçları yorumlanmıştır.

Kullanılmış kamyonların yenileriyle değiştirilmesi işlemi, sorunun ortaya çıktığı bayilerin ana şirketinde gerçekleştirilmektedir. Bu işlem kapsamında kamyon sahipleri, araçlarını yenileriyle değiştirmek istediklerinde, kamyonun markası fark etmeksizin bayiye gelir ve öncelikle mevcut araçlarının durumlarını anlatarak satış danışmanından bir fiyat alırlar. Eski araç için verilen fiyat, alınmak istenen yeni araç için bir indirim tutarı olarak düşünülebilir. Bu işlem için belli bir fiyatlandırma politikası tanımlanmıştır ve bu çerçevede bir katalog hazırlanmıştır. Satış danışmanları bazen bu kataloğa bakıp bir fiyat verirler bazen de kısa bir analiz işlemi gerçekleştirilerek fiyatı belirlerler. Bazen ise araç bilgileri şirkete gönderilir ve şirket araç için bir fiyat belirler. Ancak bu fiyatlandırma için belirli standardize edilmiş bir sistem yoktur, her bayi müşteriye farklı bir fiyat verebilir. Bizim için problem de bu noktada baş gösterir. Fiyatta mutabık olduğunda müşteri ile anlaşılır ve araç şirketin görevlendirdiği expert (uzman) tarafından incelemeye alınır. Bu araç için yeniden bir fiyat belirlenir ve müşteriye iletir. Bu noktada uzmanın belirlediği fiyat, en başta müşteriye verilen fiyattan çoğu zaman düşük, en iyi ihtimalle aynı seviyede çıkıyor. Beklentisinin altında bir teklif alan müşteri anlaşmaktan vazgeçiyor veya müşteriyi kaybetmemek adına araç değerinden daha yüksek bir fiyatla alınmış oluyor. Bu iki durum da firma için zarar anlamına geliyor.



Şekil 1. Ön fiyatlandırma süreci

Genel olarak ikinci el bir kamyonun fiyatlandırılmasında sorunun kaynakları üç kategoride toplanabilir 1) Müşteri kaynaklı problemler 2) Bayi kaynaklı problemler 3) Şirket kaynaklı problemler. Müşteri kaynaklı problemler genel olarak araç sahiplerinin piyasaya hakim olamamalarından kaynaklanıyor. Piyasadaki ikinci el kamyon alımlarında fiyatları bilmeyen müşteri niyetini ve beklentilerini ortaya koymaktan çekiniyor; düşük bir teklifte bulunduğu firmanın hali hazırda vermeyi planladığı daha yüksek teklifi kaybetmek istemiyor.

Bazen de tam tersi müşteri çok yüksek beklentilere sahip olduğu için verilen fiyattan hiçbir türlü memnun olmayabiliyor. Bunu sebebi de bayilerden aldıkları tekliflerin tutarsızlıkları olarak gösterilebilir. Diğer bir sorunda müşterilerin şirkete araçları hakkında yeterli ve doğru bilgiyi sağlayamaması. Bunun gerekçesi olarak aracın satılması aşamasında bayiinin satışı mümkün olduğu kadar hızlı yapabilmek adına tüm bilgileri sistemine taşınamaması ve şirketi müşterinin söylemleri kadar bilgilendirmesi gösterilebilir. Örneğin araç resmi talep eden şirkete bu resim gönderilmeyip sağlam olarak tanımlanmalı ve bu doğrultuda fiyat alındığında, daha sonra bu araç expert kontrolü için şirkete gittiğinde belirgin daha düşük niteliklere sahip olduğu tespit edildiğinde ilk teklif düşecektir. Bayi ve şirket gerçek bilgiyi elde etmeyeceği için doğru fiyat ve tatmin edici sonuçlar sağlayamazlar.

Bayi kaynaklı problemler genel olarak firma kıstaslarına yeterince itimat edilmemesinden kaynaklanıyor. Örneğin her tip kamyon ve 20 yıldan daha eski olanları takas için kabul edilemez. Ancak bayilerdeki satış danışmanları bu gibi kriterlere yeteri kadar önem vermiyor. Sadece yeni araç satmaya odaklanıyorlar. Ancak bu durumda satıcılar araç için bir fiyat belirlediğinde şirket, kriterler e uymayan bu araç için fiyat dahi vermeyebilir ve neticesinde markaya duyulan güvenin sarsılması söz konusu olur.

Bazı kısıtlamalarla birlikte, teklif geçerlilik koşulları müşterilere sunulan tüm teklifler için geçerlidir. Fakat bazen bayiler, müşterinin ön kontrol formunu kendilerinin doldurmasına izin veriyorlar ve bu formları kontrol etmeden şirkete iletiyorlar ve bu yüzden eksik veya hatalı doldurulmuş formlar süreçleri manipüle edebiliyor. Ayrıca, bayiler, satış sürecini hızlandırmak için müşteri formları düzenlenirken öznel olabiliyorlar; şirketten daha yüksek fiyat teklifi almak için şirketi ikna etmek için değerleri yukarı çekmeye çalışıyorlar. Bu sayede müşteriyi kaçırmayıp ona yeni araç satılabilmeyi umuyorlar.

Ön kontrol formundaki okunamayan veya eksik bilgi ile karşılaşıldığında sürecin baştan başlaması gerekiyor. Bu da zaman ve maliyet üzerinde olumsuz bir etki yaratıyor.

Şirketler her yıl, kullanılmış kamyonlar için ikinci el pazar fiyatlarının yer aldığı kataloglar yayınlamaktadır ancak bu fiyatlar iyi şartlara sahip kamyonlar için geçerlidir. Kullanılmış her kamyon farklı özelliklere sahip olduğu için, bu ön fiyat listesi aracın durumunun değerlendirilmesinden kullanıldığında, müşteri ana şirket tarafından yapılan uzmanlık sürecinden önce daha yüksek bir fiyat alır ve beklentisi yükselir. Expert incelemesinin ardından, genellikle teklifler düşüş yönündedir. Bayi tarafındaki bir diğer sorun bayiler arasında uyumsuzluk. Aynı ana şirkete ait olsalar bile, aynı müşteriye aynı kamyon için farklı tekliler sunulabilir. Bayiler arasındaki tutarsızlık firmaya güven kaybına neden olur. Aslında bu sorun yukarıdaki tüm sıkıntılarının doğan bir problem olarak gösterilebilir. Bayilerin ilk fiyat verme süreçlerinde belirli bir süreç tanımlanmadığı için veya ortak kullanıma uygun bir sistem oluşturulmadığı için, aynı araç için fiyat varyasyonu artıyor.

Şirket kaynaklı problemler öncelikle teknolojik uyumsuzluktan kaynaklanıyor. Şirket içinde IUCCA ve SAP olarak iki farklı veri depolama ve işleme sistemi var. Bu iki yazılım paketlerinin özelliklerinin farklı olması nedeniyle işlemler zorlaşmaktadır. İkinci sorun, tüm sürecin tamamlanmasının çok zaman almasıdır. Teklifin müşteriye yapılabilmesi için önce üst düzey yöneticiler tarafından onaylanması gerekir; fakat hiyerarşik onay zinciri çok uzundur. Son olarak, faiz ve vergi öncesi kardan zarar etme sorunu söylenebilir. Takas süreci kâr edemez, aksine ikinci el piyasada para kaybederler. Herhangi bir şirket için bu kabul edilemez.

Bütün bu problemlerin maksimum düzeyde bertaraf edilebilmesi için uzman kontrolünden önceki fiyatlandırma süreci eğer standardize edilebilirse, müşterilere çok daha kısa sürede ve herkes için aynı kriterler dikkate alınarak bir fiyat verilebilir. Aynı zamanda uzun hiyerarşik onay zincirini kırmaya

yardımcı bir adım atılmış olur. Bu amaçla bu çalışmada çeşitli makine öğrenmesi yöntemleri kullanılarak bir tahminleme modeli oluşturulmuş ve kullanılan modeller çeşitli değerlendirme ölçütleri açısından kıyaslanmıştır. Güncel literatürde Türkiye’deki ikinci el piyasasına yönelik bazı çalışmalar mevcutsa da, herhangi bir yapay zeka algoritması veya istatistiksel bir yöntem kullanılarak gerçek bir veri setiyle fiyat standardizasyonu probleminin çözülmeye çalışıldığına rastlanılmamıştır (Daştan, 2016; Ecer, 2013; Özçalıcı, 2017).

Çalışmanın geri kalan kısmı şu şekilde dizayn edilmiştir; Bölüm 2’de kullanılan yöntemlerin metodolojisi veri setinin oluşturulma süreci açıklanmıştır. Bölüm 3’te çalışmanın bulguları detaylı bir şekilde açıklanmış ve Bölüm 4’te tartışma ve sonuç başlığı altında çalışmanın sonuçları değerlendirilmiştir.

## MATERYAL VE METOT (MATERIALS AND METHODS)

Bu çalışmada makine öğrenmesi yöntemlerinden yapay sinir ağları (YSA) ve Destek Vektör Makineleri (DVM), istatistiksel yöntemlerden doğrusal regresyon yöntemleri kullanılmıştır. Her bir yöntemin matematiksel farklılıklarının probleme farklı katkılar sunacağını düşündüğümüzden benzer algoritmalar yerine bu şekilde bir algoritma seçimine gidilmiştir. Kullanılan veri makine öğrenmesi tekniklerinden optimum fayda sağlamak amacıyla 0-1 arasında normalize edilmiş ve kullanılan yöntemin değerlendirme ölçütleri bakımından güvenli sonuç verip vermediği çapraz sorgulama tekniği ile test edilmiştir. Aşağıdaki alt başlıklarda sırasıyla veri setinin oluşturulması, algoritmaların matematiksel altyapıları, değerlendirme ölçütleri ve çapraz sorgulama tekniği detaylı bir şekilde açıklanmıştır.

Çizelge 1. Ham veri seti

Table 1. Raw dataset

No	Marka	Tip	Model	Hasar Kaydi	Hasar sayısı	Özellikler	Üst yapı	Lastiklerin durumu (%)	Aksesuar	Müşteri Beklentisi	Bayi Öngörüsü	Verilen Ön Teklif	Yeni Araç Bilgisi
1	11	3240 C	2011	1 Adet Hasar Mevcuttur	1	Lastikler %70 - Retarder -Saç kasa	Saç Kasa	70	Retarder			145000	1840 LS
2	11	1840 LS	2011	Kayıt Bulunmamaktadır	0	Lastikler %70	Saç Kasa	70				140000	1840 LS
3	11	1841 LS	2011	2 Adet Hasar Mevcuttur	2	Lastikler %70	Saç Kasa	70				140000	1840 LS
5	7	CARGO 25:	2005	1 Adet Hasar Mevcuttur	1	Çift ilave açık ahşap- Lastikler %70	Açık Ahşap Kasa	75				40000	3340 K
7	16	HD75	2008	3 Adet Hasar Mevcuttur	3	Araç kargoda çalışıyor- Ön Lastikler %80, Arka Bitik, Baskı balata yeni,Kapalı saç kasa	Kapalı Saç Kasa	80		43000		25000	1518 UK
12	10	33420	2002	Kayıt Bulunmamaktadır	0	Araç çift çeker çekici olup silobas nakliyesinde çalışmıştır	Çekici			75000		30000	1840 LS
13	11	1840 LS	2007	2 Adet Hasar Mevcuttur	2	HL7						95000	1840 LS
14	11	3228 C	2005	Kayıt Bulunmamaktadır	0	Lastikler %80- Kondisyonlu-Açık ahşap, Eski Tip	Açık Ahşap Kasa	80				85000	1840 LS
15	11	2523 CD	2004	7 Adet Hasar 16300 TL	7	Saç Damper- 1 Milyonda Motor Yapılmış	Saç Damper					40000	1518 UK
16	1	PRO 1142	2008	Kayıt Bulunmamaktadır	0	Lastikler - Önler Yeni, Arka %70- Sorunsuz		70		85000		50000	1840 LS
17	11	3028 K	2006	7 Adet Hasar 13800 TL	5	Saç Damper-Lastikler %90-%60		75				86000	3340 K

## Veri setinin oluşturulması (Generation of dataset)

Bölüm 1’de bahsedilen sorunların çözüm önerileri getirilebilmesi adına yapılacak çalışmalar için 2013 yılı içerisinde aracını yenisiyle takas etmek amacıyla şirkete gelip eski aracı için uzmandan fiyat

alan tüm müşterilerin araç bilgileri ve yapılan teklifler 2 farklı veri tabanından çekilerek düzenlenmiş, veriler raporlaştırılmıştır. Elde edilen ortalama 13000 araçlık kümeden, işlemlerin analizinin yapılmasını kolaylaştırmak adına, 1500 tanesi veri seti olarak düzenlenmiştir. Çizelge 1'de ham veri seti gösterilmiştir.

Ham veri seti işleme sokulmadan önce gerekli olan ön işleme adımları uygulanmıştır. Aynı özelliklere sahip 2 farklı model araçtan birinin daha düşük fiyatlandırıldığı tespit ediliyor. Bu tip araçların yaptıkları kilometre az da olsa fazla da olsa hep daha düşük fiyata sahip oluyorlar. Bu tip araçlar için 2, diğerleri için 1 değeri atanarak 'Model Tipi' adında yeni bir sütun oluşturulmuştur. Özellik sütunundaki değerler girilirken karakteristik açıdan farklı birden fazla özellik için açıklama yapılmış, sayısal ve sözel karakterler bir arada kullanılmıştır. Bu sütundaki değerlerden aynı grup toplanabilecekler belirlenip 'PTO' ve 'Lastiklerin Durumu' olarak 2 ayrı sütun açılmıştır. Bu değerle sadece sayısal ifadeler kullanılarak girilmiştir. Böylece veriler arasında bütünlük ve anlam kolaylığı yakalanmıştır. 'Hasar Kaydı' sütunu tamamen kaldırılmış, sadece 'Hasar Sayısının tutulması yeterli görülmüştür. Aksesuar için de 'Klima' adlı bir sütun açılmıştır. İlk kümedeki özelliklerin bazıları ise problemin çözümünde etkisiz görülmüştür. Bunlar 'Müşterinin Beklentisi', 'Bayinin Öngörüsü', 'Verilen Ön teklif' ve 'Yeni Aracın Bilgisi'dir. Bu sütunlar çalışmada değerlendirmeye katılmamıştır. Çizelge 2'de veri setinin analiz edilmeye uygun son hali örnek olarak gösterilmiştir.

Çizelge 2. Veri setinin analiz edilmeye uygun son hali

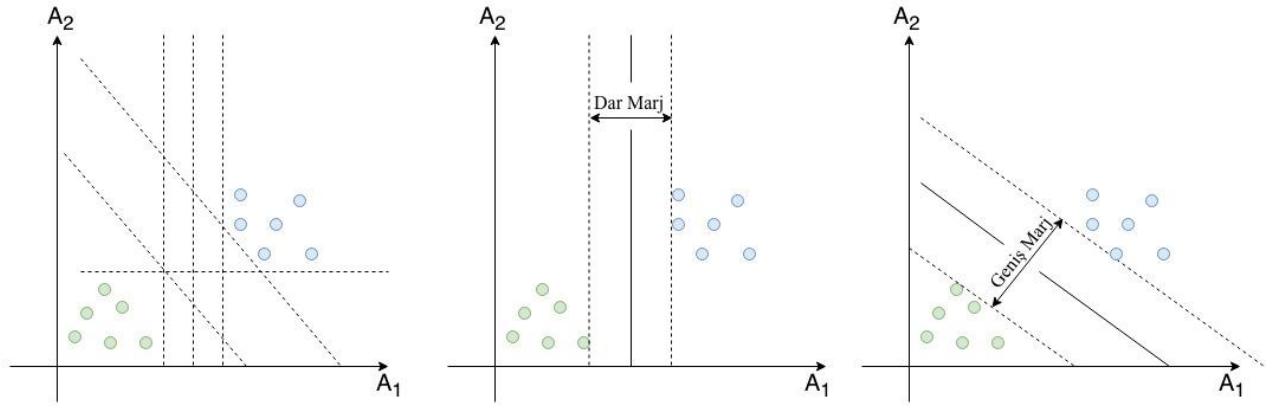
Table 2. Final version of the dataset suitable for analysis

No	Marka	Marka Grubu	Tip	Model Yılı	Km	Hasar Sayısı	Lastiklerin Durumu (%)	Klima	PTO	Teklif
1	11	2	3240 C	2011	116750	1	70	0	0	145000
2	11	2	1840 LS	2011	246630	0	70	0	0	145000
3	11	2	1840 LS	2011	107129	2	70	0	0	140000
5	7	2	CARGO 2524	2005	450000	1	75	0	0	40000
7	16	1	HD 75	2008	190000	3	80	0	0	25000
12	10	1	33420	2002	334290	0	100	0	0	30000
13	11	2	1840 LS	2007	263000	2	100	0	0	95000
14	11	2	3228 C	2005	380000	0	80	0	0	85000
15	11	2	2523 CD	2004	250000	7	100	0	0	40000
16	1	1	PRO 1142	2008	490000	0	70	0	0	50000
17	11	2	3028 K	2006	248517	5	90	0	0	86000
1308	11	1	1518 UK	2010	780000	2	100	0	0	72000
5315	11	2	1840 LS	2009	406000	2	60	0	0	130000
21	11	2	2124 C	2010	50000	0	100	0	0	95000

## Algoritmalar (Algorithms)

### Destek Vektör Makinaları (Support Vector Machines)

Destek vektör makineleri (DVM) (Cortes & Vapnik, 1995) , tüm iyi bilinen veri madenciliği algoritmalarında en sağlam ve doğru yöntemlerdendir. Vapnik tarafından 1990'larda geliştirilen DVM'ler, istatistiksel öğrenme teorisini temel alan sağlam bir teorik temele dayanır, yalnızca bir düzine eğitim örnekleri gerektirir ve genellikle boyut sayısına duyarlıdır. Son on yılda, DVM'ler hem teori hem de uygulamada hızlı bir şekilde geliştirildi. İki sınıflı doğrusal olarak ayrılabilir bir öğrenme için DVM'nin amacı, verilen örneklerin iki sınıfını, en iyi genelleme sağlayan kanıtlanmış en büyük marj (aralık) ile ayıran bir hiper düzlemi bulmaktır. Genelleme yeteneği, bir sınıflandırıcının, eğitim verisinde iyi sınıflandırma performansına sahip olmakla kalmayıp, eğitim verileri ile aynı dağılımdan gelecek verilere ilişkin yüksek tahmin doğruluğunu garanti eder. Sezgisel olarak, bir marj, bir hiper düzlem tarafından tanımlanan iki sınıf arasındaki boşluk veya ayırım miktarı olarak ifade edilebilir. Geometrik olarak kenar boşluğu, hiper düzlem üzerindeki herhangi bir noktaya en yakın veri noktaları arasındaki en kısa mesafeye karşılık gelir. Şekil 2'de DVM yönteminin çalışma prensibi gösterilmiştir.



Şekil 2. DVM yönteminin çalışma prensibi

Figure 2. Working principle of DVM method

Matematiksel olarak somutlaştırmak için elimizde  $X = \{(x_i, y_i)\}_{i=1}^l : (x_i, y_i) \in R^{n+1}$ ,  $l$  ve  $n$  sırasıyla toplam örnek sayısı ve toplam özellik (sütün sayısı),  $y_i \in \{1, -1\}$  veri setinin olduğunu varsayılmaktadır. Bu veri seti için linear bir fonksiyon  $w$ : bilinmeyen katsayıları ve  $b$ : değişken olmayan değer parametreleri kullanılarak Denklem 1’deki gibi oluşturulabilir.

$$f(x) = w^T \phi(x_i) + b \quad (1)$$

Destek vektör makineleri veri seti gurupları arasında Şekil 2’de gösterildiği gibi maksimum düzeyde bir hiper düzlem oluşturmaya çalışır. Dolayısıyla aynı mantık üzerine regresyon problemi için DVM yönteminin oluşturacağı  $f(x)$  fonksiyonun gerçek değerlerle olabildiği kadar örtüşmesi mantığı söz konusudur. Bu hedef doğrultusunda optimizasyon problemi aşağıdaki gibi düzenlenmiştir.

$$f \min \frac{1}{2} \|w\|^2 \quad (2)$$

$$\text{s.t.} \begin{cases} y_i - (w^T(x_i) - b) \geq \varepsilon & i=1,2,\dots,l \\ y_i - (w^T(x_i) - b) \leq -\varepsilon & i=1,2,\dots,l \end{cases}$$

Bu optimizasyon probleminde  $\varepsilon$  izin verilen en büyük hata miktarı olarak tanımlanır. Tahmin edilen bütün  $y = (y_i - (w^T(x_i) - b))$  değerleri ile gerçek  $y$  değerleri arasındaki fark  $\varepsilon$  değerini aşmadığı sürece bu problem konveks bir optimizasyon problemidir. Ancak her problem için bu şartları sağlamak çok kolay olamayacağından optimizasyon problemine yapay değişkenler eklenerek hiper düzlem sınırları daha esnek hale getirilebilir. Bu durumda problemin formülasyonu aşağıdaki gibi olmaktadır.

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (3)$$

$$\text{s.t.} \begin{cases} y_i - (w^T(x_i) - b) \geq \varepsilon + \xi_i & i=1,2,\dots,l \\ (w^T(x_i) + b) + y_i \geq \varepsilon + \xi_i^* & i=1,2,\dots,l \\ \xi_i^*, \xi_i \geq 0 \end{cases}$$

#### Doğrusal Regresyon (Linear Regression):

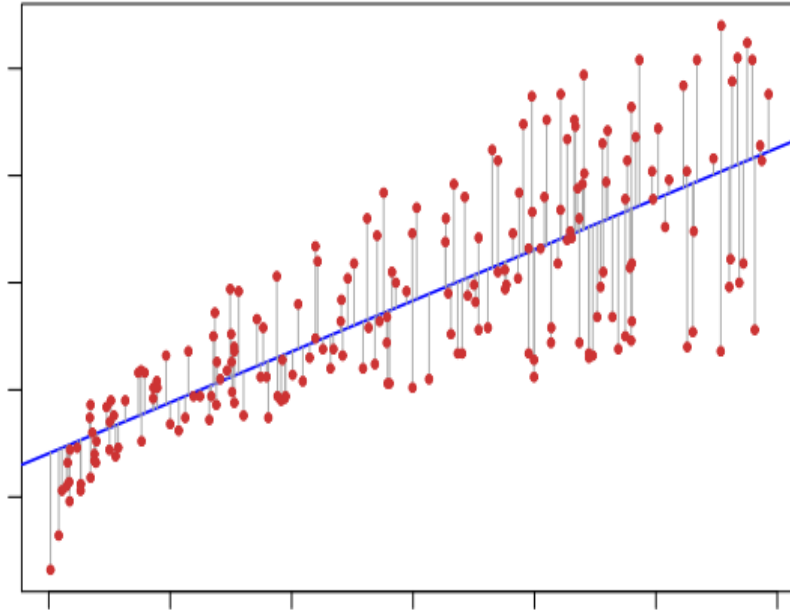
Basit doğrusal regresyon analizi (Neter, Wasserman, & Kutner, 1989), iki sürekli değişken arasındaki ilişkiyi nicelendiren istatistiksel bir tekniktir. Bu değişkenler, bağımlı değişken veya tahmin etmeye çalıştığınız değişken ve bağımsız veya tahmin edici değişken olarak ifade edilebilir. Her noktadan kareli

hatayı asgariye indiren verilerle bir doğru bularak çalışır. Şekil 3'te bahsedilen durum ifade edilmiştir. Matematiksel olarak ifade edebilmek için elimizde  $X = \{(x_i, y_i)\}_{i=1}^l : (x_i, y_i) \in R^{n+1}$  veri setinin olduğunu varsayalım. Bağımsız değişkenler  $X = \{(x_i)\}_{i=1}^l$  ile bağımsız değişkenler  $y_i$  arasındaki ilişki doğrusal ilişki aşağıdaki gibi tanımlanabilir

$$y_i = x_i^T w + \varepsilon_i \quad (4)$$

Doğrusal regresyon tahmin edilen değerlerle gerçek değerler arasındaki hatayı minimum seviyede tutmak üzere çalışır. Bu bakımdan optimizasyon problemi aşağıdaki şekilde formülize edilebilir.

$$\min_n \frac{\sum_{i=1}^l (y_i - \bar{y}_i)^2}{l} \quad (5)$$



Şekil 3. Doğrusal regresyon yönteminin çalışma prensibi

Figure 3. Working principle of linear regression method

#### Yapay Sinir Ağları (Artificial Neural Networks):

Sinir ağları, karmaşık ve çok boyutlu verileri basitleştirir, değerler arasında kolay işlem yapılmasını sağlar. Doğrusal modellere uymayan herhangi bir regresyon problemi, bu yöntem kullanılarak modellenebilir (Rokach & Maimon, 2008). YSA mimarisi Şekil 4'te gösterildiği gibi birbirleriyle bağlantılı bir dize nöronlardan oluşmaktadır. Bu bağlantılardaki amaç girdilerle çıktılar arasında doğrusal olmayan ilişkileri açığa çıkarmaktır. Her bir katmanın girdileri bir sonraki katmanın ağırlıklı toplamları olacak şekilde girdilerini oluşturur ve her bir girdi nöronlarda bazı aktivasyon fonksiyonları ile belirli aralıkta, örneğin 0-1, bir değer dönüştürülür. Çıktı katmanında ki hesaplamalardan dolayı sistem geriye dönük olarak maliyet fonksiyonunu -ki bu fonksiyon gerçek değerle tahmin edilen değer arasındaki farkı minimize etmeye çalışır. Matematiksel olarak ifade etmek Şekil 4'teki YSA modelini ele almak gerekirse: l katmanındaki n nöronunun girdisi Denklem 6 gösterildiği gibi hesaplanır.

$$y_i^n = \varphi \left[ \sum_{j=1}^p w_{ij}^n(t) y_j^{n-1}(t) + \gamma_i^n \right] \quad (6)$$



Bu denklemde  $\varphi(\cdot)$  aktivasyonu fonksiyonunu,  $w_{ij}^n$  bağlantı katsayısını,  $t$  iterasyonu,  $\gamma_i^n$  ise sabit değişkeni ifade eder.  $n$  katmanlı bir yapıda ağırlıklar  $0 \leq \epsilon \leq 1$  olmak koşulu ile aşağıdaki gibi revize edilir.

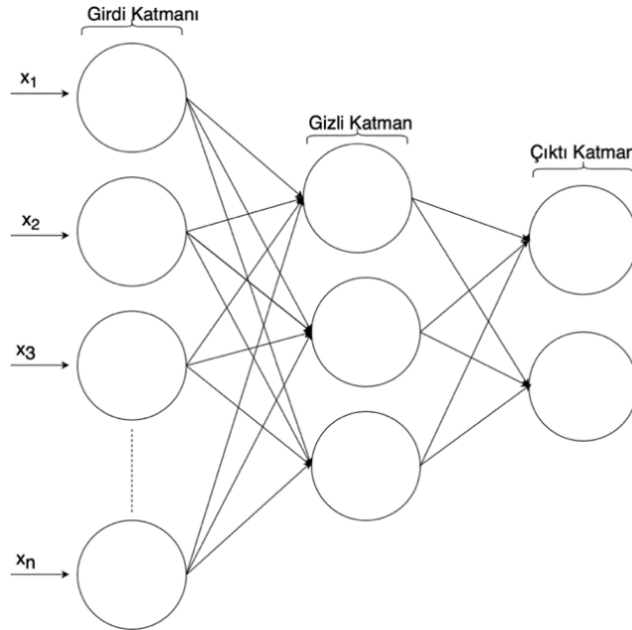
$$\Delta w_{ji}^n(t) = \epsilon \lambda_j^n(t) y_i^{n-1}(t) \quad (7)$$

Bu problemde  $\epsilon$  öğrenme oranını ve  $\lambda_j^n(t) \equiv \partial E_t / \partial u_j^n$  lokal hatayı gösterir. Algoritmanın hata oranını minimize etmek için ağırlıklar  $0 \leq \alpha \leq 1$  olmak koşulu ile aşağıdaki gibi optimize edilir. Son olarak çıktı katmanının hata oranı ise Denklem 8’de gösterildiği gibi optimize edilir.

$$\lambda_j^N(t) = [d_j(t) - y_j^N(t)] \varphi[u_j^N(t)] \equiv e_j(t) \varphi[u_j^N(t)] \quad (8)$$

Bu problemde  $d_j(t)$  hedef çıktısının değerini gösterir.

Analiz ettiğimiz problemde dikkat edilmesi gereken nokta sınıflandırma problemlerinin aksine son katmanda sadece 1 adet nöron bulunmaktadır. Ayrıca çalışmada oluşturulan yapay sinir ağı modelinde 1 gizli katman ve gizli katmanda 11 adet nöron (veri özellik sayısı + çıktı katmanı nöron sayısı) kullanılmış ve öğrenme optimizasyonu geri yayılım (backpropagation) algoritması ile gerçekleştirilmiştir.



Şekil 4. Yapay sinir ağı modeli  
Figure 4. Artificial neural network model

#### Değerlendirme ölçütleri (Evaluation metrics)

Çalışma boyunca algoritma performansları çeşitli değerlendirme ölçütlerine göre kıyaslanmıştır. Kullanılan değerlendirme ölçütleri Çizelge 3’te gösterilmiştir.



### Çapraz Doğrulama ve Veri Bölme Yöntemi (Cross Validation and Data Division Method)

Çapraz doğrulama yöntemi algoritmanın çalışma zamanının ve sonuçlarının optimize edilmesine katkı sağlayan bir yöntemdir. Veri rastgele  $k$  farklı parçaya bölünür. Her bir grup test verisi olarak kullanılmak üzere ayrılır ve geriye kalan  $k-1$  grubun tamamı eğitim seti olarak kullanılır. Başka bir deyişle algoritma  $k$  kez veriyi eğitir ve sonuçlarını test eder. Uygulanan metodun doğruluğu ise her bir test verisinin sonuçlarının ortalamasıdır. Şekil 5'te çapraz doğrulama yapısı gösterilmiştir.

Matematiksel olarak ifade etmek için, elimizde  $D$  veri setinin olduğunu varsayalım.  $D$  seti denklem 15'teki eşitliği sağlayacak şekilde rastgele  $k$  parçaya ayrılır

$$\bigcup_{i=1}^k \{D_1, D_2, \dots, D_k\} = D \quad (9)$$

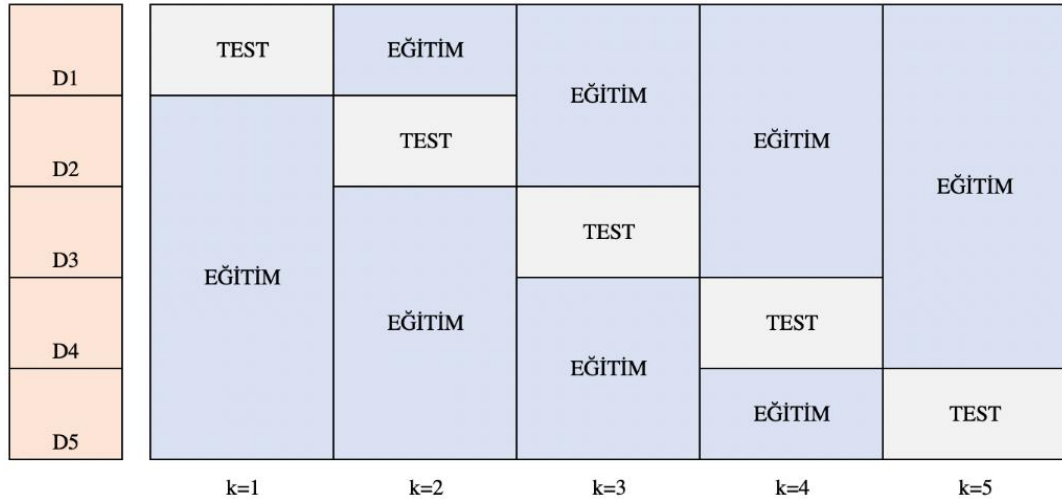
Çizelge 3. Değerlendirme ölçütleri

Table 3. Evaluation metrics

Korelasyon Katsayısı ( $R^2$ )	$\left( \frac{n \sum y \cdot y' - (\sum y) (\sum y')}{\sqrt{n(\sum y^2) - (\sum y)^2} \sqrt{n(\sum y'^2) - (\sum y')^2}} \right)^2 \quad (10)$
Ortalama Mutlak Hata (MAE)	$\frac{1}{n} \sum_{i=1}^n  y - y'  \quad (11)$
Kök Mutlak Hata (RAE)	$\frac{\sum_{i=1}^n  y' - y }{\sum_{i=1}^n  \bar{y} - y } \quad (12)$
Kök Rölatif Karesel Hata (RRSE)	$\sqrt{\frac{\sum_{i=1}^n (y' - y)^2}{\sum_{i=1}^n (\bar{y} - y)^2}} \quad (13)$
Kök Ortalama Kare Hata (RMSE)	$\sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (14)$
Ortalama Mutlak Yüzdesele Hata (MAPE)	$\frac{100\%}{n} \sum_{i=1}^n \left  \frac{e_i}{y_t} \right  \quad (15)$

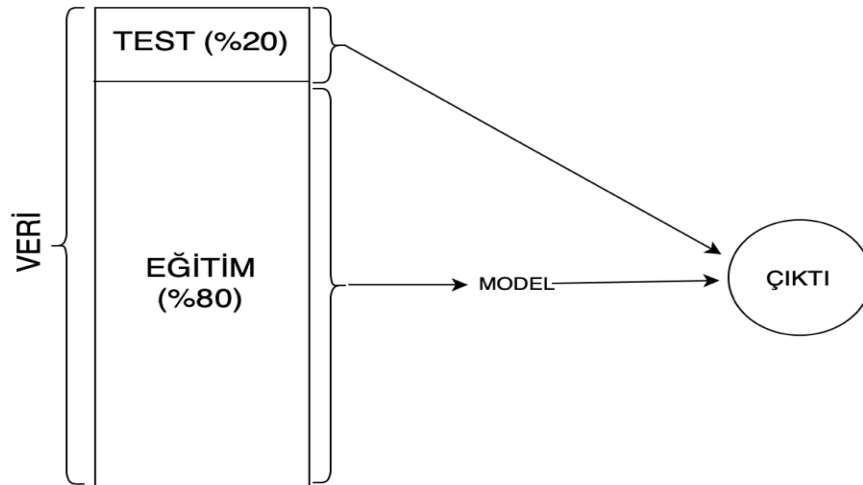
Yukarıda belirttiğimiz gibi eğitim süreci boyunca her bir  $D_i$  test veri olarak ayrılır ve eğitim her bir  $D_i$  test seti olarak kullanılıncaya kadar devam eder. Son olarak algoritmanın genel performansı  $p^*$  Denklem 16'da gösterildiği gibi her bir eğitimin performanslarının ( $p_i$ ) ortalaması olarak hesaplanır.

$$p^* = \frac{\sum_{i=1}^k p_i}{k} \quad (16)$$



Şekil 5. 5-fold Çapraz doğrulama yapısı  
Figure 5. 5-fold cross validation structure

Öte yandan çapraz doğrulama yönteminden farklı olarak veri bölme yöntemi veriyi belli yüzdelerle eğitim ve test seti olacak şekilde ikiye böler. Örneğin Şekil 6’da gösterildiği gibi veri %80 ve %20 olmak üzere ikiye bölünür ve %80’lik kısmı eğitim %20’lik kısmı test verisi olarak kullanılır. Algoritmanın performansı test verisinde gösterdiği performans ölçütü olarak belirlenir.



Şekil 6. Veri bölme yöntemi yapısı  
Figure 6. Data division method structure

#### ARAŞTIRMA SONUÇLARI (RESEARCH RESULTS)

Bu bölümde algoritmaların detaylı sonuçları ve yukarıda belirtilen değerlendirme ölçütleri kapsamında kıyaslama sonuçları verilmiştir. Algoritma sonuçlarının güvenilir olması ve genelleştirilebileceğini ispatlamak açısından belirtilen üç algoritma da 10-folds, 8-folds, 5-folds çapraz doğrulama ve veri bölme yöntemi ile çalıştırılmış ve sonuçları kıyaslanmıştır. Her bir yöntemin minimum hata vermesi için ızgara sistemi yöntemi ile parametre optimizasyonu yapılmıştır. Öte yandan bütün metodlar için girdi değerleri z değeri ile standardize edilmiştir.

YSA yöntemi için kullanılan parametreler ve elde edilen optimum değerler sırasıyla Çizelge 4 ve Çizelge 5’te gösterilmiştir.

**Çizelge 4.** Yapay sinir ağı yönteminde kullanılan parametreler**Table 4.** Parameters used in artificial neural networks method

Parametre Sembolleri	Tanım
L	Öğrenme Hızı
M	Momentum
N	Eğitim Tur Sayısı
V	Test verisinin yüzdelik değeri
S	Tekrar edilebilirlik için üretilmiş rastgele sayı
E	Hata eşik değeri.

**Çizelge 5.** Yapay sinir ağı yönteminde kullanılan parametrelerin optimum değerleri**Table 5.** Optimum values of parametrs used in artificial neural network method

Parametreler	L	M	N	V	S	E
Değerler	0.3	0.2	500	0	0	20

Yapay sinir ağları yönteminin çeşitli metotlarla (örneğin 10-folds) çalıştırılarak elde edilen sonuçlar Çizelge 6' da gösterilmiştir. Söz konusu yöntem en iyi performansı veri setinin %80 veri bölme durumu uygulandığında göstermiştir. Değerlendirme ölçütlerinde en iyi performans değerleri koyu fontla gösterilmiştir. Çizelge 6' da görüldüğü gibi YSA tahmini değerleri ile gerçek değerler arasında yüksek bir korelasyon vardır ( $R^2=0.92$ ). Aynı zamanda YSA modelinin herhangi bir araç modeli için vereceği değer ile aracın gerçek değeri arasında MAE kıstası göz önünde bulundurulduğunda +/- 10658.0037 liralık bir fark oluşmaktadır ki otonom bir tahminleme değeri için kabul edilebilir bir değer olarak nitelendirilebilir

**Çizelge 6.** Yapay sinir ağları sonuçları**Table 6.** Artificial neural networks results

Değerlendirme ölçütleri	$R^2$	MAE	RMSE	RAE	RRSE	MAPE
Çapraz doğrulama folds= 5	0.7157	19378,0878	28451,1314	64,1250%	74,4125%	23,99%
Çapraz doğrulama folds= 8	0.7047	17696,2071	28372,0690	58,5586%	74,1946%	23,05%
Çapraz doğrulama folds= 10	0.7961	14915,7703	23816,8406	49,3483%	62,2880%	19,69%
Veri bölme = 90%	0.7521	2188,90308	2638,86343	73,7364%	73,3979%	30,88%
Veri bölme = 80%	<b>0.9202</b>	<b>10658,0037</b>	<b>15413,4091</b>	<b>33,8439%</b>	<b>40,8894%</b>	<b>15,30%</b>

Doğrusal regresyon yönteminde minimum hatayı elde etmek için parametre optimizasyonu yapılmıştır. Kullanılan parametreler ve elde edilen değerler Çizelge 7 ve Çizelge 8' de gösterilmiştir.

**Çizelge 7.** Doğrusal Regresyon yönteminde kullanılan parametreler

**Table 7.** Parameters used in linear regression method

Parametre Sembolleri	Tanım
S	Veri özelliklerinin seçimi için gerekli olan metod. Örnek: S=0 M5 metodunu ifade eder.
R	Ridge optimizasyon parametresi
Num Decimal Places	Çıktılardaki maksimum ondalık rakam adeti.

**Çizelge 8.** Doğrusal Regresyon yönteminde kullanılan parametrelerin optimum değerleri

**Table 8.** Optimum values of parametrs used in linear regression method

Parametreler	S	R	Num Decimal Places
Değerler	0	1.00E-08	500

YSA yönteminden farklı olarak doğrusal regresyon yöntemi en iyi sonuçları 90% veri bölme yöntemi uygulandığında vermektir. Yapay sinir ağları ve doğrusal regresyon yönteminden Çizelge 9’da gösterildiği gibi kıyaslanan bütün değerlendirme ölçüleri bakımından üstünlük sağlamaktadır.

**Çizelge 9.** Doğrusal regresyon sonuçları

**Table 9.** Linear regression results

Method	R <sup>2</sup>	MAE	RMSE	RAE	RRSE	MAPE
Çapraz doğrulama folds= 5	0.8867	7840,4264	17696,4911	25,9451%	46,2843%	9,75%
Çapraz doğrulama folds= 8	0.8899	7676,2335	17442,2522	25,4015%	45,6125%	9,56%
Çapraz doğrulama folds= 10	0.8899	7669,2722	17450,318	25,3735%	45,6377%	9,53%
<b>Veri bölme = 90%</b>	<b>0.9614</b>	<b>6861,8214</b>	<b>9922,5663</b>	<b>23,1151%</b>	<b>27,5989%</b>	<b>9,91%</b>
Veri bölme = 80%	0.9393	7848,424	12968,78	24,9222%	34,4042%	9,78%

YSA yöntemiyle kıyasladığımızda Doğrusal Regresyon yöntemi tahminleri ile gerçek değerler arasında R<sup>2</sup>=0.9614 ile çok yüksek bir korelasyon yakalamış ve YSA’dan daha iyi bir performans yakalamıştır. Bu çıkarımı -yine örnek olarak MAE kıstasını göz önünde bulundurursak- tahmin edilen fiyatla gerçek fiyat arasındaki farka bakarak da görebiliriz. Doğrusal Regresyon modelinin herhangi bir araç modeli için vereceği değer ile aracın gerçek değeri +/- 6861.8214 liralık bir fark oluşmaktadır ki bu değer YSA modelinin verdiği değerinin çok altındadır.

Doğrusal regresyon yöntemi için ayrıca istatistiksel z-test uygulanmıştır. Eğitim modeli için kullanılan 1500 örnek normal olarak dağılmaktadır ve z-test için kurulan hipotez gerçek test setinin ortalaması, tahmin edilen değerlerin ortalamasından istatistiksel olarak farklı değildir. Bu bağlamda  $\mu_0$  = gerçek test değerlerinin ortalaması ve  $\mu_1$  = tahmin edilen değerlerin ortalaması olarak düşünüldüğünde hipotezimizi aşağıda belirtildiği gibi kurabiliriz.

$$H_0: \mu_0 = \mu_1$$
$$H_1: \mu_0 \neq \mu_1$$

p değeri 0.05 alınarak uygulanan z-testi sonuçları Çizelge 10’da verilmiştir.

**Çizelge 10.** Doğrusal regresyon için istatistiksel z-test sonuçları**Table 10.** Statistical z-test results for linear regression

Parametreler	Actual	5-folds	8-folds	10-folds	90%	80%
Ortalama	80575	77065.5	76900	76900	79947.5	78973.75
Bilinen Varyans	1543071154	1382885800	1375694872	1375363077	1383963071	1491381152
Gözlem Sayısı	1500	1500	1500	1500	150	300
Örnek Kümeler Ortalama Farkı	0	0	0	0	0	0
Alfa	0.05	0.05	0.05	0.05	0.05	0.05
z değeri	0.41	0.41	0.43	0.43	0.07	0.18
P(Z<=z) tek yönlü	0.34	0.34	0.33	0.33	0.47	0.43
Kritik z değeri (tek yönlü)	1.64	1.64	1.64	1.64	1.64	1.64
P(Z<=z) (iki yönlü)	0.68	0.68	0.67	0.67	0.94	0.85
Kritik z değeri (çift yönlü)	1.96	1.96	1.96	1.96	1.96	1.96

Çizelge 10'da verilen sonuçlara göre uygulanan her stratejide (örnek: 10-folds,%90 veri bölme) doğrusal regresyon çıktıları gerçek değerlerle aynı istatistiksel ortalamaya sahip. Dolayısıyla hiçbir stratejide  $H_0: \mu_0 = \mu_1$

hipotezi reddedilemez. Bir başka deyişle doğrusal yönteminin sonuçları şans eseri üretilmemiş, bir öğrenme mekanizmasıyla eğitim setinden elde edilen model üzerinden tahminler yapılmıştır.

Uygulanan son yöntem Destek vektör makinalarıdır. DVM metodu için kullanılan parametreler ve parametre optimizasyon sonuçları Çizelge 11 ve Çizelge 12'de gösterilmiştir.

**Çizelge 11.** Destek vektör makinaları yönteminde kullanılan parametreler**Table 11.** Parameters used in support vector machines

Parametre Sembolleri	Tanım
C	Hata beklenti oranı
N	Veri normalizasyon çeşidi. Örnek: N=0 normalizasyon, N=1 standardize
T	Eğitim Tur Sayısı
W	Tekrar edilebilirlik için üretilmiş rastgele sayı
Kernel	Kernel fonksiyonu
E	Kernel fonksiyonunun üstsel değeri

**Çizelge 12.** Destek vektör makinaları yönteminde kullanılan parametrelerin optimum değerleri**Table 12.** Optimum values of paramtrs used in support vector machines method

Parametreler	C	N	T	W	Kernel	E
Değerler	1.1	0	0.001	1.0	Poly	1.0

Çizelge 13'de gösterildiği gibi DVM metodu doğrusal regresyon yöntemiyle benzer olarak %90 veri bölme yöntemiyle en başarılı sonucu vermiştir. Ek olarak iki yöntem arasında benzer metot (veri bölme 90%) kullanıldığında DVM nispeten daha başarılı sonuç vermiştir. Hem gerçek değerlerle tahmin edilen değerler arasında daha yüksek bir korelasyon sağlanmış ve hem de bunun doğal bir sonucu olarak hata miktarı da azalmıştır. DVM metodunun herhangi bir araç modeli için vereceği değer ile aracın gerçek değeri +/- 6227.9912 liralık bir fark oluşmaktadır ki bu değer hem YSA hem de Doğrusal Regresyon yöntemlerinin tahmin ettiği değerlerin altındadır.

**Çizelge 13.** Destek vektör makineleri sonuçları

**Table 13.** Support vector machines results

Method	R <sup>2</sup>	MAE	RMSE	RAE	RRSE	MAPE
Çapraz doğrulama folds=5	0.8942	6983,154	17113,7095	23,1083%	44,7600%	14,53%
Çapraz doğrulama folds=8	0.8921	7036,84	17274,70	23,2856%	45,1744%	14,60%
Çapraz doğrulama folds=10	0.8911	7105,9141	17342,2816	23,5097%	45,3552%	15,04%
<b>Veri bölme = 90%</b>	<b>0.9638</b>	<b>6227,9912</b>	<b>9590,313</b>	<b>20,9799%</b>	<b>26,6747%</b>	<b>10,86%</b>
Veri bölme = 80%	0.9434	7162,2449	12501,5883	22,7433%	33,1648%	12,98%

Kullanılan yöntemlerin Çizelge 3’de verilen değerlendirme ölçüleri için gösterdikleri performans değerleri bütünsel bir grafik elde etmek için 0-1 arasında normalize edilerek Şekil 7’de görselleştirilmiştir.

#### TARTIŞMA (DISCUSSION)

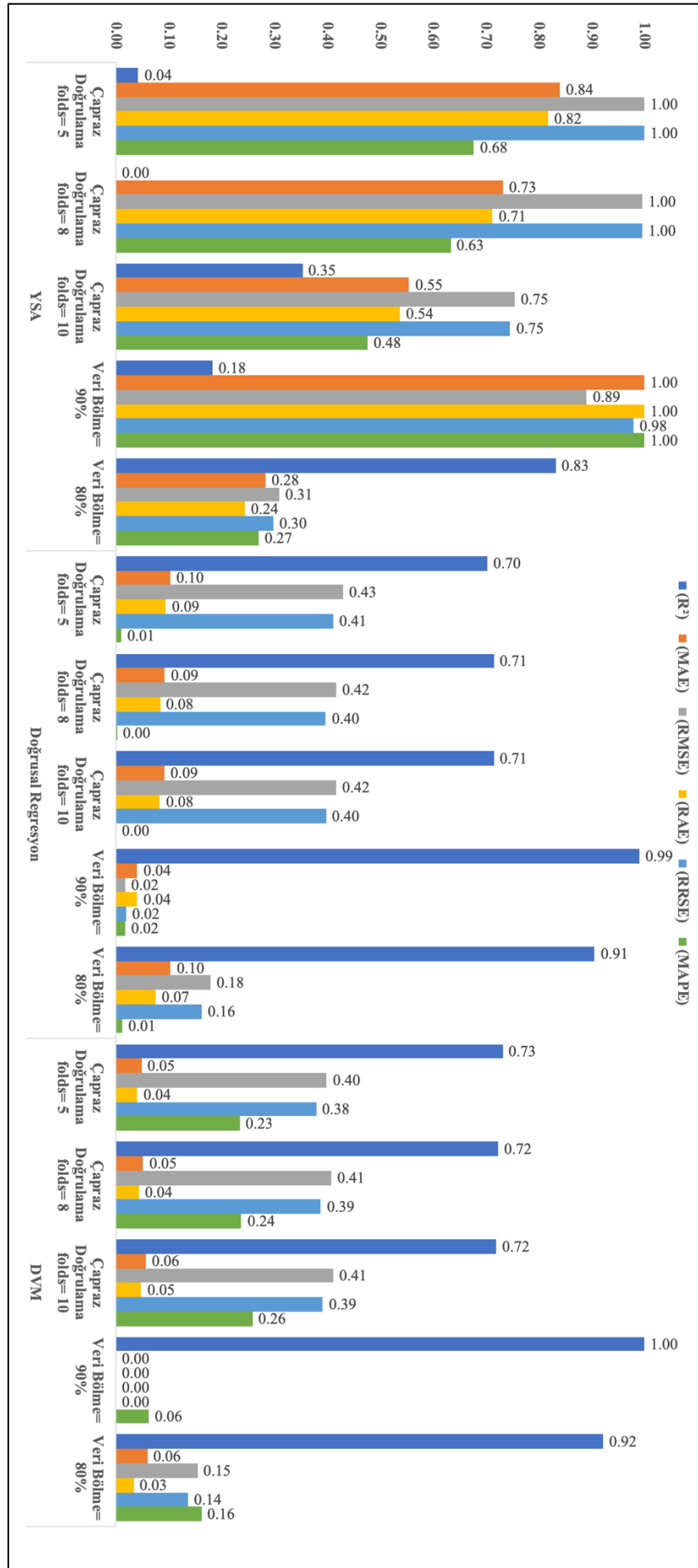
Bu çalışmada irdelenen sorunun kaynağı olan otomotiv firmasında, standart olmayan ve neticesinde şirkete maddi zarar ile zaman kaybı yaşatan ve müşteri-şirket güven bağının zayıflamasına sebep olan ikinci el kamyonların takasında ön fiyatlama sürecinin incelenmesine karar verilmiştir. Öncelikle söz konusu süreçte teklif veriler araçlarda hangi kriterlerin fiyata etki ettiği belirlenmiş; bu kriter değerleri ve teklifler kullanılarak veri seti oluşturulmuş; veriler Python scikit-learn ve Weka makine öğrenmesi platformları kullanılarak analiz edilmiştir. Ancak güncel literatürde benzer bir çalışma olmadığından ve kullanılan veri setinin özgün olmasından kaynaklı çalışma sonuçlarını diğer çalışmalarla kıyaslamak mümkün olmamaktadır.

Amaç, fiyatlar belirlenirken kullanılan özellik değerleri arasındaki ilişkinin en iyi seviyede tahmin edilmesidir. Bunun için Python scikit-learn ve Weka kütüphanesinden üç farklı algoritma -Yapay Sinir Ağları, Doğrusal Regresyon ve Destek Vektör Makineleri- kullanılmıştır ve her biri için beşer test yöntemi uygulanmıştır. Bu sonuçların tamamı Çizelge 6, Çizelge 9 ve Çizelge 13’de gösterilmiş ayrıca Şekil 7’de görselleştirilmiştir.

Bu çizelgelerde yorumlama için önemli kriterler Correlation Coefficient (Korelasyon katsayısı) değerinin yüksek ve hata oranlarının düşük olmasıdır. Korelasyon katsayısı -1 ile 1 arasındaki herhangi bir değeri alabilir. Bu katsayısı negatif olması ters yönlü bir ilişkiyi (değişken değerleri artınca sonuç azalır veya değişken değerleri azalınca sonuç artar), pozitif olması da aynı yönlü ilişkiyi (değişken değerleri artınca sonuç da artar veya değişken değerleri azalınca sonuç da azalır) ifade eder. Katsayı -1 veya +1’e ne kadar yakınsa ilişki o kadar kuvvetlidir diyebiliriz.

Bölüm 3’de açıklandığı üzere en iyi sonuçlar DVM algoritmasıyla veri bölme yöntemi %90 uygulanarak elde edilmiştir. Bu değer 0,9638 ile 1’e oldukça yakındır. Bu sonuç bize, SMO algoritmasının araçların özellikleri ve fiyatlarıyla ilişkilerini en iyi şekilde öğrenebildiğini gösterir. Aynı şekilde DVM seçilen bütün kıstaslara göre yine en az hata veren algoritmadır.

Sonuç olarak bu algoritmadan destek alınarak oluşturulan yeni bir ön fiyatlama sisteminde, müşteriye çok daha kısa sürede tutarlı ve her bayide aynı teklifler verilebilecektir. Ayrıca, doğru ve hızlı fiyatlandırma nedeniyle müşteri güven ve memnuniyetinde de artış olacaktır. Aynı zamandan tüm fiyatlandırma verileri otomatik olarak alınabilecektir.



Şekil 7: YSA, DVM ve Doğrusal regresyon yöntemlerinin performanslarının karşılaştırılması

Figure 7. Performance comparison of ANN, SVM and linear regression methods



**KAYNAKLAR (REFERENCES)**

- Aggarwal, C. C., & Zhai, C. (2012). *Mining text data*. Springer Science & Business Media.
- Ahmed, N. K., Atiya, A. F., Gayar, N. El, & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5–6), 594–621.
- Berry, M. W., & Castellanos, M. (2004). Survey of text mining. *Computing Reviews*, 45(9), 548.
- Cao, L.-J., & Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks*, 14(6), 1506–1518.
- Chan, T.-H., Jia, K., Gao, S., Lu, J., Zeng, Z., & Ma, Y. (2015). PCANet: A simple deep learning baseline for image classification? *IEEE Transactions on Image Processing*, 24(12), 5017–5032.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Daştan, H. (2016). TÜRKİYE’DE İKİNCİ EL OTOMOBİL FİYATLARINI ETKİLEYEN FAKTÖRLERİN HEDONİK FİYAT MODELİ İLE BELİRLENMESİ. *İktisadi ve İdari Bilimler Fakültesi Dergisi*, 18(1), 303–327.
- Ecer, F. (2013). Türkiye’de 2. El otomobil fiyatlarının tahmini ve fiyat belirleyicilerinin tespiti.
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108.
- Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*.
- Jurafsky, D. (2000). *Speech and language processing: An introduction to natural language processing*. Computational Linguistics, and Speech Recognition.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 160, 3–24.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436.
- Manning, C. D., Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT press.
- Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55–60).
- Nasrabadi, N. M. (2007). Pattern recognition and machine learning. *Journal of Electronic Imaging*, 16(4), 49901.
- Neter, J., Wasserman, W., & Kutner, M. H. (1989). *Applied linear regression models*.
- Özçalıcı, M. (2017). Veri Madenciliğinde Birliktelik Kuralları ve İkinci El Otomobil Piyasası Üzerine Bir Uygulama. *ODÜ Sosyal Bilimler Araştırmaları Dergisi (ODÜSOBİAD)*, 7(1), 45–58.
- Rokach, L., & Maimon, O. Z. (2008). *Data mining with decision trees: theory and applications* (Vol. 69). World scientific.
- Steinbach, M., Karypis, G., & Kumar, V. (2000). A comparison of document clustering techniques. In *KDD workshop on text mining* (Vol. 400, pp. 525–526). Boston.
- Tan, A.-H. (1999). Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases* (Vol. 8, pp. 65–70). sn.
- Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., & Li, J. (2014). Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 157–166). ACM.
- Wang, N., & Yeung, D.-Y. (2013). Learning a deep compact image representation for visual tracking. In *Advances in neural information processing systems* (pp. 809–817).