

Ev Fiyat Tahmini

Hilal Avşar*

1 Giriş

Ev fiyatları tahmini, gayrimenkul piyasasında önemli bir rol oynamaktadır. Emlak sektöründe, ev fiyatları tahmini, birçok farklı faktörü dikkate alarak bir evin değerini belirlemek için kullanılmaktadır. Bu faktörler arasında evin konumu, büyüklüğü, yaşına ve özelliklerine, konut piyasasının durumuna, faiz oranlarına, vergi politikalarına ve benzeri faktörlere bağlı olarak değişiklik gösterebilir. Genelde emlak sektöründe çalışan ve tecrübe sahibi kişilerin sezgisel olarak yaptıkları tahminler yanında günümüzde ev özellikleri ve fiyatları hakkında bilginin özellikle internette olması nedeniyle artık daha bilimsel yöntemlerle ev fiyatı tahmini yapılabilmektedir.

Ev fiyatları tahmini için birçok farklı model kullanılabilir. Bu modeller arasında en yaygın olanı regresyon analizidir. Regresyon analizi, ev fiyatlarındaki değişkenliği, büyüklük, bulunduğu kat, ısıtma şekli, toplu ulaşımaya yakınlık gibi ev özellikleri ile tahmin etmeye çalışır. Bu modelleme yöntemi, ev özellikleri, coğrafi faktörler ve ekonomik göstergeler gibi birçok değişkeni de kullanarak ev fiyatlarını tahmin etmek için kullanılabilir. Regresyon tahminleri en temel regresyon yöntemi olan linear regresyondan daha gelişmiş makine öğrenmesi modellerine kadar birçok farklı metodla da yapılabilir.

Ev fiyatları tahmini için kullanılan diğer bir modelleme yöntemi, yapay sinir ağıdır. Yapay sinir ağı, birçok farklı değişkeni kullanarak karmaşık bir model oluşturur ve bu modeli kullanarak ev fiyatlarını tahmin eder. Bu yöntem, regresyon analizine göre daha karmaşık olmasına rağmen, daha doğru sonuçlar verebilir çünkü regresyon analizlerinde genelde değişkenlerin fazlalılığı nedeniyle göz ardı edilen etkileşimler yapay sinir ağları ile daha rahat çözümlenebilir.

Ev fiyatları tahmini modelleme, gayrimenkul sektöründe büyük öneme sahip bir konudur. Regresyon analizi ve yapay sinir ağı gibi farklı modelleme yöntemleri kullanılarak ev fiyatları tahmini yapılabilir. Ancak, modelin doğruluğu için kullanılan verilerin kalitesi çok önemlidir.

1.1 Çalışmanın Amacı

Çalışmanın amacı, ev özellikleri verilen bir versiyetinde, bu özellikleri kullanarak ev fiyatını tahmin etmektir. İstatistiksel yöntem olarak multipl lineer regresyon kullanılacaktır. Veriseti

*21080201, [Github Repo](#)

Carnegie Mellon Üniversitesi'nden alınmıştır (boston). Veriseti biri bağımlı, 13'ü bağımsız olmak üzere toplamda 14 değişken içermektedir. Gözlem sayısı fazla olması nedeniyle 220 gözleme indirilmiştir.

1.2 Literatür

Literatürde ev fiyatı tahmini üzerine kullanılan tekniklerin çeşitliliği nedeniyle birçok çalışma vardır. Verisetinin çalışıldığı makalede havakirliliğinin ev fiyatlarını etkileyen önemli bir faktör olduğu bulunmuştur (harrison1978hedonic). Bir başka çalışmada ise ev fiyat tahmininde en doğru sonucu veren metod araştırılmış ve gradient boosting metodunun lineer regresyon, lasso regresyon, ridge regresyon gibi metodlara göre daha doğru tahminde bulunduğu görülmüştür (madhuri2019house). Ülkemizden yayınlanan iki çalışmada ise yapay sinir ağlarının ev fiyat tahmininde en doğru sonuçları verdikleri görülmüştür (Ozdemir2022-vc, yilmazel2018konut)

2 Veriseti

Çalışmada kullanılan veriseti incelendiğinde 200 gözlem ve 14 değişkene sahip olduğu görülmektedir.

Verisetindeki değişkenler incelendiğinde bağımlı değişkenin (medv) ev fiyatlarının median değeri (1000 dolar katı olarak) olduğu görülmüştür. 13 bağımsız değişkenden sayısal kodlu olmayanlar sayısala çevrildiğinde eksik gözlem sayısına bakıldı. 26 eksik veri olduğu görüldüğünden bu eksik verilere sahip gözlemler çalışma dışında bırakıldı ve geriye 188 gözlem kaldığı görüldü. 1 bağımsız değişken (chas) kategorik kodlandığı görüldü.

Bağımsız sayısal değişkenler şunlardır:

crim: kişi başına düşen suç oranı

zn: 25.000 fit karenin üzerindeki parseller için imarlı meskun arazi oranı.

indus: kasaba başına perakende olmayan iş alanlarının oranı

nox: nitrik oksit konsantrasyonu (10 milyonda parça)

rm: konut başına düşen ortalama oda sayısı

age: 1940'tan önce inşa edilen sahibi tarafından kullanılan birimlerin oranı

dis: beş Boston istihdam merkezine ağırlıklı mesafeler

rad: radyal otoyollara erişilebilirlik indeksi

tax: 10.000 \$ başına tam değerli emlak vergisi oranı

ptratio: şehre göre öğrenci-öğretmen oranı

b: $1000(Bk - 0.63)^2$ burada Bk, şehre göre siyahların oranıdır

lstat: Nüfusun % alt durumu

Sayısal değişkenlerin özet istatistikleri Tablo 1’de gösterilmektedir.

Table 1: Sayısal Verilerin Özet İstatistikleri

	Ortalama	Standard Deviasyon	Ortanca	Min	Maks
age	66.6643617	28.1949128	73.25000	2.90000	98.90000
b	376.0301064	49.1429200	392.78500	70.80000	396.90000
crim	0.4580199	0.7004628	0.12809	0.00632	3.53501
dis	4.0223420	1.8490212	3.75370	1.34590	9.22290
induz	9.8032979	7.0420319	8.14000	0.46000	25.65000
lstat	11.3301596	5.8795173	10.28500	1.73000	30.81000
nox	0.5216867	0.1058005	0.51000	0.39800	0.87100
ptratio	18.1148936	2.2726440	18.00000	12.60000	21.20000
rad	4.1276596	1.3423106	4.00000	1.00000	8.00000
rm	6.2793511	0.5882691	6.15250	4.90300	8.37500
tax	323.1968085	77.1222974	307.00000	188.00000	469.00000
zn	10.2712766	22.1255128	0.00000	0.00000	95.00000

Kategorik değişkenler incelendiğinde tek kategorik değişken olan char değişkeni ile Charles Nehri kukla değişkenidir (yol nehri sınırlıyorsa 1; aksi takdirde 0).

Table 2: Kategorik Verilerin Özet İstatistikleri

Değişken	Sayı ve Yüzde
chas	6 (3.2%)

3 Yöntem ve Veri Analizi

Boston’da bulunan evlerin median değerlerini tahmin eden model için multiple linear regresyon modeli kullanılmıştır. Model ilk olarak değerlendirildiğinde anlamlı bir model olduğu görülmüştür ($F(13,174)=78,15$, $p<0,001$).

Anlamlı bağımsız değişkenler değerlendirildiğinde crim(kişi başına düşen suç oranı), rm (konut başına düşen ortalama oda sayısı), age (1940’tan önce inşa edilen sahibi tarafından kullanılan birimlerin oranı), dis (beş Boston istihdam merkezine ağırlıklı mesafeler), rad (radyal otoyollara erişilebilirlik indeksi), tax (10.000 \$ başına tam değerli emlak vergisi oranı), ptraito (şehre göre öğrenci-öğretmen oranı), b (şehre göre siyahların oranıdır) ve lstat (Nüfusun % alt durumu) değişkenlerinin ev median fiyatlarını açıklamak için anlamlı olduğu görülmüştür.

Modelin açıklayabilirliği %84,28 olarak bulunmuştur, yani bağımsız değişkenler bağımlı değişkendeki varyansın %84,28’ini açıklamaktadır.

Multiple lineer regresyon modeli ise şu şekilde bulundu:

$$EvMedianFiyati = \beta_0 + \beta_1 crim + \beta_2 rm + \beta_3 age + \beta_4 dis + \beta_5 rad + \beta_6 tax + \beta_7 ptratio + \beta_8 b + \beta_9 lstat$$

Bağımsız değişkenlerin bağımlı değişken üzerine etkisi şöyle açıklanabilir:

Kişi başına suç oranında 1 birim artında median ev değeri 1484,4 artmaktadır.

Konut başına düşen ortalama oda sayısında 1 birim artında median ev değeri 9131,8 artmaktadır.

1940'tan önce inşa edilen sahibi tarafından kullanılan birimlerin oranında 1 birim artış median ev değerinde 38,9 azalmaya neden olmaktadır.

Beş Boston istihdam merkezine ağırlıklı mesafelerinde 1 birim artış median ev değerinde 1088,676 azalışa neden olmaktadır

Radyal otoyollara erişilebilirlik indeksinde 1 birim artış median ev değerlerinde 484,6 artışa neden olmaktadır.

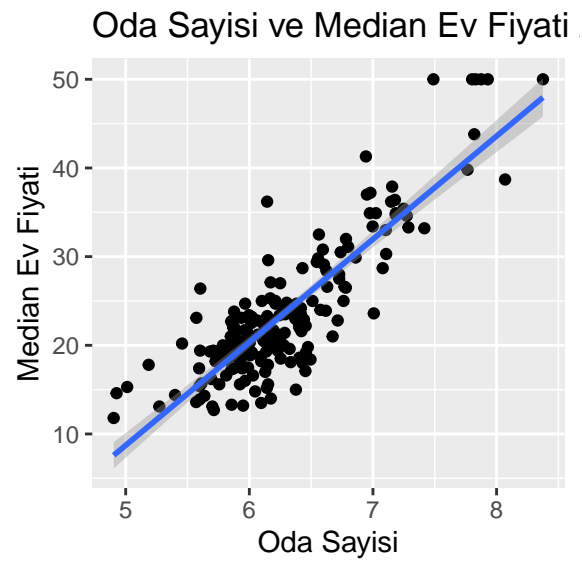
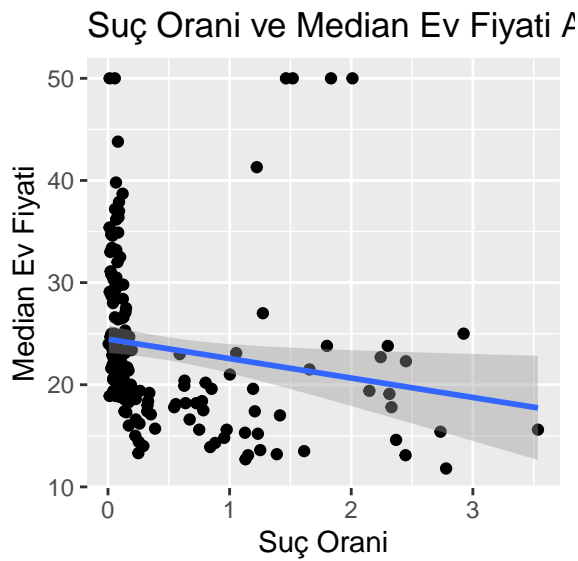
10.000 \$ başına tam değerli emlak vergisi oranında 1 birim artış median ev değerlerinde 19,1 azalışa neden olmaktadır.

Şehre göre öğrenci-öğretmen oranında 1 birim artış median ev değerlerinde 647,3 azalışa neden olmaktadır.

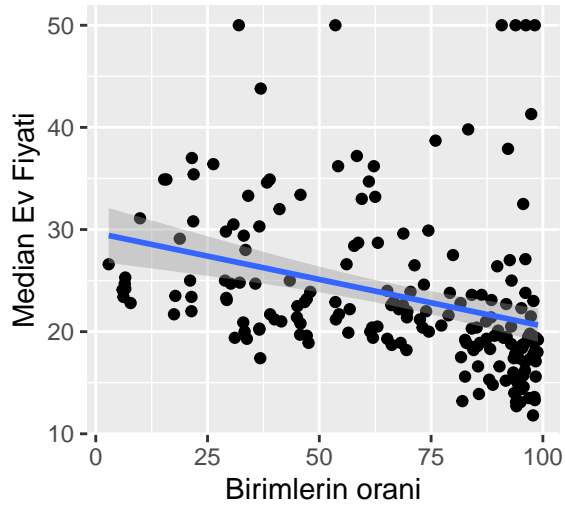
Şehre göre siyahların oranında 1 birim artış median ev değerlerinde 19,4 artışa neden olmaktadır.

Nüfusun % alt durumunda 1 birim artış median ev değerlerinde 221,3 azalışa neden olmaktadır.

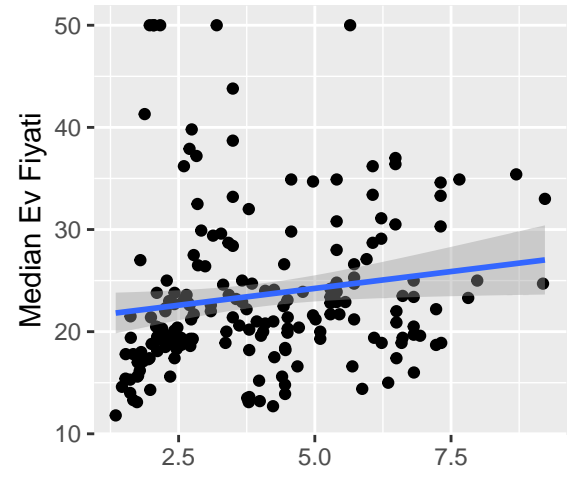
Grafikle anlamlı bağımsız değişkenlerin her birinin median ev fiyatları ile ilişki saçılım grafiği ile incelendiğinde ise:



1940'tan önce inşa edilen sahibi

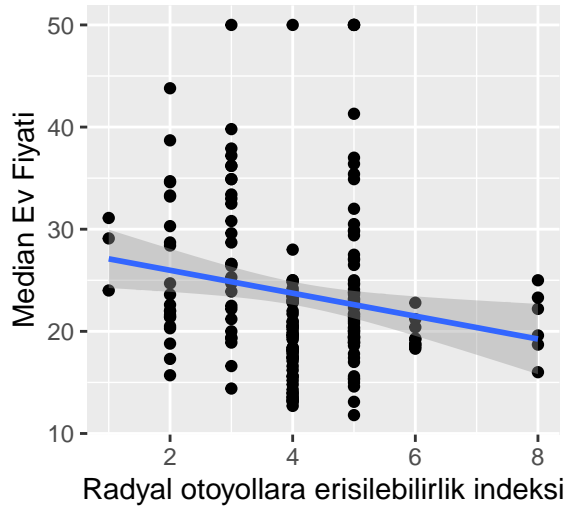


Bes Boston istihdam merkezine

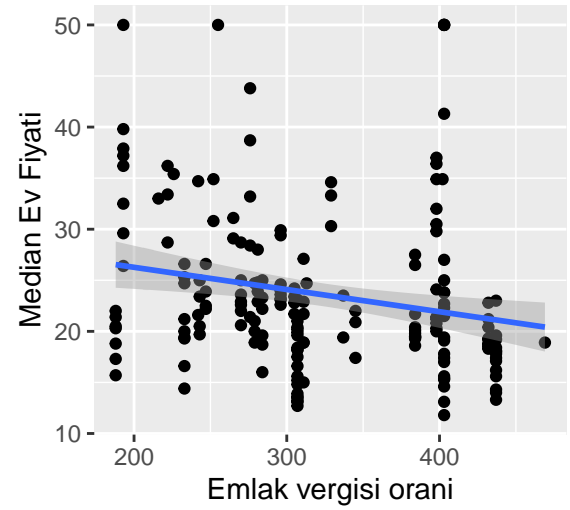


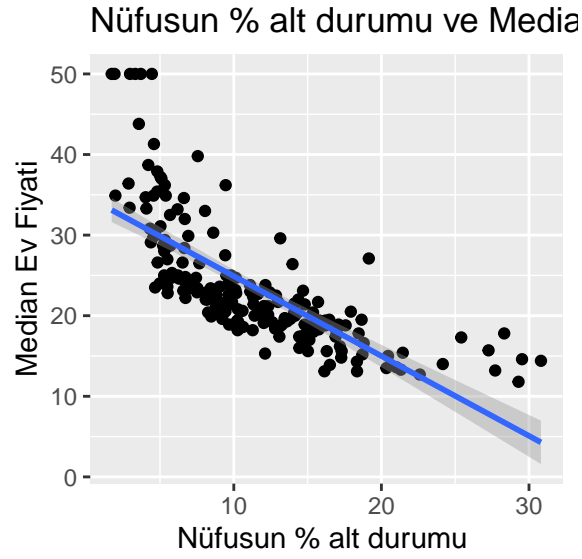
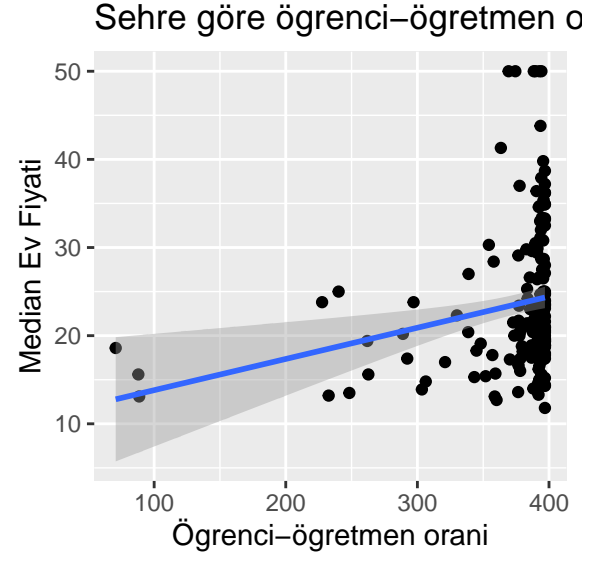
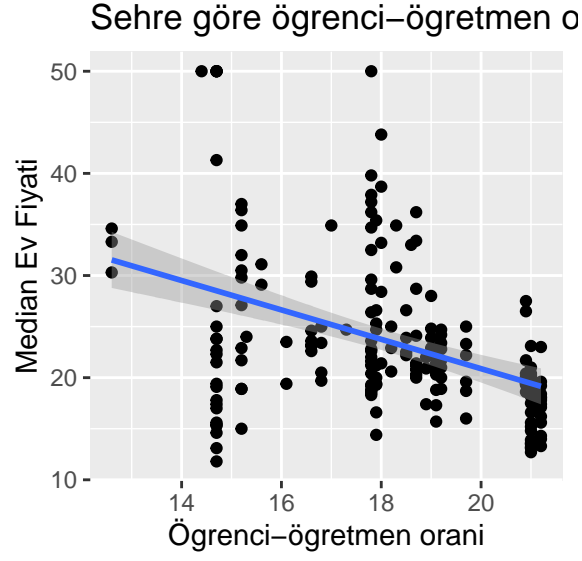
Bes Boston istihdam merkezine ağırlıklı mes

Radyal otoyollara erişilebilirlik in



10.000 \$ başına tam değerli em





4 Tartışma

Ev fiyatı tahmininde median ev fiyatını etkileyen faktörler olarak suç oranı, ortalama oda sayısı, istihdam yerlerine mesafe, otoyollara erişim, emlak vergisi oranı, öğrenci-öğretmen oranı, sayahların oranı ve alt statüdeki nüfus oranının etkili olduğu görülmektedir.

5 Kaynakça

The Boston housing dataset. Web sayfası: <https://www.cs.toronto.edu/~delve/data/boston/bostonDetail.html> Erişim tarihi: 14.04.2023

Madhuri, C. R., Anuradha, G., & Pujitha, M. V. (2019, March). House price prediction using regression techniques: A comparative study. In 2019 International conference on smart structures and systems (ICSSS) (pp. 1-5). IEEE.

Harrison Jr, D., & Rubinfeld, D. L. (1978). Hedonic housing prices and the demand for clean air. *Journal of environmental economics and management*, 5(1), 81-102.

Ozdemir, M., Yildiz, K., & Büyüktanir, B. (2022). Derin Öğrenme ile Konut Fiyat Tahminlemesi: Türkiye'deki Sakarya İli Üzerinde Bir Çalışma. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*.

Afşar, A., Yilmazel, Ö., & Yilmazel, S. (2018). KONUT FİYAT TAHMİNİNDE YAPAY SİNİR AĞLARI YÖNTEMİNİN KULLANILMASI. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*.