

ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

DOKTORA TEZİ

BÜYÜK VERİ ANALİZİNDE ETKİN İSTATİSTİKSEL YÖNTEMLER

Serdar Cihat GÖREN

İSTATİSTİK ANABİLİM DALI

ANKARA
2023

Her hakkı saklıdır

ÖZET

Doktora Tezi

BÜYÜK VERİ ANALİZİNDE ETKİN İSTATİSTİKSEL YÖNTEMLER

Serdar Cihat GÖREN

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Olçay ARSLAN

Bu tez çalışmasında, büyük veri kullanarak hedonik regresyon modeli ile bir ürünün fiyatını etkileyen faktörleri tespit edip her bir özelliğin fiyata olan katkısını ölçmek amaçlanmıştır. Büyük veri kaynağından elde edilen dizüstü bilgisayar ürün grubuna ait fiyat ve özellik bilgilerinden yararlanarak yapılan analizler çerçevesinde En Küçük Kareler (EKK) ve robust (Huber M ve MM tahmin edici) tahmin yöntemleri karşılaştırılmış ve en etkin tahmin edici ön plana çıkarılmıştır. Verilerde gözlenebilecek aykırı değerlerin etkisini en aza indiren robust yöntemler kullanılarak verilerdeki aykırı değerlerin analiz sonucunu en az şekilde etkilemesi hedeflenmiştir. Çalışmada kullanılan veri seti çerçevesinde; dizüstü bilgisayarın fiyatını ve özelliklerinin fiyata olan katkısını en etkin şekilde tahmin eden modelin Tam Logaritmik robust hedonik regresyon modeli olduğu gözlenmiştir.

Temmuz 2023, 83 sayfa

Anahtar Kelimeler: Büyük veri, web kazıma, hedonik regresyon, robust regresyon, değişken seçimi.

ABSTRACT

Ph.D. Thesis

EFFECTIVE STATISTICAL METHODS IN BIG DATA ANALYSIS

Serdar Cihat GÖREN

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Prof. Dr. Olçay ARSLAN

This thesis aims to use big data in order to determine the features affecting the price of a product with the hedonic regression model and to estimate the contribution of each feature to the price by using robust regression estimation methods. For the analysis, the price and features of the laptop product groups were obtained from the big data source by using the web scraping method. Four different alternatives of the hedonic regression model are used to determine the features affecting the price of the laptops. The contribution of each feature to the laptop price is estimated by using the robust estimation methods (M and MM) and the Ordinary Least Squares (OLS) estimation method, and the resulting estimates are compared for both methods. By using robust methods that minimize the effect of outliers that can be observed in the data, it is aimed that the outliers in the data will have the least effect on the analysis result. In the framework of the data set used in the study, it is observed that the effective model is the Logarithmic Robust Hedonic Regression Model.

July 2023, 83 pages

Keywords: Big data, web scraping, hedonic regression, robust regression, variable selection.

TEŞEKKÜR

Doktora eğitim hayatıma başladığım ilk günden itibaren her zaman yanımda olupengin bilgi ve tecrübeleriyle beni yönlendirip her konuda desteğini gösteren, sadece akademik açıdan değil hayatın her alanıyla bana yol gösterici olan kendisi ile birlikte çalışmanın benim için büyük bir şans olduğunu düşündüğüm çok değerli danışman hocam Sayın Prof. Dr. Olçay ARSLAN'a (Ankara Üniversitesi İstatistik Anabilim Dalı) saygı, sevgi ve şükranlarımı sunarım.

Tez hazırlama sürecinde derin bilgi ve tecrübeleriyle her konuda desteklerini esirgemeyen Sayın Prof. Dr. Mehmet YILMAZ'a (Ankara Üniversitesi İstatistik Anabilim Dalı) ve Sayın Doç. Dr. Fatma Zehra DOĞRU'ya (Giresun Üniversitesi İstatistik Anabilim Dalı) saygılarımı ve teşekkürlerimi iletirim.

Doktora eğitim sürecimde görev yapmakta olduğum Türkiye İstatistik Kurumu'ndaki yöneticilerim ve tüm çalışma arkadaşlarıma saygılarımı ve teşekkürlerimi sunarım.

Maddi-manevi, her zaman her konuda destekleriyle, varlıklarıyla ve sonsuz sevgileriyle beni hiçbir zaman yalnız bırakmayan sevgili babam Fehmi GÖREN, sevgili annem Serpil GÖREN ve sevgili ablam Ayşegül GÖREN'e sonsuz teşekkürlerimi bir borç bilirim.

Serdar Cihat GÖREN
Ankara, Temmuz 2023

İÇİNDEKİLER

TEZ ONAY SAYFASI	
ETİK.....	i
ÖZET.....	ii
ABSTRACT	iii
TEŞEKKÜR	iv
SİMGELER DİZİNİ	vii
ŞEKİLLER DİZİNİ	viii
ÇİZELGELER DİZİNİ	x
1. GİRİŞ.....	1
2. TEMELLER VE KAYNAK ÖZETLERİ	4
3. MATERYAL VE YÖNTEM	7
3.1 Büyük Veri.....	7
3.2 Web Kazıma	8
3.3 Regresyon Analizi	9
3.3.1 Çoklu doğrusal regresyon	10
3.4 Hedonik Regresyon	11
3.4.1 Hedonik regresyon model alternatifleri.....	14
3.4.1.1 Tam doğrusal model (DogDog)	15
3.4.1.2 Tam logaritmik model (LogLog).....	15
3.4.1.3 Doğrusal logaritmik model (DogLog).....	16
3.4.1.4 Logaritmik doğrusal model (LogDog).....	17
3.5 Tahmin Yöntemleri.....	18
3.5.1 En Küçük Kareler (EKK) tahmin edicileri	18
3.5.2 Robust (Dayanıklı) istatistiksel yöntemler	19
3.5.2.1 Huber M tahmin edicisi	20
3.5.2.2 MM tahmin edicisi.....	22
3.6 Değişken Seçim Yöntemleri.....	23
3.6.1 Ridge regresyon	23
3.6.2 Lasso regresyon.....	24
3.6.3 Elastiknet	25
4. ARAŞTIRMA BULGULARI	27
4.1 Tam Logaritmik Model (LogLog)	29
4.1.1 Değişkenlerin önem düzeyi	30
4.1.2 Korelasyon analizi	34
4.1.3 Korelasyon analizi sonrası değişkenlerin önem düzeyi.....	37
4.1.4 Veride aykırı değer olması durumu	41
4.2 Logaritmik Doğrusal Model (LogDog)	42
4.2.1 Değişkenlerin önem düzeyi	43
4.2.2 Korelasyon analizi	46
4.2.3 Korelasyon analizi sonrası değişkenlerin önem düzeyi.....	48
4.2.4 Veride aykırı değer olması durumu	52
4.3 Tam Doğrusal Model (DogDog).....	53
4.3.1 Değişkenlerin önem düzeyi	54
4.3.2 Korelasyon analizi	57
4.3.3 Korelasyon analizi sonrası değişkenlerin önem düzeyi.....	59
4.3.4 Veride aykırı değer olması durumu	63
4.4 Doğrusal Logaritmik Model (DogLog)	64

4.4.1 Değişkenlerin önem düzeyi	65
4.4.2 Korelasyon analizi	68
4.4.3 Korelasyon analizi sonrası değişkenlerin önem düzeyi.....	70
4.4.4 Veride aykırı değer olması durumu	74
4.5 Karşılaştırma	75
5. SONUÇ	78
KAYNAKLAR	80
ÖZGEÇMİŞ.....	83

SİMGELER DİZİNİ

$\psi(\cdot)$	ρ fonksiyonun birinci türevi
ε	Regresyon hata terimi
β	Regresyon katsayısı
$\rho(\cdot)$	Robust amaç fonksiyonu
ln	e tabanında logaritma
x	Regresyon bağımsız değişken
y	Regresyon bağımlı/yanıt değişkeni

Kısaltmalar

DogDog	Tam Doğrusal Model
DogLog	Doğrusal Logaritmik Model
EKK	En Küçük Kareler
Eurostat	Avrupa İstatistik Ofisi
EXT	External
INT	Internal
LAD	Least Absolute Deviations
Lasso	Least Absolute Shrinkage and Selection Operator
LMS	Least Median Square
LogDog	Logaritmik Doğrusal Model
LogLog	Tam Logaritmik Model
LTS	Least Trimmed Squares
M	En çok olabilirlik tipi
MM	İki aşamalı M
NLP	Doğal Dil İşleme
TÜFE	Tüketici Fiyat Endeksi

ŞEKİLLER DİZİNİ

Şekil 4.1 Veri deseni	27
Şekil 4.2 Ridge önemli değişkenler (LogLog).....	31
Şekil 4.3 Lasso önemli değişkenler (LogLog).....	31
Şekil 4.4 Elastiknet önemli değişkenler (LogLog)	32
Şekil 4.5 EKK önemli değişkenler (LogLog)	32
Şekil 4.6 Robust önemli değişkenler (LogLog).....	33
Şekil 4.7 Korelasyon grafiği (LogLog).....	36
Şekil 4.8 Korelasyon analizi sonrası Ridge önemli değişkenler (LogLog)	38
Şekil 4.9 Korelasyon analizi sonrası Lasso önemli değişkenler (LogLog).....	38
Şekil 4.10 Korelasyon analizi sonrası Elastiknet önemli değişkenler (LogLog).....	39
Şekil 4.11 Korelasyon analizi sonrası EKK önemli değişkenler (LogLog).....	39
Şekil 4.12 Korelasyon analizi sonrası Robust önemli değişkenler (LogLog).....	40
Şekil 4.13 Ridge önemli değişkenler (LogDog)	43
Şekil 4.14 Lasso önemli değişkenler (LogDog).....	44
Şekil 4.15 Elastiknet önemli değişkenler (LogDog).....	44
Şekil 4.16 EKK önemli değişkenler (LogDog).....	45
Şekil 4.17 Robust önemli değişkenler (LogDog).....	45
Şekil 4.18 Korelasyon grafiği (LogDog)	47
Şekil 4.19 Korelasyon analizi sonrası Ridge önemli değişkenler (LogDog).....	49
Şekil 4.20 Korelasyon analizi sonrası Lasso önemli değişkenler (LogDog)	49
Şekil 4.21 Korelasyon analizi sonrası Elastiknet önemli değişkenler (LogDog).....	50
Şekil 4.22 Korelasyon analizi sonrası EKK önemli değişkenler (LogDog)	50
Şekil 4.23 Korelasyon analizi sonrası EKK önemli değişkenler (LogDog)	51
Şekil 4.24 Ridge önemli değişkenler (DogDog).....	54
Şekil 4.25 Lasso önemli değişkenler (DogDog)	55
Şekil 4.26 Elastiknet önemli değişkenler (DogDog).....	55
Şekil 4.27 EKK önemli değişkenler (DogDog)	56
Şekil 4.28 Robust önemli değişkenler (DogDog)	56
Şekil 4.29 Korelasyon grafiği (DogDog).....	58
Şekil 4.30 Korelasyon analizi sonrası Ridge önemli değişkenler (DogDog).....	60
Şekil 4.31 Korelasyon analizi sonrası Lasso önemli değişkenler (DogDog).....	60
Şekil 4.32 Korelasyon analizi sonrası Elastiknet önemli değişkenler (DogDog)	61

Şekil 4.33 Korelasyon analizi sonrası EKK önemli değişkenler (DogDog).....	61
Şekil 4.34 Korelasyon analizi sonrası Robust önemli değişkenler (DogDog).....	62
Şekil 4.35 Ridge önemli değişkenler (DogLog)	65
Şekil 4.36 Lasso önemli değişkenler (DogLog).....	66
Şekil 4.37 Elastiknet önemli değişkenler (DogLog).....	66
Şekil 4.38 EKK önemli değişkenler (DogLog).....	67
Şekil 4.39 Robust önemli değişkenler (DogLog).....	67
Şekil 4.40 Korelasyon grafiği (DogLog)	69
Şekil 4.41 Korelasyon analizi sonrası Ridge önemli değişkenler (DogLog).....	71
Şekil 4.42 Korelasyon analizi sonrası Lasso önemli değişkenler (DogLog).....	72
Şekil 4.43 Korelasyon analizi sonrası Elastiknet önemli değişkenler (DogLog).....	72
Şekil 4.44 Korelasyon analizi sonrası EKK önemli değişkenler (DogLog)	73
Şekil 4.45 Korelasyon analizi sonrası Robust önemli değişkenler (DogLog)	73

ÇİZELGELER DİZİNİ

Çizelge 4.1 Değişken tanımları	28
Çizelge 4.2 Tahmin sonuçları (LogLog).....	29
Çizelge 4.3 Tahmin yöntemlerinin model standart hataları (LogLog)	30
Çizelge 4.4 Değişken önem düzeyi sonuçları (LogLog).....	34
Çizelge 4.5 Bağımsız değişkenlerin VIF değerleri (LogLog).....	35
Çizelge 4.6 Bağımsız değişkenlere ait korelasyon matrisi (LogLog)	35
Çizelge 4.7 Korelasyon analizi sonrası tahmin sonuçları (LogLog).....	37
Çizelge 4.8 Korelasyon analizi sonrası model standart hataları (LogLog).....	37
Çizelge 4.9 Korelasyon analizi sonrası değişken önem düzeyi sonuçları (LogLog)	41
Çizelge 4.10 Veride aykırı değer olduğunda model standart hataları (LogLog)	41
Çizelge 4.11 Tahmin sonuçları (LogDog)	42
Çizelge 4.12 Tahmin yöntemlerinin model standart hataları (LogDog)	43
Çizelge 4.13 Değişken önem düzeyi sonuçları (LogDog)	46
Çizelge 4.14 Bağımsız değişkenlerin VIF değerleri (LogDog)	46
Çizelge 4.15 Bağımsız değişkenlere ait korelasyon matrisi (LogDog).....	47
Çizelge 4.16 Korelasyon analizi sonrası tahmin sonuçları (LogDog)	48
Çizelge 4.17 Korelasyon analizi sonrası model standart hataları (LogDog)	48
Çizelge 4.18 Korelasyon analizi sonrası değişken önem düzeyi sonuçları (LogDog)....	52
Çizelge 4.19 Veride aykırı değer olduğunda model standart hataları (LogDog).....	52
Çizelge 4.20 Tahmin sonuçları (DogDog)	53
Çizelge 4.21 Tahmin yöntemlerinin model standart hataları (DogDog)	54
Çizelge 4.22 Değişken önem düzeyi sonuçları (DogDog).....	57
Çizelge 4.23 Bağımsız değişkenlerin VIF değerleri (DogDog).....	57
Çizelge 4.24 Bağımsız değişkenlere ait korelasyon matrisi (DogDog)	58
Çizelge 4.25 Korelasyon analizi sonrası tahmin sonuçları (DogDog).....	59
Çizelge 4.26 Korelasyon analizi sonrası model standart hataları (DogDog).....	59
Çizelge 4.27 Korelasyon analizi sonrası değişken önem düzeyi sonuçları (DogDog) ...	63
Çizelge 4.28 Veride aykırı değer olduğunda model standart hataları (DogDog)	63
Çizelge 4.29 Tahmin sonuçları (DogLog)	64
Çizelge 4.30 Tahmin yöntemlerinin model standart hataları (DogLog)	65
Çizelge 4.31 Değişken önem düzeyi sonuçları (DogLog)	68
Çizelge 4.32 Bağımsız değişkenlerin VIF değerleri (DogLog)	68

Çizelge 4.33 Bağımsız değişkenlere ait korelasyon matrisi (DogLog).....	69
Çizelge 4.34 Korelasyon analizi sonrası tahmin sonuçları (DogLog)	70
Çizelge 4.35 Korelasyon analizi sonrası model standart hataları (DogLog)	70
Çizelge 4.36 Korelasyon analizi sonrası değişken önem düzeyi sonuçları (DogLog)....	74
Çizelge 4.37 Veride aykırı değer olduğunda model standart hataları (DogLog).....	74
Çizelge 4.38 Model standart hataları	75
Çizelge 4.39 Veride aykırı değer olduğunda model standart hataları	76
Çizelge 4.40 Tüm modeller için MAE değerleri.....	76
Çizelge 4.41 Tüm modeller için RMSE değerleri.....	76
Çizelge 4.42 Tüm modeller için R^2 değerleri.....	76
Çizelge 4.43 Tüm modeller için en önemli 3 değişken.....	77
Çizelge 4.44 Tam logaritmik robust tahmin edici sonuçları	77

1. GİRİŞ

İstatistik üreticileri kullanıcılarına kaliteli ve doğru istatistikler sunmayı amaçlar. Birçok çalışmada olduğu gibi istatistik üretimi de maliyet, insan kaynağı ve zaman gerektirmektedir. Anket ile veri toplama yöntemi geleneksel ve en yaygın kullanılan yöntemlerdendir. Bunun yanı sıra idari kayıt ve büyük veri gibi veri toplama yöntemleri de yaygın olarak kullanılmaktadır. İdari kayıt ve büyük veri ile veri toplamak maliyet, insan kaynağı ve zaman gibi gereksinimleri oldukça aza indirmektedir. Ayrıca anket ile meydana gelen cevaplayıcılar üzerindeki yükü de büyük ölçüde azaltmaktadır. Veri hacminin büyük olması doğru ve güvenilir istatistikler üretilmesine de imkan sağlamaktadır.

Teknoloji sektöründe dizüstü bilgisayarların pazar payı oldukça yüksektir. Günümüzde ciddi satış rakamlarına sahip olan bu ürünler artık bir ihtiyaç haline gelmiştir. Dizüstü bilgisayar üreticisi birçok marka bulunmakta ve bu piyasada ciddi bir rekabet ortamı yaratmaktadır. Dizüstü bilgisayar üreticileri, müşteri profillerini belirleyerek bu müşteri profillerine uygun özelliklerde bilgisayarlar üretmekte ve satışa sunmaktadır. Tüketiciler ise kendi kullanımına ve bütçesine uygun, kaliteli bir dizüstü bilgisayar satın almak amacındadır. Dizüstü bilgisayar üreticilerinin ve tüketicilerinin buluşma noktası olan teknoloji mağazalarında yapılan kampanyalar, promosyonlar, reklamlar ve birden fazla marka model olması tüketicilerin satın alışlarındaki karar verme süreçlerini zorlaştırmaktadır. Bu rekabet ortamı tüketicilerin yanlış tercihle dizüstü bilgisayar satın alma riskini de artırmaktadır. Bu noktada tüketicilerin bilgisayar markalarını, özelliklerini ve fiyatlarını detaylı şekilde araştırmaları faydalı olacaktır.

Tüketiciler dizüstü bilgisayar fiyatlarını ve modellerini incelemek istediklerinde, teknoloji marketlerin internet sitelerini tercih etmektedirler. Ama kullanımlarına uygun bilgisayarın ne olduğunu ve bu bilgisayar için en doğru fiyatın ne olması gerektiği noktasında gerekli bilgiye her zaman sahip olamayabilirler. Bu çalışmanın amacı teknolojik marketlerin internet fiyatlarından yararlanarak dizüstü bilgisayar fiyatlarını

etkileyen özellikleri belirlemek ve her bir özelliğin fiyata olan etkisini belirleyerek araştırmacılara sunmaktır.

Tez çalışmasının amacı doğrultusunda web kazıma yöntemi ile Türkiye’de faaliyet gösteren zincir teknoloji marketlerin web sitelerinden elde edilen dizüstü bilgisayar fiyat ve özellik verisi kullanılmış olup bu veri setinden faydalanılarak dizüstü bilgisayarın fiyatını etkileyen her bir özelliğin fiyat tahmini yapılmaya çalışılmıştır. Kurulan hedonik regresyon model alternatiflerine uygulanan EKK ve robust yöntemler ile fiyat tahminleri elde edilmiştir. Bu analizlere ek olarak dizüstü bilgisayar özelliklerinin birbiri arasındaki ilişki olup olmadığını tespit etmek amacıyla korelasyon analizi yapılmıştır. Ayrıca hangi değişkenlerin en önemli dizüstü bilgisayar özelliği olduğunu ön plana çıkarmak amacıyla değişken önem düzeyleri ölçülmüştür.

Elde edilen veri setinde yaşanabilecek potansiyel hatalı veri girişi ve ölçüm hatası gibi sorunlarda EKK yöntemlerinin etkin sonuç vermediği bilinmektedir. Tez çalışmasında; hatalı veri girişi ve ölçüm hatası gibi kaygıları en aza indirmek için robust istatistiksel yöntemlere odaklanılmıştır. Çünkü robust yöntemler hatalı veriler ile meydana gelebilecek aykırı değer sorunlarını en aza indirerek bu kaygıları ortadan kaldırmaktadır. Robust yöntemlerin bu tarz çalışmalarında yaygın olarak kullanılması aykırı değer potansiyeli olan veri setlerinde etkin sonuçlar elde etmeye yarayacaktır.

Tez çalışmasının “Temeller ve Kaynak Özetleri” bölümünde tez çalışması kapsamında yapılan literatür çalışması ve benzer çalışmalardan bahsedilmiştir.

Tez çalışmasının devamında “Materyal ve Yöntem” bölümünde öncelikle büyük verinin tanımı ve kullanım alanlarından bahsedilmiştir. Tez çalışmasında kullanılan veri setinin elde edilme yöntemi olan web kazıma (web scraping) tekniği ve regresyon analizi tanıtılmıştır. Ardından tez çalışmasının uygulama safhasında fiyat tahmininde kullanılan hedonik regresyon modeli ve model alternatifleri tanıtılarak hedonik regresyonda parametre tahminlerinde kullanılan EKK ve robust istatistiksel yöntemleri incelenmiştir. Ayrıca ilgili bölümün son kısmında fiyat tahminini en çok etkileyen değişkenleri ön plana çıkarmak amacıyla değişken seçim yöntemleri ele alınmıştır.

Tez çalışmasının “Araştırma Bulguları” bölümünde elde edilen veri seti çerçevesinde kurulan hedonik regresyon model alternatifleri ile birlikte tahmin sonuçları verilmiştir. “Sonuç” bölümünde çalışma sonucunda elde edilen bulgular değerlendirilmiştir.

2. TEMELLER VE KAYNAK ÖZETLERİ

Bir mal veya hizmetin deęerini bileşenlerine ayırarak belirlemek için kullanılan yöntemlerden biri hedonik regresyon yöntemidir. Hedonik regresyon yöntemine dair birçok akademik çalışma bulunmaktadır.

Colweell ve Dilmore (1999) çalışmasına göre; hedonik regresyonun ilk kullanıcısının Haas (1922) olduğu ifade edilmektedir. Haas (1922) yapmış olduğu çalışmada tarım arazisi fiyatını şehir merkezine uzaklık ve şehir merkezi büyüklüğü deęişkenleri kullanarak hedonik regresyon modeli ile tahmin etmiştir.

Sheppard (1997) çalışmasına göre; ilerleyen dönemlerde Waugh (1929)'ın ürünlerin fiyatına kalitenin etkisinin ölçüldüğü ilk çalışma olduğu ifade edilmiştir. Court (1939)'ın "hedonik" terimini ilk kullanan çalışma olduğu, heterojen malları karakterize etmek ve bireysel tercihlere yönelik talepleri belirlemek için hedonik modellerin kullanıldığı belirtilmiştir. Griliches (1961, 1971) çalışmalarında malların heterojenliğini belirlemek için hedonik analiz ve tekniklerini daha geniş olarak kullandığı ifade edilmiştir.

Yapılan literatür çalışmasında hedonik regresyon ile fiyat tahmininin geniş bir kullanım alanı olduğu görülmüştür. Yakın geçmişte hedonik regresyon modelinin sıklıkla konut fiyatlarını ve konut özelliklerinin fiyata olan etkisini belirlemek üzerine kullanıldığı gözlenmektedir.

Diewert vd. (2010), Diewert vd. (2014), Hülagu vd. (2016), Jiang vd, (2014), Selim (2008), konut fiyatları ve endeksleri için hedonik regresyon modelini kullanarak konut fiyatlarını etkileyen özellikleri belirleyip EKK yöntemi ile hangi özelliğin ne kadar katkısı olduğunu analiz etmişlerdir.

Konut dışında farklı ürünler için de hedonik regresyon modeli kullanılarak analizler yapılmıştır.

Fixler vd. (1999) hedonik regresyon modelini Amerika'daki Tüketici Fiyat Endeksleri'nde kalite düzeltme yöntemi olarak kullanmışlardır.

Manoel vd. (2009) çalışmasında; kişisel masaüstü bilgisayarların fiyatlarını ayırt eden özellikleri hedonik regresyon modeli ile belirlenmiş ve EKK yöntemi ile hangi özelliğin fiyata ne kadar katkısı olduğu ölçülmüştür.

McCormack (2013) çalışmasında; sıfır arabalar için hedonik regresyon modeli ile fiyatı etkileyen özellikler ön plana çıkarılmış ve EKK yöntemi ile hangi özelliğin fiyata ne kadar katkısı olduğu ölçülmüştür.

Hedonik regresyon ile yapılmış birçok çalışmada tahminler EKK yöntemi ile yapılmıştır. Fakat veride aykırı değerlerin ve hatalı verilerin olduğu durumlarda EKK yönteminin etkin sonuçlar vermediği bilinmektedir. Bu durumda robust istatistiksel yöntemler daha etkin sonuçlar vermektedir. Çünkü robust istatistiksel yöntemler aykırı değerlere karşı sağlamdır.

Dehon vd. (2006) robust hedonik regresyon üzerine bir çalışma yürütmüşlerdir. Çalışmada Harry Potter oyun kartları verisini kullanarak robust yöntemlerden LTS (Least Trimmed Squares- En Küçük Budanmış Kareler) ve M-tahmin edicileri ile EKK yöntemini karşılaştırmış ve aykırı değerlerin olması durumunda en etkin tahmin edicinin LTS olduğuna kanaat getirmişlerdir.

Bourassa vd. (2013) yapmış oldukları çalışmada konut fiyatları için robust hedonik regresyon modelini kullanarak konut fiyatlarını etkileyen özellikleri belirleyip robust istatistiksel yöntemler ile hangi özelliğin ne kadar katkısı olduğunu analiz etmişlerdir. 3 robust istatistiksel yöntem (2 MM tahmin edici ve 1 S tahmin edici) ve EKK yönteminin etkinliği karşılaştırılmıştır. Çalışma sonucunda robust yöntemlerin EKK yöntemine göre aykırı değerler olması durumunda daha etkin sonuçlar verdiği gözlenmiştir.

Bulut ve Zaman (2018) çalışmasında; Türkiye’de “Vosvos” ya da “Kaplumbağa” olarak da adlandırılan Beetle araçların fiyatını etkileyen faktörleri robust hedonik modellerle belirlemişlerdir. Tahminler için EKK ve robust istatistiksel yöntemlerden; M-tahmin ediciler (Huber, Tukey, Hampel), MM, LTS, LMS (Least Median Square- En Küçük Medyan Kareler) ve LAD (Least Absolute Deviations- En Küçük Mutlak Sapmalar) birbirleri ile karşılaştırılmıştır. En etkin sonuçları veren tahmin edicinin LAD robust yöntemi olduğu gözlenmiştir.

Bu tez çalışmasında teknolojik marketlerden elde edilen dizüstü bilgisayara ait fiyat ve özellik bilgisini içeren büyük veri kullanılmıştır. Hedonik regresyon yöntemi ile belirlenen model alternatifleri ile birlikte EKK ve robust istatistiksel yöntemlerden Huber-M ve MM tahmin edicileri dizüstü bilgisayarın fiyatını etkileyen özellikleri belirlemek ve her bir özelliğin dizüstü bilgisayar fiyatına olan katkısını ölçmek için hesaplanmıştır. Ayrıca değişken seçim yöntemlerinden Lasso, Ridge ve Elastik net yöntemleri kullanılarak kurulan regresyon modellerinde yer alan değişkenlerin önem düzeyi de ön plana çıkarılmıştır.

3. MATERYAL VE YÖNTEM

3.1 Büyük Veri

Gelişen ve değişen dünyamızda istatistiki bilgiye olan talep ve ihtiyaç giderek artmaktadır. İstatistiki bilginin kullanıcılar ile buluşturulması için anket, idari kayıt ve büyük veri gibi veri toplama yöntemleri geliştirilmiştir. Bu veri toplama yöntemleri arasında yer alan büyük veri günümüzde oldukça popüler hale gelmiştir. Çünkü istatistik üreticileri veriyi alandan toplamak yerine hazırda var olan verileri kullanarak sonuca ulaşmayı hedeflemektedir. Böylelikle istatistik üreticileri hem insan kaynaklar, hem mali kaynaklar hem de zaman açısından ciddi fayda sağlamaktadır.

Büyük verinin birçok tanımı olmasına karşılık, genel olarak geleneksel kullanıcıların depolayıp işleyemeyeceği kadar büyüklükte, terabyte perabyte'ler ile ölçülebilen verilerdir. Bu verilere örnek olarak ulaştırma, sosyal medya, market verileri verilebilir.

Aktan (2018) makalesinde; büyük veri kavramının ilk kez Cox ve Ellsworth (1997) çalışmasında kullanıldığı ifade edilmiştir. Bu çalışmada, bilimsel veri görselleştirme çalışması yapılmıştır. Makale içerisinde veri görselleştirme için büyük hacimli verilere ihtiyaç olduğu ve bu verilerin bilgisayar bellek kapasitelerini zorladığından bahsedilmiştir. Web'de bulunan veriler, web sayfaları, web veritabanları, e-postalar, tweet'ler, blog gönderileri, fotoğraflar, videolar vb. şeklinde dağıtılan yapılandırılmış, yarı yapılandırılmış ve yapılandırılmamış nicel ve nitel verilerden oluşur.

Bir veri setinin büyük veri olarak tanımlanabilmesi için veri hacminin büyüklüğünden farklı gereksinimlerin de sağlanması gerekmektedir. Bu gereksinimler literatürde ilk olarak Doug Laney tarafından ön plana çıkarılmıştır. Doug Laney 2001 yılında yayımlanmış olduğu makalesinde büyük veriyi 3V (Volume, Velocity ve Variety) boyutunda ele almıştır. İnternette yer alan veriler genellikle Zettabayt (milyarlarca gigabayt) cinsinden ölçülen büyük hacimle (Volume) karakterize edilir. Bu geniş veri havuzları çeşitli (Variety) biçimlerde gelir ve çeşitli teknolojik standartlara dayanır.

(Basoglu ve White 2015). Ayrıca bu veriler statik değildir ve aşırı hızla (Velocity) üretilir (Goes 2014).

Büyük veri üzerine devam eden çalışmalar neticesinde 3V boyutuna 2V (Veracity ve Value) boyutları da eklenerek 5V kavramı ortaya çıkmıştır. Burada “Volume” veri hacmini ve büyüklüğünü, “Variety” verinin çeşitliliğini, “Velocity” verinin hızını, “Veracity” verinin geçerliliğini, “Value” ise verinin değerini ifade etmektedir.

Büyük veri elde etmenin yollarından biri web kazıma (web scraping) yöntemidir. Web scraping web sitelerinden bilgi çıkartmanın bilgisayar programı tekniğidir. Bir sonraki bölümde web kazıma tekniğinden bahsedilecektir.

3.2 Web Kazıma

Büyük verinin 5V özelliği ve doğruluğu göz önüne alındığında, bu verilerin toplanması ve işlenmesi, geleneksel araştırmacılar veya büyük araştırma ekipleri tarafından manuel olarak neredeyse imkansızdır. Büyük veriyi toplamak için kullanılan teknoloji genellikle web kazıma olarak adlandırılır Web kazıma, büyük verinin analiz edilmesi amacıyla internetten otomatik olarak elde edilmesi ve düzenlenmesi için teknoloji araçlarının kullanılması olarak tanımlanmaktadır (Krotov ve Silva 2018).

Web kazıma, bir web sitesini manuel olarak kopyalamak yerine bazı bilgileri otomatik olarak almak için kullanılan teknikler kümesidir. Web kazımanın amacı, belirli türden bilgileri aramak, ayıklamak ve yeni web sayfalarında bir araya getirmektir. Özellikle kazıyıcılar, yapılandırılmamış verileri dönüştürmeye ve bunları yapılandırılmış veritabanlarına kaydetmeye odaklanır (Vargiu ve Urru 2013).

Web kazıma metodunda web sitesi analizi ve web sitesi taraması oldukça önemlidir. Web sitesi analizi, gerekli verilerin nasıl depolandığını anlamak amacıyla bir web sitesinin veya bir veri deposunun temel yapısının incelenmesini gerektirir. Bu, web mimarisinin temel düzeyde anlaşılmasını gerektirir. Web sitesi taraması, web sitesinde

otomatik olarak gezinen ve gerekli verileri alan bir komut dosyası geliştirmeyi ve çalıştırmayı içerir. Bu tarama uygulamaları veya komut dosyaları genellikle R ve Python gibi programlama dilleri kullanılarak geliştirilir. Gerekli veriler seçilen web deposundan ayrıştırıldıktan sonra, bu verilerin daha fazla analiz edilmesini sağlayacak şekilde temizlenmesi, önceden işlenmesi ve düzenlenmesi gerekir. İlgili verilerin hacmi göz önüne alındığında, zaman kazanmak için programatik bir yaklaşım da gerekli olabilir. R ve Python gibi birçok programlama dili, verileri temizlemek ve düzenlemek için yararlı olan Doğal Dil İşleme (NLP) kitaplıkları ve veri işleme işlevleri içerir (Krotov ve Silva 2018).

Web kazıma için en önemli güncel kullanımlardan biri, işletmelerin rakiplerinin fiyatlandırma faaliyetlerini izlemesidir. Kısa zaman ve minimum manuel çabayla tüm bir site genelinde fiyatlandırma verilerine erişilebilmektedir. Şirketler ayrıca analitik odaklı yatırım için veri elde etmek için web kazımayı kullanır. Web kazıma, her türden istatistiksel ve bilimsel araştırma için merkezi bir araç haline gelmiştir (Snell ve Menaldo 2016).

Bu tez çalışmasında web kazıma yöntemi kullanılarak teknoloji marketlerin dizüstü bilgisayarlar için fiyat ve özellik verisi elde edilmiş ve bu veri setinden yararlanılarak hedonik regresyon modelleri kurulmuştur. Çeşitli teknoloji sitelerinden web kazıma yöntemi ile belirli bir süre veri toplanmıştır. Web sitelerinin yapısı farklı olduğu için Python'da her siteye özel Spider kodu kullanılmıştır. Ayrıca tıklama teknolojisi olarak Selenium, veri toplama teknolojisi olarak Scrapy kütüphanelerinden yararlanılmıştır. Sadece belirli bir süre ve internet sitelerinin veri trafiğini artırmayacak şekilde kullanıldığından internet sitelerinde herhangi bir engelleme ile karşılaşılmamıştır.

3.3 Regresyon Analizi

Regresyon analizi değişkenler arasındaki ilişkiyi modellemek için kullanılan bir yöntemdir. Regresyon analizinin en önemli alanlarından biri değişkenler arasındaki ilişkiyi açıklayabilecek doğru modele karar vermektir. Modele karar verebilmek için

değişkenler arasındaki ilişkinin ön analizinin yapılmasıdır. Bu analize dayalı olarak model belirlenmelidir (Montgomery vd. 2001).

3.3.1 Çoklu doğrusal regresyon

Y bağımlı değişkeni, k sayıda ($k \geq 1$) bağımsız değişkenle ilişkili olduğu durumda model çoklu doğrusal regresyon olarak tanımlanmaktadır. $k=1$ olması durumunda ise Basit Doğrusal Regresyon olarak tanımlanmaktadır. Çoklu doğrusal regresyon yapısı aşağıda verilmektedir:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon . \quad (3.1)$$

Bu model aşağıdaki gibi yazılabilir:

$$y = x^T \beta + \varepsilon . \quad (3.2)$$

$$x = \begin{bmatrix} 1 \\ x_1 \\ \cdot \\ x_k \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \beta_k \end{bmatrix}$$

Burada β_j , $j=0,1,\dots,k$ parametreleri regresyon katsayıları olarak bilinir. β_j parametresi x_i ($i \neq k$) bağımsız değişkenleri sabit tutulduğu zaman bağımsız değişkenindeki bir birimlik değişime karşılık gelen y ($y \in R$) yanıt değişkenindeki beklenen değişimi göstermektedir. Bu nedenle, β_j , $j=0,1,\dots,k$ parametreleri genellikle kısmi regresyon katsayıları olarak adlandırılır (Montgomery vd. 2001).

Çoklu doğrusal regresyon modelinde parametre tahmini yapabilmek için $(x_1, y_1), \dots, (x_n, y_n)$ örnekleme ihtiyacımız olacaktır. Örneklem için aşağıdaki örneklem modeli kurulur:

$$y_i = x_i^T \beta + \varepsilon_i, \quad i = 1, \dots, n. \quad (3.3)$$

Burada ε_i 'ler için aşağıdaki varsayımları sağlanır:

- * Hata terimlerinin (ε_i) ortalaması sıfırdır.
- * Hata terimleri (ε_i) ilişkisizdir.
- * Hata terimleri (ε_i) ile x_i 'ler ile ilişkisizdir.

Örneklem modeli için matris gösterimi aşağıda yer almaktadır:

$$Y = X\beta + \varepsilon. \quad (3.4)$$

Burada Y $n \times 1$ 'lik bağımlı değişken vektörü, X $n \times p$ 'lik bağımsız değişken matrisi ve ε $n \times 1$ 'lik hata vektörüdür. ε hata vektörü için yukarıda yer alan koşullarımız $E(\varepsilon) = 0$ ve $\text{cov}(\varepsilon) = \sigma^2 I$ olur.

3.4 Hedonik Regresyon

Hedonizm, insan davranışlarının haz sağlayacak bir biçimde planlanması gerektiğini ve haz verene yönelmenin en uygun davranış biçimi olduğunu öne süren felsefi görüştür (Gillian ve Davashish, 1999). Hedonik terimi ise, hedonizm kelimesinden türemektedir (Hans 2004).

Hedonik fiyat ürün kalitesi sorunlarını değerlendirmek için uzun yıllardır kullanılan istatistiksel bir tekniktir. Hedonik fiyat analizinin amacı, bir ürünün birim fiyatının özelliklerine göre nasıl değiştiğini belirlemek, yani hedonik fiyat fonksiyonunu tahmin etmektir. 1960'ların başlarında çok fazla tercih edilmeyen bu teknik son yıllarda özellikle tüketici fiyat endeksi hesaplamalarında oldukça yaygın olarak kullanılmaktadır.

Belirli ürünler için fiyat ve kalite ilişkisini analiz etmek için literatürde yer alan çok sayıda araştırmada hedonik fiyat yöntemleri kullanılmıştır. Otomobiller, teknolojik ürünler, emlak gibi ürünlerin fiyatlarını özellikleri ile bağdaştırmak için hedonik fiyat yöntemleri kullanılmıştır (Unwin 1999).

Her hedonik fiyat modelinin başlangıç noktası hedonik hipotezdir. Bu hipotezin özü, her malın tüm özelliklerinin bir dizi ile karakterize edilmesidir. Bir hedonik fiyat fonksiyonu, bir ürünün ekonomik olarak ilgili özellikleri ile fiyatı arasındaki denge ilişkisini tanımlar. Bir malın fiyatının, onun belirleyici özelliklerinin ve rastgele bir hata teriminin bir fonksiyonu olduğu varsayılır. Bu hedonik fiyatlar, yeni mallar için fiyatları tahmin etmek, bir malın fiyatındaki kalite değişikliklerini düzeltmek ve farklılaştırılmış ürünlerin tüketici ve üretici değerlemelerini ölçmek için kullanılabilir (Hulten 2003).

Hedonik regresyon, bir mal veya hizmetin değerini bileşenlerine ayırarak belirlemek için kullanılan bir yöntemdir. Her bileşenin değeri daha sonra regresyon analizi ile ayrı ayrı belirlenir (McCormack 2013).

Hedonik regresyon modeli yapı olarak çoklu doğrusal regresyon ile tamamen aynıdır. Hedonik kavramı tüketici tatminine dayanan bir kavram olduğu için malların fiyatları ve özellikleri arasındaki ilişkinin analiz edilmesi ile özdeşleşmiştir.

Tüketici Fiyat Endeksinin (TÜFE) hesaplanmasında yer alan ürün ve hizmetlerdeki kalite değişikliklerinin etkilerini ortadan kaldırmak için hedonik regresyon modelleri istatistik kurumları tarafından kullanılmaktadır. Bu konuda birçok çalışma yer

almaktadır. TÜFE hesaplamaları kapsamında konut, giyim ve elektronik gibi ürün gruplarında sıkça kullanılmaktadır.

Eurostat (2018) dokümanında hedonik regresyonun tanımı “*Genel olarak hedonik regresyon yöntemi, kalite değişiminin fiyat ve ürün özelliklerinin bir fonksiyonu olarak ifade eden bir regresyon denklemine dayandığı anlamına gelir.*” olarak yapılmıştır. Eurostat (2018) dokümanında, Tüketici Fiyat Endeksi kapsamındaki ürünlerin kalite değişimlerini tespit etmek ve düzeltmek amacıyla belli ürün grupları için hedonik regresyon modeli önerilmektedir.

Chow 1960’lı yıllarda hedonik regresyon modelini teknolojik ürünlerde ilk kullanan kişidir. Chow, hedonik regresyon modelini bilgisayar fiyatlarında kalite düzeltme yöntemi olarak uygulamıştır. Micheals, bilgisayar fiyatını harddisk ve ram özellikleriyle modellemiştir. Berndt ve Griliches kalite düzeltmeli bilgisayar fiyatlarını tahmin etmek için hedonik regresyon yöntemlerini kullanmıştır (Lee vd. 2003).

Hedonik fiyat fonksiyonları; üretilen malların kalitesindeki değişiklikleri hesaba katan genel fiyat endekslerinin oluşturulmasında ve heterojen malların özelliklerine yönelik tüketici taleplerinin analizinde olmak üzere iki temel amaç için kullanılmaktadır (Sheppard, 1997). Sheppard’ın tanımlamış olduğu bu amaçlar hedonik regresyon modeli ile dizüstü bilgisayarların özelliklerinin fiyata olan katkılarının ölçülmesi bir nevi bu ürünün özelliklerine yönelik tüketici talebinin nasıl olacağını da bize gösterecektir.

Hedonik fiyat modeli aşağıdaki gibidir:

$$p = h(c_i). \quad (3.5)$$

Yukarıda belirtilen fonksiyon p bir malın fiyatını, $h(c_i)$ ise ilgili malın özelliklerinin hedonik fonksiyonu olarak adlandırılır. Hedonik fonksiyon regresyon analizinde kullanılan parametre tahmin yöntemleri ile tahmin edilir (McCormack 2013).

Bu tez çalışmasında 2020 yılının son çeyreğine ait dizüstü bilgisayar satış fiyatı ve özelliklerinin yer aldığı büyük veri setinden yararlanılmış olup dizüstü bilgisayarın fiyatı ile sahip olduğu özellikler üzerine hedonik fiyat modeli kurulmaktadır. Dizüstü bilgisayarın özelliklerinin fiyat üzerindeki etkisinin ölçülmesi için regresyon modeli ile her bir özelliği açıklayıcı değişken olarak tanımlayarak öğrenmek mümkündür.

3.4.1 Hedonik regresyon model alternatifleri

Literatürdeki çalışmalar incelendiğinde hedonik regresyon model konusunda birçok alternatif bulunmaktadır. Bunlar; Tam Doğrusal Model (DogDog), Tam Logaritmik Model (LogLog), Doğrusal Logaritmik Model (DogLog) ve Logaritmik Doğrusal Model (LogDog)dir.

Model alternatifleri bağımlı ve bağımsız değişkenlerin yapısına göre değişkenlik göstermektedir. Bağımlı ve bağımsız değişkenlerin doğrusal ve logaritmik yapıda olması kurulan hedonik regresyon modelinin çeşitlenmesine ve parametre tahminlerinin etkinliğinin karşılaştırılmasına imkan vermektedir. Ayrıca kurulan hedonik regresyon model tipleri analiz edilen ürünün yapısına göre farklı etkin sonuçlar göstermektedir.

Hedonik regresyon modelinde bağımlı değişken ürünün fiyatı iken bağımsız değişkenler de ürünün özellikleri olmaktadır. Ürünün özelliklerinden oluşan bağımsız değişkenlerin bir kısmı nicel, bir kısmı da nitel değişkenler olabilmektedir. Nicel değişkenler modelin içerisinde doğrudan yer alabilirken nitel değişkenler kukla değişken olarak tanımlanması gerekmektedir. Bu durumda kukla değişken olarak tanımlanan nitel değişkenlerin özellik olarak ilgili üründe bulunması durumunda 1 bulunmaması durumunda 0 değerini almaktadır. Bu tez çalışmasında odaklanılan ürün dizüstü bilgisayar olduğu için dizüstü bilgisayarın fiyatı bağımsız değişken olarak tanımlanırken fiyatı etkileyen özellikler ise bağımlı değişkenler olarak tanımlanmıştır.

Hedonik fiyat fonksiyonunda dört farklı fonksiyonel yapısı bulunmaktadır. Bu modeller aşağıda yer almaktadır:

- Tam Doğrusal Model (DogDog)
- Tam Logaritmik Model (LogLog)
- Doğrusal Logaritmik Model (DogLog)
- Logaritmik Doğrusal Model (LogDog)

3.4.1.1 Tam doğrusal model (DogDog)

Bağımlı değişken ve bağımsız değişkenler arasında tam doğrusal ilişki olması durumunda gerçekleşen hedonik regresyon model yapısıdır. Bu model çoklu doğrusal regresyon ile yapı olarak tamamen aynıdır. Bu tez çalışmasında kullanılan ürün grubu olan dizüstü bilgisayar için bağımlı değişkenin dizüstü bilgisayar fiyatı ile bağımsız değişkenlerin (İşlemci hızı, Ram, Hard disk, Ekran boyutu, İşlemci- Intel i7, İşlemci- Intel i5, İşlemci- Intel i3, Ekran kartı) doğrusal formda olduğu durumudur.

$$P = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p + \varepsilon . \quad (3.6)$$

3.4.1.2 Tam logaritmik model (LogLog)

Bağımlı değişken ve bağımsız değişkenler arasında tam logaritmik ilişki olması durumunda gerçekleşen hedonik regresyon model yapısıdır. Bu tez çalışmasında kullanılan ürün grubu olan dizüstü bilgisayar için bağımlı değişkenin dizüstü bilgisayar fiyatı ile bağımsız değişkenlerin (İşlemci hızı, Ram, Hard disk, Ekran boyutu, İşlemci- Intel i7, İşlemci- Intel i5, İşlemci- Intel i3, Ekran kartı) logaritmik formda olduğu durumudur. Tam logaritmik hedonik regresyon modeli üstel olarak aşağıdaki gibi elde edilir:

$$P = e^\alpha \prod_{i=1}^p X_i^{\beta_i} . \quad (3.7)$$

Daha sonra modelin logaritması alınır ve parametre tahminlerinde kullanılan forma dönüştürülür:

$$\ln[P] = \ln[e^\alpha] + \ln\left[\prod_{i=1}^p X_i^{\beta_i}\right]. \quad (3.8)$$

$$\ln P = \alpha + \sum_{i=1}^p \beta_i \ln X_i. \quad (3.9)$$

$$\ln P = \alpha + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \dots + \beta_p \ln X_p + \varepsilon. \quad (3.10)$$

3.4.1.3 Doğrusal logaritmik model (DogLog)

Bağımlı değişkenin doğrusal ve bağımsız değişkenlerin ise logaritmik olması durumunda gerçekleşen hedonik regresyon model yapısıdır. Bu tez çalışmasında kullanılan ürün grubu olan dizüstü bilgisayar için bağımlı değişkenin dizüstü bilgisayar fiyatının doğrusal, bağımsız değişkenlerin (İşlemci hızı, Ram, Hard disk, Ekran boyutu, İşlemci- Intel i7, İşlemci- Intel i5, İşlemci- Intel i3, Ekran kartı) logaritmik formda olduğu durumdur. Doğrusal logaritmik hedonik regresyon modeli üstel olarak aşağıdaki gibi elde edilir:

$$e^P = e^\alpha \prod_{i=1}^p X_i^{\beta_i}. \quad (3.11)$$

Daha sonra modelin logaritması alınır ve parametre tahminlerinde kullanılan forma dönüştürülür:

$$\ln[e^P] = \ln[e^\alpha] + \ln\left[\prod_{i=1}^p X_i^{\beta_i}\right]. \quad (3.12)$$

$$P = \alpha + \sum_{i=1}^p \beta_i \ln X_i. \quad (3.13)$$

$$P = \alpha + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \beta_3 \ln X_3 + \dots + \beta_p \ln X_p + \varepsilon . \quad (3.14)$$

3.4.1.4 Logaritmik doğrusal model (LogDog)

Bağımlı değişkenin logaritmik ve bağımsız değişkenlerin ise doğrusal olması durumunda gerçekleşen hedonik regresyon model yapısıdır. Bu tez çalışmasında kullanılan ürün grubu olan dizüstü bilgisayar için bağımlı değişkenin dizüstü bilgisayar fiyatının logaritmik, bağımsız değişkenlerin (İşlemci hızı, Ram, Hard disk, Ekran boyutu, İşlemci- Intel i7, İşlemci- Intel i5, İşlemci- Intel i3, Ekran kartı) doğrusal formda olduğu durumudur. Logaritmik doğrusal hedonik regresyon modeli üstel olarak aşağıdaki gibi elde edilir:

$$P = e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p} \quad (3.15)$$

Daha sonra modelin logaritması alınır ve parametre tahminlerinde kullanılan forma dönüştürülür:

$$\ln[P] = \ln \left[e^{\alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p} \right] . \quad (3.16)$$

$$\ln P = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_p X_p + \varepsilon \quad (3.17)$$

McCormack (2013) çalışmasında sıfır araba fiyatlarını etkileyen özelliklerin belirlenmesinde Logaritmik Doğrusal (LogDog) hedonik regresyon modeli kullanmıştır.

Bulut ve Zaman (2018); Türkiye’de “Vosvos” ya da “Kaplumbağa” olarak da adlandırılan Beetle araçların fiyatını etkileyen faktörlerini yukarıdaki dört farklı hedonik modelle analiz etmiştir. Çalışma sonucunda en etkin sonuçların Logaritmik Doğrusal Model (LogDog) ile elde edildiği belirtilmiştir.

Selim (2008); Türkiye’deki konut fiyatlarını belirleyen konut özelliklerini hedonik regresyon modeli ile belirlenmiştir. Çalışmada Logaritmik Doğrusal Model (LogDog) kullanılmış ve tahminler EKK yöntemi ile gerçekleştirilmiştir.

Jiang vd. (2014) çalışmasında ise konut fiyatlarını belirleyen konut özellikler Tam Doğrusal Model (DogDog) hedonik regresyon modeli kurularak EKK yöntemi ile elde edilmiştir.

Bir sonraki bölümde bu tez çalışması kapsamında; hedonik regresyon modelinde kullanılan tahmin yöntemlerinden bahsedilmektedir.

3.5 Tahmin Yöntemleri

3.5.1 En Küçük Kareler (EKK) tahmin edicileri

En Küçük Kareler (EKK) yöntemi, regresyon analizlerinde kullanılmakta ve diğer yöntemlerin de temelini oluşturmaktadır. Stigler (1981), çalışmasında EKK’nın ilk kullanıcılarının Carl Friedrich Gauss ve Adrien Marie Legendre olduğunu fakat hangisinin ilk kullanıcı olduğunun bir tartışma konusu olduğundan bahsedilmiştir. EKK ile ilgili Legendre’nin 1805 yılında yayını bulunurken, Gauss’un 1809 yılında yayını olduğu ifade edilmiştir. Fakat Gauss’un EKK yöntemini ilk olarak 1795 yılında kullandığına dair ciddi kanıtların olduğu ve EKK’nın mucidinin Gauss olduğu kanısına varılmıştır. Stigler (1981) çalışmasında ayrıca EKK yöntemi modern istatistiğin otomobili olarak tanımlanmış ve EKK yöntemini ilk bulan kişiyi otomobilin mucidi Henry Ford ile özdeşleştirilmiştir.

Doğrusal regresyon modeli X matrisinin tam ranklı olması varsayımı altında “3.3.1. Çoklu Doğrusal Regresyon” bölümünde irdelenmiştir. Regresyon parametre vektörü için “3.3.1. Çoklu Doğrusal Regresyon” bölümünde verilen varsayımlar altında EKK tahmin edicisi şu şekilde edilir:

$$\hat{\beta} = \min_{\beta} \|Y - X\beta\|^2. \quad (3.19)$$

$$\hat{\beta} = (X'X)^{-1} X'Y. \quad (3.20)$$

Literatürde doğrusal regresyonda olduğu gibi hedonik regresyon modelinde de parametre tahminleri genellikle EKK yöntemi ile yapılmaktadır. EKK her ne kadar en yaygın istatistiksel yöntemlerden biri olsa da veride aykırı değerlerin olduğu durumlarda etkin sonuçlar vermemektedir. Veride aykırı değerlerin olması durumunda robust istatistiksel yöntemler EKK yöntemine göre daha etkin sonuçlar vermektedir.

EKK yöntemi kullanılarak tahmin edilen hedonik regresyon modelleri aykırı değerlere karşı oldukça hassastır çünkü EKK karesel artıkların toplamını en aza indirir. Bu durum tahminlerin aykırı değerlerden ve özellikle büyük aykırı değerlerden orantısız şekilde etkileneceği anlamına gelir. Az sayıda aykırı değer olduğu durumda bile büyük bir etkiye sahip olabilir. EKK'nın aksine robust yöntemler artıkların boyutuna göre otomatik olarak ağırlığını azaltır ve bu nedenle daha verimli tahminler üretir (Bourassa vd. 2013).

Bu tez çalışmasında hedonik regresyon modelindeki parametre tahminleri hem EKK hem de robust yöntemler ile yapılmış ve etkinlikleri birbirleri ile karşılaştırılmıştır. Bir sonraki bölümde robust yöntemler tanıtılmıştır.

3.5.2 Robust (Dayanıklı) istatistiksel yöntemler

Robust terimi istatistik literatürüne Box (1953) tarafından sunulmuştur. Modern robust istatistik Tukey (1960), Huber (1964) ve Hampel (1968)'in öncü çalışmalarıyla ortaya çıkmış ve zamanla yoğun bir şekilde geliştirilmiştir.

Veri setinde aykırı değerler olması durumunda robust yöntemlerin kullanılması önerilmektedir. Bir önceki bölümde de bahsedildiği üzere aykırı değerlerin etkisini azaltan robust istatistiksel yöntemler tahminlerin daha güvenilir olmasını sağlar.

Robust yöntemler verilerin çoğunluğunun davranışını modellemeyi amaçladığından, farklı piyasa koşullarına farklı tepkiler verirler. Piyasa koşullarındaki ani indirimler, icradan satışlar gibi fiyatı ciddi derecede düşüren sorunlu satış fiyatları gerçekleşebilmektedir. Bu işlemleri tespit etmek hem zaman hem maliyet gerektirir. Bazı durumlarda ise bu işlemleri tespit etmek imkansız olabilmektedir. Robust yöntemler bu tür sorunların üstesinden gelmek için etkili bir araçtır. Genel olarak, robust tahmin ediciler, aykırı değerlere daha az ağırlık verecek şekilde artıkların bazı işlevlerini en aza indirir (Bourassa vd. 2013).

Bu tez çalışmasında robust yöntemlerden sadece Huber'in M tahmin edicisi ve MM tahmin edicisine odaklanılmış ve elde edilen parametre tahminleri EKK yöntemi ile karşılaştırılmıştır.

3.5.2.1 Huber M tahmin edicisi

Robust regresyon için yaygın olarak kullanılan M tahmin yöntemi ilk robust yöntem olmak ile birlikte Huber (1964, 1973) tarafından önerilmiştir. Bu tahmin edici sınıfı, maksimum olabirlik tahmin edicisinin bir genellemesi olarak kabul edilmiştir. M tahmin edici yöntemi, EKK amaç fonksiyonundan daha az hızla artan bir amaç fonksiyonunu minimize etmek için tasarlanmıştır. M tahmin edicileri daha büyük standartlaştırılmış artıklara sahip gözlemlere azalan ağırlıklar verir (Maronna vd 2006).

Huber M tahmin edicisi aşağıda yer alan amaç fonksiyonunun β 'ya göre minimum edilmesiyle elde edilebilir:

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^n \rho \left(\frac{Y_i - X_i^T \beta}{\sigma} \right). \quad (3.21)$$

Burada $\rho(\cdot)$ fonksiyonu negatif olmayan, azalmayan bir fonksiyon olmak ile birlikte $\rho(0) = 0$ 'dır. M tahmin edicileri yukarıdaki fonksiyonu minimize etmeyi amaçladığı

için eğer ρ türevlenebilir ise β 'ya göre 1. türev alınıp 0'a eşitlenerek aşağıdaki eşitlik elde edilir:

$$\frac{1}{n} \sum_{i=1}^n \psi \left(\frac{Y_i - X_i^T \beta}{\sigma} \right) X_i = 0, \quad \psi(\cdot) = \rho'(\cdot). \quad (3.22)$$

İteratif yöntemler ile çözülen bu fonksiyon ile $\hat{\beta}$ elde edilir. Literatürde çeşitli $\rho(\cdot)$ fonksiyonları yer almaktadır. Bu çalışmada yaygın olarak kullanılan $\rho(\cdot)$ fonksiyonlarından Huber ve Tukey'in $\rho(\cdot)$ fonksiyonları kullanılmıştır. Bu fonksiyonlar aşağıda yer almaktadır.

Huber'in $\rho(\cdot)$ fonksiyonu aşağıda yer almaktadır:

$$\rho(r) = \begin{cases} r^2, & |r| \leq k \\ 2k|r| - k, & |r| > k \end{cases}. \quad (3.23)$$

Tukey'in $\rho(\cdot)$ fonksiyonu aşağıda yer almaktadır:

$$\rho(r) = \begin{cases} \frac{c^2}{6} (1 - (1 - r^2)^3), & |r| \leq c \\ \frac{c^2}{6}, & |r| > c \end{cases}. \quad (3.24)$$

Burada k ve c robustluğu ayarlama sabitleri olarak adlandırılır. %95'lik etkinlik için k=1.345 ve c=4.685 olmalıdır. Ayrıca r ise aşağıdaki eşitlikteki gibidir:

$$r = \left(\frac{Y_i - X_i^T \beta}{\sigma} \right). \quad (3.25)$$

Etki fonksiyonu, bir tahmin edicinin sağlamlığını ölçmenin yoludur ve robust yöntemlerde sınırlandırılmış bir etki fonksiyonuna sahip olmak beklenmektedir. M tahmin yönteminde, ψ fonksiyonu etki fonksiyonunun şeklini belirler. Bu nedenle, ψ artmayan bir fonksiyon ise M tahmin edicisi sınırlandırılmış etki fonksiyonuna sahip olacaktır. M tahmin yönteminde sınırlandırılmış ψ fonksiyonu yalnızca büyük artıkları kontrol eder; yani tahmin edici, Y yönündeki aykırı değerlere karşı dayanıklı olur. X yönündeki aykırı değerleri kontrol etmek için, genelleştirilmiş M tahmin veya MM tahmin yöntemleri gibi diğer robust tahmin yöntemlerinin kullanılması önerilmektedir (Moranna vd. 2006).

3.5.2.2 MM tahmin edicisi

Yohai (1987) tarafından geliştirilen MM tahmin edicisi, M tahmin edicisinin özel bir formudur. MM-tahmini, normallik varsayımı altında yüksek kırılma noktasına sahip bir tahmin edicidir. MM tahmin edicisi iki adımlı M tahmin edicisi olarak da bilinir. Öncelikle σ için M tahmini, aşağıdaki denklem çözülerek hesaplanır:

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left(\frac{Y_i - X_i^T \hat{\beta}_s}{\hat{\sigma}_s} \right) = 0.5. \quad (3.26)$$

Bu aşama yüksek kırılma noktalı robust σ tahmini verir. σ tahmini kullanarak aşağıdaki eşitliğin β 'ya göre minimum yapılır:

$$Q_n(\beta) = \frac{1}{n} \sum_{i=1}^n \rho_1 \left(\frac{Y_i - X_i^T \hat{\beta}_s}{\hat{\sigma}_s} \right). \quad (3.27)$$

Burada $\rho(\cdot)$ fonksiyonu için Tukey'in ρ fonksiyonu kullanılmıştır ve $\rho_1(\cdot) \leq \rho_0(\cdot)$ 'dir. Eğer $\rho(\cdot)$ türevlenebilir bir fonksiyon ise $\rho(\cdot)$ türevi $\psi(\cdot)$ olmak üzere MM tahmin edicisi $\hat{\beta}_{MM}$ aşağıdaki eşitliğin çözümü ile elde edilir:

$$\frac{1}{n} \sum_{i=1}^n \psi_1 \left(\frac{Y_i - X_i^T \hat{\beta}_{MM}}{\hat{\sigma}_s} \right) X_i = 0. \quad (3.28)$$

3.6 Değişken Seçim Yöntemleri

Hedonik regresyon analizindeki ana amaç ürünün fiyatını etkileyen özelliklerin fiyata olan katkısını bir nevi özelliklerin de fiyatını tahmin etmektir. Ürünlerin özelliklerinden bazıları fiyata doğrudan etki yaparken bazılarının fiyata çok etkisi sınırlı olmaktadır. Bu noktada özelliklerin fiyata olan katkısı önemli bir husustur. Ürün özelliklerinin fiyata olan katkısını ölçmek ve en kritik değişkenlerin seçilmesi için literatürde belirli yöntemler kullanılmıştır. Ridge, Lasso ve Elastiknet literatürde en yaygın kullanılan değişken seçim yöntemleridir.

Bu tez çalışmasında dizüstü bilgisayar fiyatını etkileyen ürün özelliklerinin önem sıralaması Ridge, Lasso ve Elastiknet yöntemleri ile gerçekleştirilmiş ve en önemli 3 özellik ön plana çıkarılmıştır. Burada Ridge, Lasso ve Elastiknet yöntemleri EKK yöntemindeki amaç fonksiyonuyla birleştirilerek uygulanmıştır. Ayrıca değişken seçimi göz önüne alınmaksızın EKK ve robust yöntemler için de en önemli 3 özellik tespit edilmiştir. Fakat bu yöntemler çerçevesinde tespit edilen önemli değişkenler literatürde ranking(sıralama) olarak adlandırılan ve hedonik regresyon ile sık olarak kullanılan yöntem ile belirlenmiştir. Bu yöntemin temel amacı parametre tahminlerinin modele katkısı baz alınarak önem düzeyini belirlemektir (Goren and Arslan 2023).

3.6.1 Ridge regresyon

Ridge regresyon, kareleri alınmış parametrelerin toplamının en aza indirilmesiyle elde edilen ve parametreleri ölçekten bağımsız hale getirmek için tüm değişkenler standardize edildiği bir yöntemdir. Ridge regresyon, ilk olarak Hoerl ve Kennard (1970) tarafından önerilmiştir.

Regresyon modelinde bağımsız değişkenler arasında çoklu bağlantının olduğu durumlarda ridge regresyon ile yansız sonuçlar elde edilebilmektedir (Myers 1986).

Ridge regresyon yaklaşımı, $(X'X)$ matrisinin köşegenine bir ridge parametresi ekleyerek yeni bir matris elde etmeye dayanır $(X'X + \lambda I)$. Korelasyon matrisindeki sistemlerin köşegeni bir ridge (sırt) olarak yorumlanabileceği için yöntemin adı ridge regresyonu olarak adlandırılmıştır (Hoerl ve Kennard 2000).

Ridge tahmin edicisi aşağıdaki kısıtlı minimizasyon problemi sonucunda elde edilir:

$$\min_{\beta} (Y - X\beta)^T (Y - X\beta), \quad \sum_{j=1}^p \beta_j^2 < k. \quad (3.29)$$

Bu kısıtlı minimizasyon problemi Lagrange yöntemi ile aşağıdaki gibi çözülür ve $\hat{\beta}_{ridge}$ elde edilir:

$$\min_{\beta} \left\{ (Y - X\beta)^T (Y - X\beta) + \lambda \sum_{j=1}^p \beta_j^2 \right\} \quad (3.30)$$

$$\hat{\beta}_{ridge} = (X^T X + \lambda I)^{-1} X^T Y. \quad (3.31)$$

3.6.2 Lasso regresyon

Lasso regresyon, Tibshirani (1996) tarafından önerilen bir değişken seçim yöntemidir. Bu seçim yönteminin ana avantajı, bazı katsayıları küçültmesi ve diğerlerini sıfıra büzmesidir. Bu nedenle, ridge regresyonunun ve diğer seçim yöntemlerinin iyi özelliklerini sürdürür (Basoğlu 2020). Ridge regresyonundan farklı olarak, belirli katsayıları sıfıra indirir (Herawati 2018). Ayrıca Lasso tahmin yöntemi, yüksek boyutlu doğrusal regresyon modelinde hem çoklu bağlantı sorununu hem de değişken seçimini birlikte ele alır (Altalbany 2021).

Lasso regresyon genellikle verimli ve hızlı algoritmaların gerekli olduğu ilaç sektörü gibi büyük veri tabanlarında kullanılır (Friedman vd. 2010).

Lasso regresyon aşağıdaki kısıtlı minimizasyon problemi sonucunda elde edilir:

$$\min_{\beta} (Y - X\beta)^T (Y - X\beta), \quad \sum_{j=1}^p |\beta_j| < k. \quad (3.32)$$

Bu kısıtlı minimizasyon problemi Lagrange yöntemi ile aşağıdaki gibi çözülür ve $\hat{\beta}_{lasso}$ elde edilir:

$$\min_{\beta} \left\{ (Y - X\beta)^T (Y - X\beta) + \lambda \sum_{j=1}^p |\beta_j| \right\}. \quad (3.33)$$

3.6.3 Elastiknet

Elastik net, Zou ve Hastie (2005) tarafından önerilen Lasso ve Ridge yöntemini tek bir amaç fonksiyonunda birleştiren bir yöntemdir. Bu yöntem parametre tahmini ve değişken seçimini aynı anda yapabilmektedir.

Elastik net aşağıdaki kısıtlı minimizasyon problemi sonucunda elde edilir:

$$\min_{\beta} (Y - X\beta)^T (Y - X\beta), \quad \sum_{j=1}^p \beta_j^2 < k_1, \quad \sum_{j=1}^p |\beta_j| < k_2. \quad (3.34)$$

Bu kısıtlı minimizasyon problemi Lagrange yöntemi ile aşağıdaki gibi çözülür ve $\hat{\beta}_{elastiknet}$ elde edilir:

$$\min_{\beta} \left\{ (Y - X\beta)^T (Y - X\beta) + \lambda_1 \sum_{j=1}^p \beta_j^2 + \lambda_2 \sum_{j=1}^p |\beta_j| \right\}. \quad (3.35)$$

Alttaki fonksiyonda $\alpha = 0$ olduğu durumda Ridge Regresyon, $\alpha = 1$ olduğu durumda ise Lasso regresyona karşılık gelmektedir. Ayrıca $\lambda = 0$ olduğu zaman EKK'ya eşit olmaktadır.

Ridge, Lasso ve Elastiknet tahmin edicilerinin ortak fonksiyonu aşağıdaki gibidir:

$$\hat{\beta}_{(\alpha)} = \min_{\beta} \left\{ (Y - X\beta)'(Y - X\beta) + \lambda \left[\alpha \sum_{j=1}^p |\beta_j| + (1-\alpha) \sum_{j=1}^p \beta_j^2 \right] \right\}. \quad (3.36)$$

$$\lambda = \lambda_1 + \lambda_2, \quad \alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}, \quad 1 - \alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2}. \quad (3.37)$$

$$\hat{\beta}_{(\alpha)} = \begin{cases} \hat{\beta}_{Ridge}, & \alpha = 0 \\ \hat{\beta}_{Elastiknet}, & 0 < \alpha < 1 \\ \hat{\beta}_{Lasso}, & \alpha = 1 \end{cases}. \quad (3.38)$$

4. ARAŞTIRMA BULGULARI

Bu tez çalışmasında dizüstü bilgisayar fiyatları hakkında etkin tahminleri veren modelin tespit edilmesi amacıyla Türkiye’de yer alan zincir elektronik marketlerde satışa sunulan dizüstü bilgisayar listesini içeren veri seti web kazıma yöntemiyle elde edilmiştir. Veri seti 2020 yılının son çeyreğine ait olmak ile birlikte yaklaşık 5000 satırdan oluşmaktadır. Veri deseninin bir kısmı aşağıdaki **şekil 4.1**’de gösterilmektedir.

	Fiyat	İşlemciHızı	Ram	Hdd	EkranBoyut	i3	i5	i7	ekrankart
2705	29490	2.6	32	1024	15.6	0	0	1	1
2706	29373	2.6	32	1024	15.6	0	0	1	1
2707	29296	2.6	32	1024	15.6	0	0	1	1
2708	28557	2.6	32	1024	15.6	0	0	1	1
2709	28557	2.6	32	1024	15.6	0	0	1	1
2710	14498	2.6	8	512	15.6	0	0	1	1
2711	14842	2.6	8	512	15.6	0	0	1	1
2712	13224	2.6	8	512	15.6	0	0	1	1
2713	13224	2.6	8	512	15.6	0	0	1	1
2714	13499	2.6	8	512	15.6	0	0	1	1

Şekil 4.1 Veri deseni

Web kazıma ile farklı teknoloji marketlerden elde edilen veriler birbirinden farklı veri deseni ile gelmiştir. Çünkü her market kendi sisteminde farklı şekillerde verileri tutmaktadır. Aynı teknoloji market içerisinde farklı marka bilgisayarlar için dahi farklı veri desenleri gözlenmiştir. Bu bağlamda gelen farklı veri desenli veri setinde detaylı bir işleme süreci gerçekleştirilmiş ve hedonik regresyon model alternatiflerini etkin bir şekilde uygulayabilmek için **şekil 4.1**’de yer alan veri deseni elde edilmiştir. Analizler R programı vasıtasıyla gerçekleştirilmiştir.

Veri setinde bir dizüstü bilgisayarın fiyatını doğrudan etkileyen işlemci, işlemci hızı, ram, harddisk, ekran kartı ve ekran boyutu özelliklerinin hedonik regresyon modelinde kullanılmasına karar verilmiştir. Burada kurulan hedonik regresyon modellerinde bağımlı değişken fiyat, bağımsız değişkenler ise işlemci, işlemci hızı, ram, harddisk, ekran boyutu ve ekran kartı olarak belirlenmiştir. Bu noktada fiyat, işlemci hızı, ram,

harddisk ve ekran boyutu deęişkenleri sayısal deęişkenler olarak tanımlanırken dięer deęişkenler kategorik (var yok (1,0)) kukla deęişken olarak tanımlanmıştır.

Analizlerde kullanılan model alternatiflerinden logaritmik formda olanlar için kukla deęişkenlerin logaritmasının anlamlı olmayacağı bilinmektedir. Bu sorunu ortadan kaldırmak amacıyla tez çalışmasında kullanılan tüm model alternatiflerinin üstel formda açık halleri tezin “3.4.1.Hedonik regresyon model alternatifleri” bölümünde ele alınmıştır.

Çalışmada kullanılan deęişkenlerin tanımı ve özellikleri aşağıdaki **çizelge 4.1**'de yer almaktadır.

Çizelge 4.1 Deęişken tanımları

Deęişken	Adı	Tipi	Türü	Ölçü Birimi
Y	Fiyat	Baęımlı	Sayısal	Türk Lirası
X_1	İşlemci hızı	Baęımsız	Sayısal	Gigahertz (GHz)
X_2	Ram	Baęımsız	Sayısal	Gigabyte (Gb)
X_3	Hard disk	Baęımsız	Sayısal	Gigabyte (Gb)
X_4	Ekran boyutu	Baęımsız	Sayısal	İnç
X_5	İşlemci- Intel i7	Baęımsız	Kukla	i7=1, dięer=0
X_6	İşlemci -Intel i5	Baęımsız	Kukla	i5=1, dięer =0
X_7	İşlemci -Intel i3	Baęımsız	Kukla	i3=1, dięer =0
X_8	Ekran kartı*	Baęımsız	Kukla	EXT=1, INT=0

*: Ekran kartının harici olması durumu EXT, dahili olması durumu INT olarak belirlenmiştir.

Hedonik regresyon modeli analizleri R programında gerçekleştirilmiştir. Regresyon analizlerinde Tam Logaritmik Model (LogLog), Logaritmik Doğrusal Model (LogDog), Tam Doğrusal Model (DogDog) ve Doğrusal Logaritmik Model (DogLog) olmak üzere temelde 4 model yapısı kullanılmıştır. Her bir model için parametreler hem robust (Huber-M ve MM tahmin edicileri) hem de EKK tahmin edilmiştir. Ayrıca her bir model için korelasyon analizi, deęişken seçim yöntemleri ve veride aykırı deęer olması durumu konuları da incelenmiştir.

4.1 Tam Logaritmik Model (LogLog)

Dizüstü bilgisayarın fiyatını etkileyen 8 faktörden oluşan tam logaritmik regresyon modeli kurulmuştur. Analizler için kurulan tam logaritmik hedonik regresyon modeli aşağıda verilmiştir:

$$\ln P_i = \alpha + \beta_1 \ln X_{i1} + \beta_2 \ln X_{i2} + \beta_3 \ln X_{i3} + \dots + \beta_8 \ln X_{i8} + \varepsilon_i, \quad i = 1, \dots, n. \quad (4.1)$$

Kurulan hedonik regresyon modelinden EKK, robust yöntemlerden Huber-M ve MM yöntemi ile elde edilen parametre tahminleri, standart hatalar ve elde edilen t değerleri **çizelge 4.2**'de yer almaktadır. Ayrıca elde edilen p değerleri de incelenmiş olup bulguları desteklediği görülmüştür.

Çizelge 4.2 Tahmin sonuçları (LogLog)

	EKK			Robust (Huber-M)			Robust (MM)		
	Tahmin	St Hata	t-değeri	Tahmin	St Hata	t-değeri	Tahmin	St Hata	t-değeri
α	9,94	0,13	78,24	9,94	0,12	84,44	10,16	0,16	64,27
X_1	0,14	0,01	16,01	0,14	0,01	17,47	0,15	0,01	11,44
X_2	0,49	0,01	62,13	0,48	0,01	65,00	0,49	0,01	47,78
X_3	0,11	0,01	15,10	0,11	0,01	16,03	0,10	0,01	11,49
X_4	-1,06	0,04	-24,27	-1,07	0,04	-26,01	-1,13	0,05	-21,46
X_5	0,51	0,01	57,40	0,50	0,01	60,11	0,50	0,01	81,92
X_6	0,24	0,01	25,92	0,23	0,01	26,39	0,23	0,01	26,70
X_7	0,12	0,01	11,64	0,09	0,01	96,14	0,10	0,01	68,38
X_8	0,08	0,01	11,62	0,08	0,01	13,00	0,08	0,01	11,49

Tahmin yöntemlerinin etkinlikleri karşılaştırmak amacıyla her bir tahmin yönteminin model standart hatası ise **çizelge 4.3**'te yer almaktadır. **Çizelge 4.3**'e göre standart hatalar birbirine oldukça yakın olmak ile birlikte en düşük standart hata değeri robust MM tahmin yöntemine aittir.

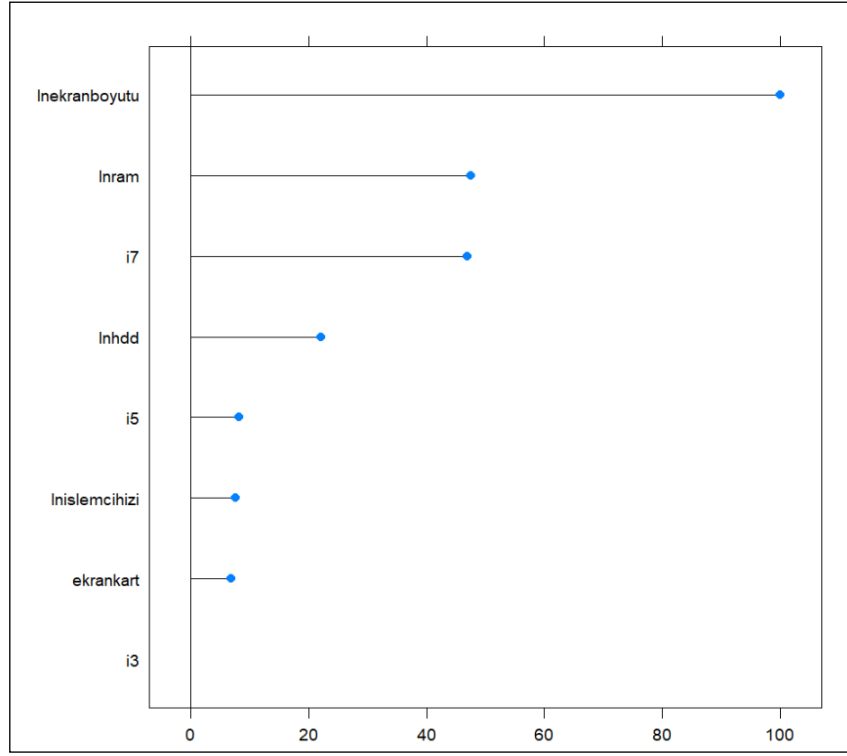
Çizelge 4.3 Tahmin yöntemlerinin model standart hataları (LogLog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası	0,13	0,11	0,10

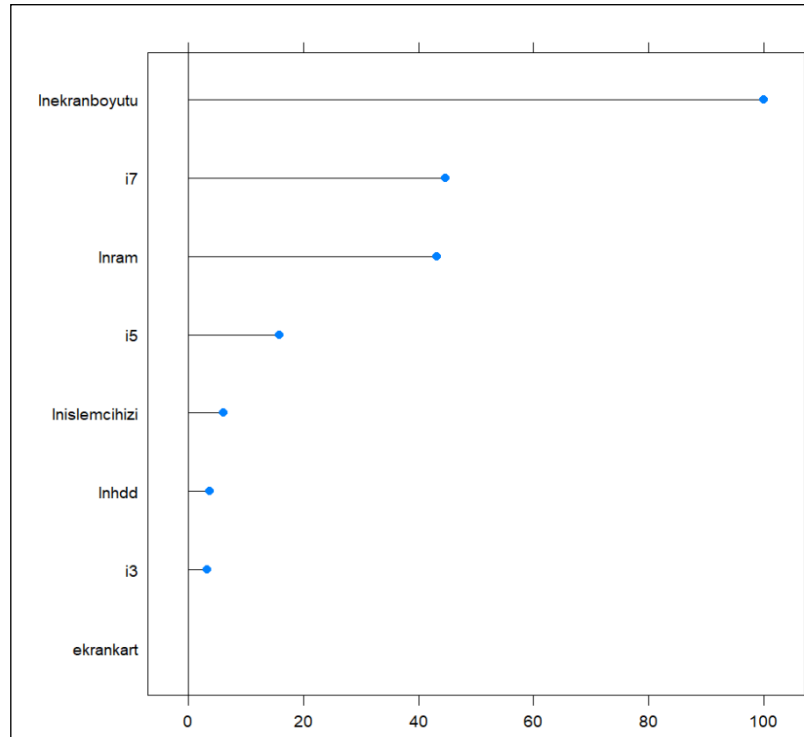
4.1.1 Değişkenlerin önem düzeyi

Dizüstü bilgisayar fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla değişken seçim yöntemlerinden olan Ridge, Lasso ve Elastiknet yöntemlerinden faydalanılmıştır. Her üç yöntem için yapılan analizler sonucunda fiyat değişkenini en çok etkileyen 3 değişken ön plana çıkarılmıştır. Buna ek olarak; EKK ve robust yöntemler için de en önemli 3 özellik tespit edilmiştir. Fakat bu yöntemler çerçevesinde tespit edilen önemli değişkenlerin yapılan parametre tahminlerinde modele katkı sıralaması baz alınarak belirlenmiştir.

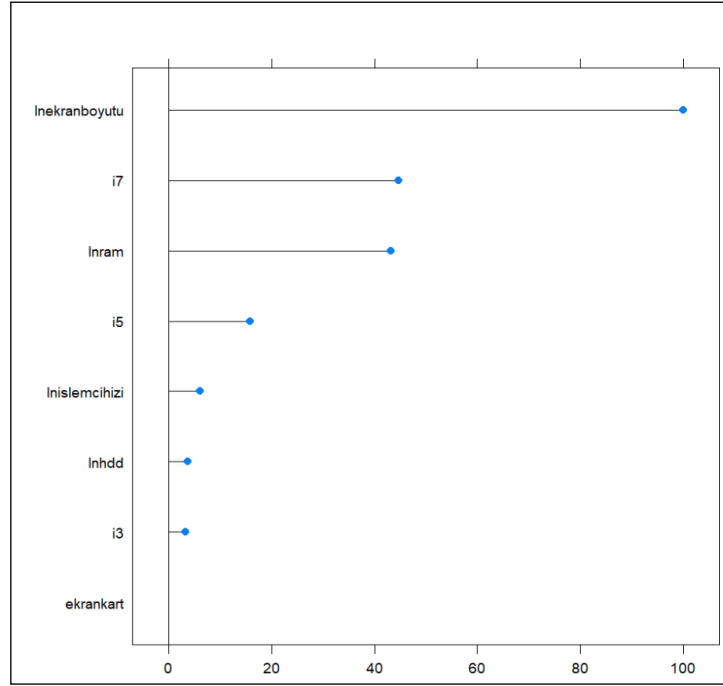
Ridge, Lasso ve Elastik net yöntemleri ile elde edilen sonuçlar **şekil 4.2-4.4**'te görülmektedir. Sonuçlara göre X_4 (ekran boyutu), X_2 (Ram) ve X_5 (İşlemci- i7) değişkenlerinin önem düzeyi bakımından ön plana çıktığı görülmektedir.



Şekil 4.2 Ridge önemli değişkenler (LogLog)

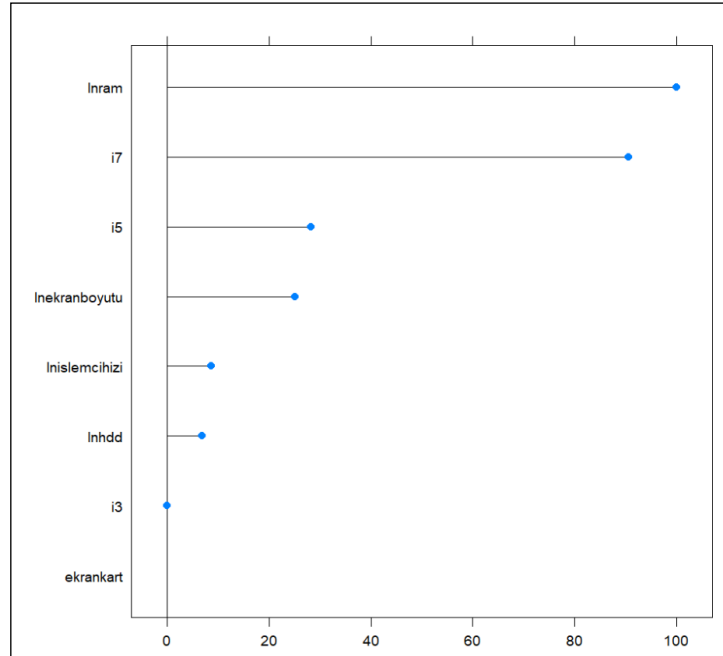


Şekil 4.3 Lasso önemli değişkenler (LogLog)

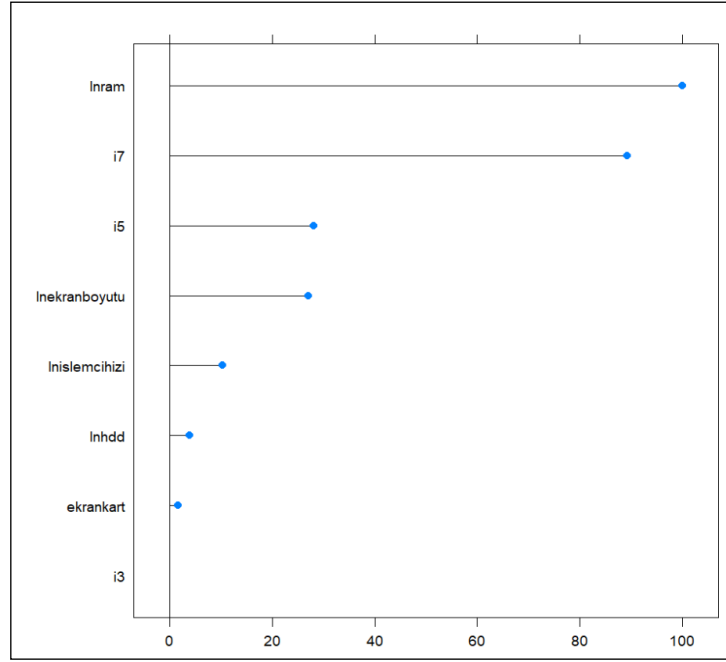


Şekil 4.4 Elastiknet önemli değişkenler (LogLog)

EKK ve robust yöntemler kapsamında X_2 (Ram), X_5 (İşlemci- i7) ve X_6 (İşlemci- i5) değişkenlerinin önem düzeyi bakımından ön plana çıktığı görülmektedir (Şekil 4.5-4.6).



Şekil 4.5 EKK önemli değişkenler (LogLog)



Şekil 4.6 Robust önemli değişkenler (LogLog)

Tam logaritmik olarak kurulan hedonik regresyon modelinde yer alan bağımsız değişkenlerin önem düzeyi birden fazla yöntem ile incelenmiş olup en önemli 3 değişken uygulanan yöntem bazında **çizelge 4.4**'te görülmektedir. **Çizelge 4.4**'te ayrıca kullanılan yöntemlerin performansını ölçen Ortalama Mutlak Hata (Mean Absolute Error-MAE), Ortalama Karekök Sapması (Root Mean Square Error-RMSE) ve R^2 gibi gösterge sonuçları da görülmektedir. İlgili göstergelerin elde edilmiş fonksiyonları aşağıdaki gibidir:

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - \hat{P}|. \quad (4.2)$$

$$RMSE = \frac{1}{n} \sqrt{\sum_{i=1}^n (P_i - \hat{P})^2}. \quad (4.3)$$

$$R^2 = 1 - \frac{\sum (P_i - \hat{P})^2}{\sum (P_i - \bar{P})^2}. \quad (4.4)$$

Analiz sonuçlarına göre kullanılan yöntemlerin hepsinin hemen hemen aynı performansta olduğu gözlenmiştir. En önemli değişkenler bütünsel olarak

incelendiğinde ise X_4 (ekran boyutu), X_2 (Ram), X_5 (İşlemci- i7) ve X_6 (İşlemci- i5) değişkenlerinin diğer bağımsız değişkenlere nazaran daha önemli olduğu görülmüştür (Çizelge 4.4).

Çizelge 4.4 Değişken önem düzeyi sonuçları (LogLog)

	En Önemli 3 bağımsız değişken	MAE	RMSE	R²
EKK	(X_2, X_5, X_6)	0,10	0,13	0,94
Robust	(X_2, X_5, X_6)	0,10	0,13	0,94
Ridge	(X_4, X_2, X_5)	0,10	0,14	0,94
Lasso	(X_4, X_5, X_2)	0,10	0,13	0,94
Elastik net	(X_4, X_5, X_2)	0,10	0,13	0,94

4.1.2 Korelasyon analizi

8 faktörün bir başka deyişle bağımsız değişkenlerin ilişkili olup olmadığı önemli bir konudur. Bağımsız değişkenlerin arasında önemli düzeyde ilişki olması çoklu bağlantı sorununu ortaya çıkarır ve regresyon modellerinde istenmeyen bir durumdur. Bu kapsamda tam logaritmik olarak kurmuş olduğumuz hedonik regresyon modelinde bağımsız değişkenler arasındaki ilişki düzeyi ölçülmüştür. Öncelikle değişkenler arasındaki potansiyel çoklu bağlantıyı tespit etmek amacıyla değişkenlerin Varyans Şişirme Faktörü (VIF) ölçülmüştür. Literatür incelendiğinde genel olarak; VIF değerinin 10'dan büyük veya eşit olması çoklu bağlantı olduğunu göstermektedir. Buna karşın bazı kaynaklarda bu değer 5'ten hatta 2,5'ten büyük olması durumunda da çoklu bağlantı şüphesinin olabileceğine rastlanmıştır. Bu bağlamda bu tez çalışmasında elde edilen VIF değerlerinden en yüksek olanları çoklu bağlantı şüphesi ile incelenmiştir.

Çizelge 4.5 Bağımsız değişkenlerin VIF değerleri (LogLog)

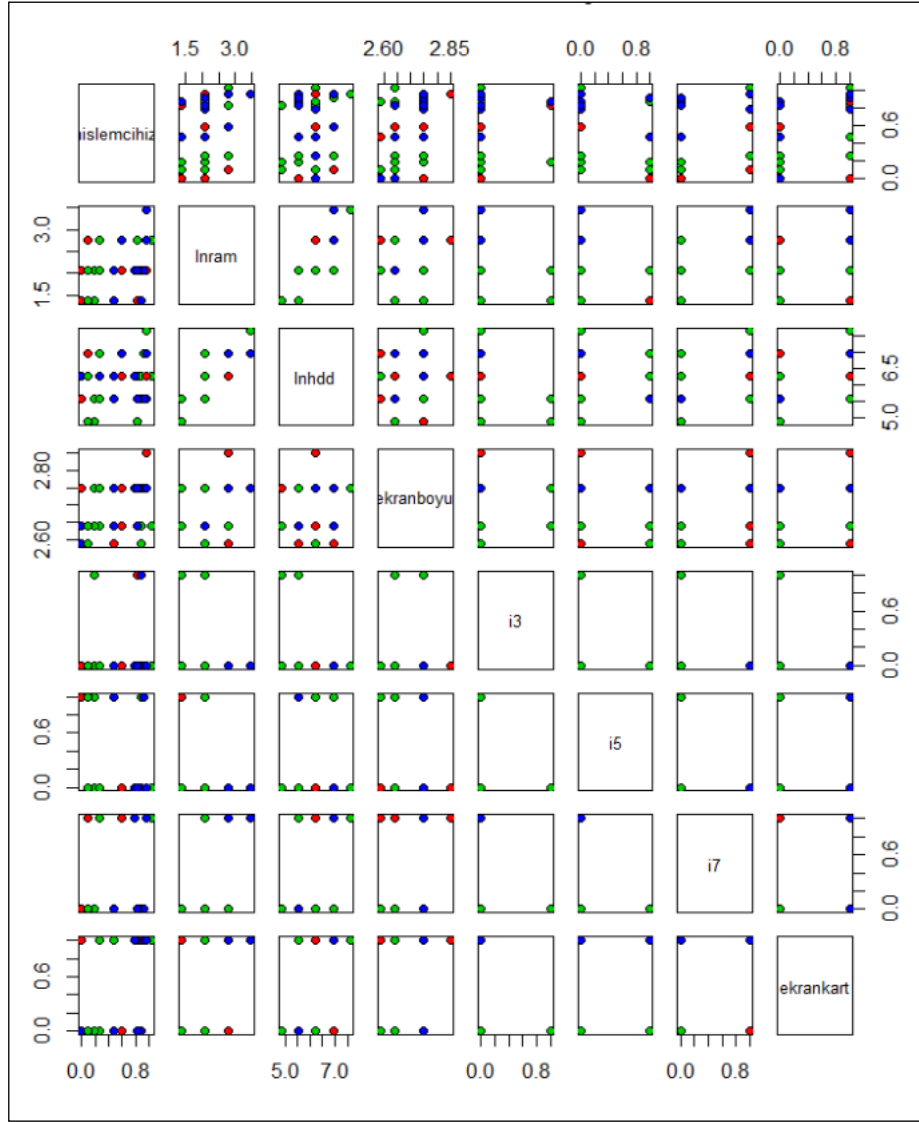
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
VIF	2,20	5,00	5,12	2,40	4,21	2,52	2,88	2,53

Çizelge 4.5'e göre X₂ (Ram) ve X₃ (Harddisk) değişkenlerinin VIF değerlerinin yüksek olduğu görülmüştür. Bu değerler iki değişken arasında önemli düzeyde bir ilişki olduğu şüphesini uyandırmıştır. Bu sebeple Pearson korelasyon katsayısı yöntemi ile değişkenlerin korelasyon matrisi oluşturulmuştur (Çizelge 4.6).

Çizelge 4.6 Bağımsız değişkenlere ait korelasyon matrisi (LogLog)

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
X ₁	1,00	0,29	0,27	0,60	0,35	-0,31	-0,15	0,59
X ₂	0,29	1,00	0,85	0,09	0,65	-0,28	-0,52	0,37
X ₃	0,27	0,85	1,00	-0,04	0,68	-0,26	-0,45	0,20
X ₄	0,60	0,09	-0,04	1,00	0,15	-0,25	0,11	0,59
X ₅	0,35	0,65	0,68	0,15	1,00	-0,60	-0,44	0,22
X ₆	-0,31	-0,28	-0,26	-0,25	-0,60	1,00	-0,18	0,00
X ₇	-0,15	-0,52	-0,45	0,11	-0,44	-0,18	1,00	-0,32
X ₈	0,59	0,37	0,20	0,59	0,22	0,00	-0,32	1,00

Şekil 4.7'de yer alan korelasyon grafiğindeki kutucuklarda yer alan noktalar iki değişken arasındaki ilişkiyi göstermektedir. Bu noktaların yoğun olduğu kutucuklar iki değişken arasında güçlü ilişki olduğunu göstermektedir.



Şekil 4.7 Korelasyon grafiği (LogLog)

Çizelge 4.6 ve **Şekil 4.7**'de görüleceği üzere X_2 (Ram) ve X_3 (Harddisk) değişkenleri arasındaki korelasyonun 0,85 yani güçlü ve pozitif yönlü olduğu görülmektedir. Bu nedenle 2 değişkenden birinin modelden çıkarılmasının devam eden analizlerin daha etkin yapılmasına olanak sağlayacaktır. Bu bağlamda **Çizelge 4.2**'de her iki değişkenin t değerleri incelendiğinde t değeri düşük olan X_3 (Harddisk) değişkeninin modelden çıkarılmasına karar verilmiştir.

X_3 (Harddisk) değişkeninin modelden çıkarılması ile birlikte elde edilen parametre tahminleri **çizelge 4.7**'de görülmektedir. Kurulan tam logaritmik hedonik regresyon

modelinde parametre tahminleri hem klasik EKK yöntemi hem de robust yöntemlerden Huber-M ve MM tahmin yöntemleri ile elde edilmiştir (**Çizelge 4.7**).

Çizelge 4.7 Korelasyon analizi sonrası tahmin sonuçları (LogLog)

	EKK			Robust (Huber-M)			Robust (MM)		
	Tahmin	St Hata	t-değeri	Tahmin	St Hata	t-değeri	Tahmin	St Hata	t-değeri
α	10,70	0,12	90,45	10,86	0,11	98,34	11,02	0,12	88,47
X_1	0,18	0,01	20,30	0,19	0,01	22,91	0,20	0,01	14,55
X_2	0,58	0,01	109,86	0,58	0,00	11,63	0,57	0,01	95,49
X_4	-1,22	0,04	-27,56	-12,67	0,04	-30,66	-1,32	0,05	-28,28
X_5	0,57	0,01	68,26	0,55	0,01	70,95	0,54	0,01	84,75
X_6	0,28	0,01	30,72	0,26	0,01	31,37	0,26	0,01	29,81
X_7	0,16	0,01	14,73	0,15	0,01	14,70	0,14	0,01	11,68
X_8	0,06	0,01	8,80	0,07	0,01	10,26	0,07	0,01	8,41

Çoklu bağlantı sorununa yol açan X_3 değişkeninin modelden çıkarılması ile elde edilen model bazında model standart hataları **çizelge 4.8**'de görülmektedir. Sonuçlar incelendiğinde X_3 değişkeninin modelden çıkarılmasının belirgin bir fark yaratmadığı sonucuna varılmıştır.

Çizelge 4.8 Korelasyon analizi sonrası model standart hataları (LogLog)

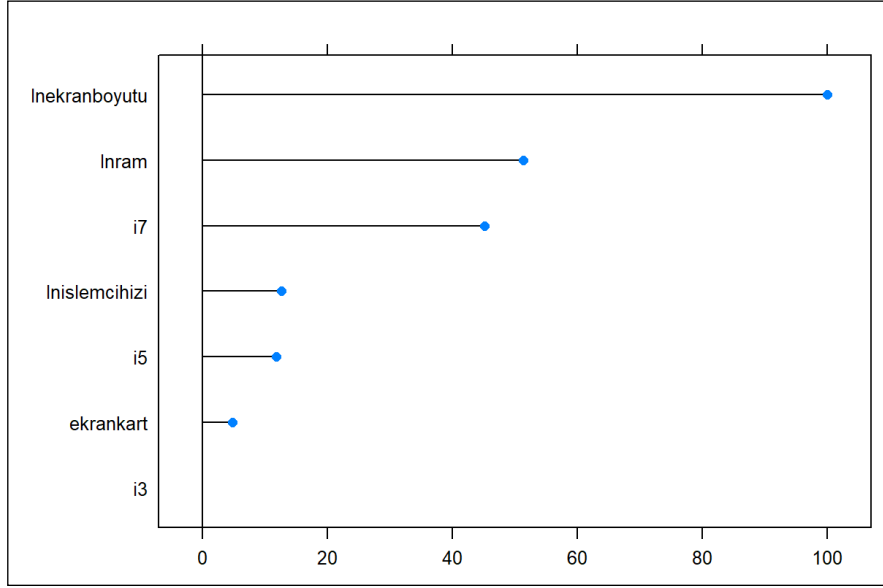
	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası-ilk	0,13	0,11	0,10
Model standart hatası-Korelasyon analizi sonrası	0,13	0,13	0,11

4.1.3 Korelasyon analizi sonrası değişkenlerin önem düzeyi

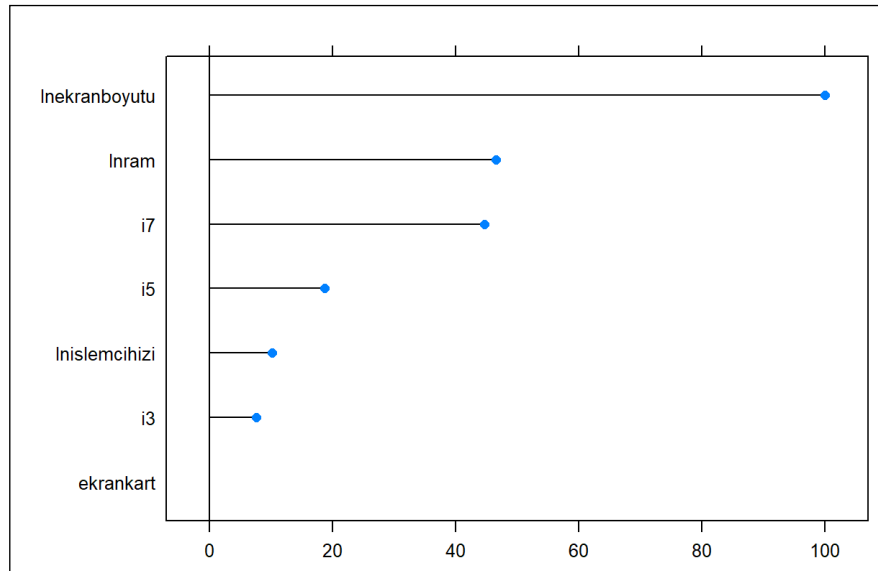
Korelasyon analizi sonrasında X_3 değişkenin çoklu bağlantı problemi nedeniyle model çıkarıldığı durumda da değişkenlerin önem düzeyi ölçülmüştür. Dizüstü bilgisayar

fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla Ridge, Lasso ve Elastiknet yöntemleri ile birlikte EKK ve robust yöntemden faydalanılmıştır.

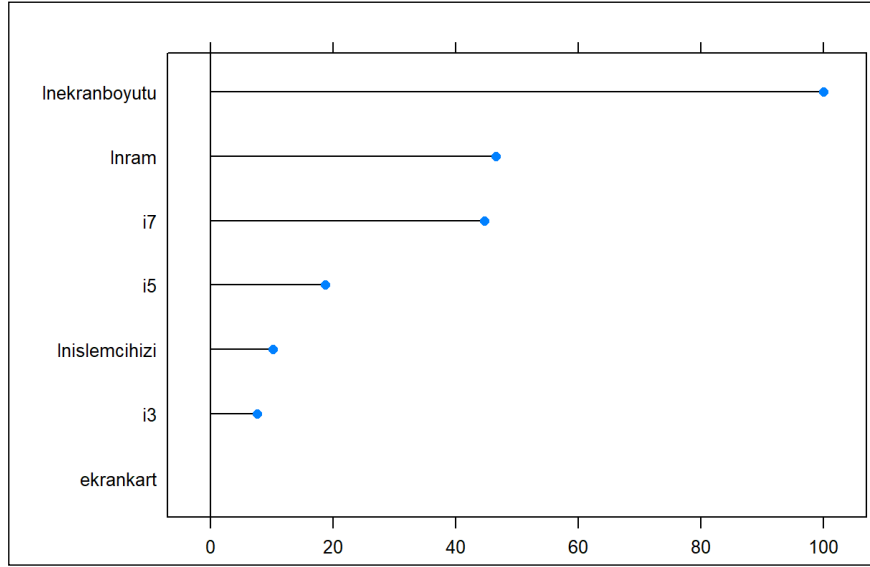
Ridge, Lasso ve Elastiknet yöntemleri ile elde edilen değişken önem düzeyi sonuçları **şekil 4.8-4.10**'da görülmektedir.



Şekil 4.8 Korelasyon analizi sonrası Ridge önemli değişkenler (LogLog)

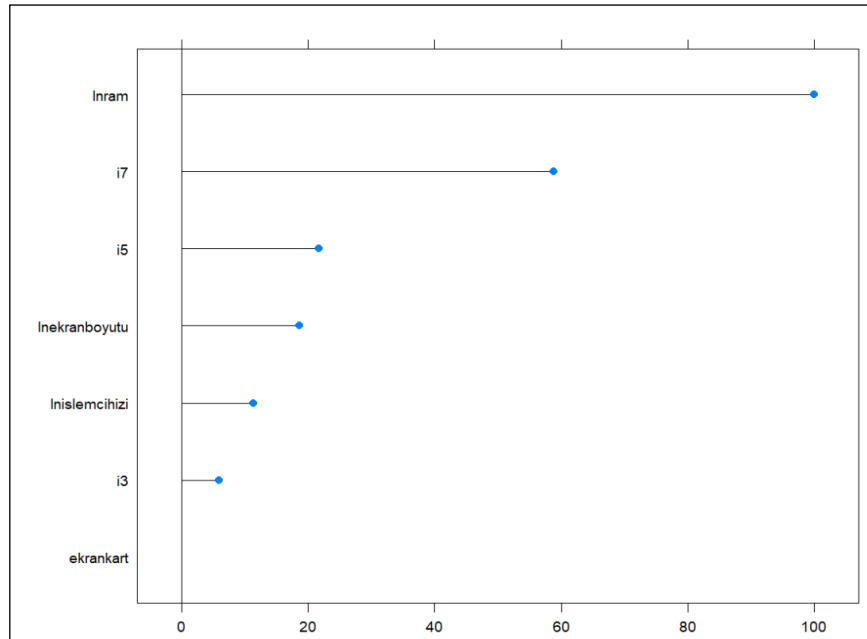


Şekil 4.9 Korelasyon analizi sonrası Lasso önemli değişkenler (LogLog)

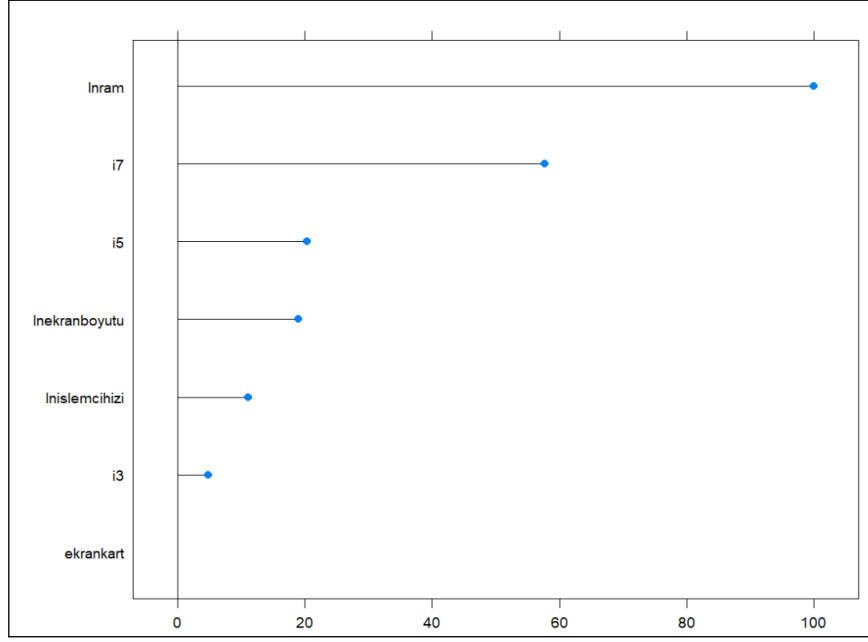


Şekil 4.10 Korelasyon analizi sonrası Elastiknet önemli değişkenler (LogLog)

EKK ve robust yöntemleri ile elde edilen değişken önem düzeyi sonuçları **şekil 4.11-4.12**'de görülmektedir.



Şekil 4.11 Korelasyon analizi sonrası EKK önemli değişkenler (LogLog)



Şekil 4.12 Korelasyon analizi sonrası Robust önemli değişkenler (LogLog)

X_3 değişkeninin çoklu bağlantı problemi nedeniyle modelden çıkarıldığı durumda tam logaritmik hedonik regresyon modeli ile elde edilen tahminler doğrultusunda Ridge, Lasso ve Elastik net yöntemlerine ek olarak klasik EKK ve robust yöntemler ile ölçülen değişkenlerin önem düzeyi ile ilgili hesaplamalar **çizelge 4.9'**da verilmiştir. **Çizelge 4.9'**da ayrıca kullanılan yöntemlerin performansını ölçen MAE, RMSE ve R^2 gibi göstergeler de hesaplanmıştır. Analiz sonuçlarına göre kullanılan yöntemlerin hepsinin hemen hemen aynı performansta olduğu gözlenmiştir. En önemli değişkenler bütünsel olarak incelendiğinde ise X_4 (ekran boyutu), X_2 (Ram), X_5 (İşlemci- i7) ve X_6 (İşlemci- i5) değişkenlerinin diğer bağımsız değişkenlere nazaran daha önemli olduğu görülmüştür. **Çizelge 4.9'**da elde edilen sonuçların **çizelge 4.4'**te elde edilen sonuçlar ile oldukça yakın olduğu görülmektedir Bu durum X_3 değişkeninin modelden çıkarılmasının sonuçlara belirgin bir etkisinin olmadığını göstermektedir.

Çizelge 4.9 Korelasyon analizi sonrası değişken önem düzeyi sonuçları (LogLog)

	En Önemli 3 bağımsız değişken	MAE	RMSE	R²
EKK	(X ₂ , X ₅ , X ₆)	0,11	0,14	0,93
Robust	(X ₂ , X ₅ , X ₆)	0,10	0,13	0,94
Ridge	(X ₄ , X ₂ , X ₅)	0,11	0,14	0,93
Lasso	(X ₄ , X ₂ , X ₅)	0,10	0,13	0,94
Elastik net	(X ₄ , X ₂ , X ₅)	0,10	0,13	0,94

4.1.4 Veride aykırı değer olması durumu

Veri setlerinde veri girişi, sistem, ölçü birimi kaynaklı hatalar meydana gelebilmektedir. Bu hataların büyük veri setlerinde tespit edilmesi kimi zaman zor olmakla birlikte gözden kaçma ihtimali de bulunmaktadır. Bu durumu ele almak amacıyla elde edilen veri setine %5 ila %10 arasında hatalı veri girişi eklenmiştir. Hatalı veri girişlerinde sıfır hatası (örneğin: 4399 TL olan fiyatın 43990 TL olması veya 13999 TL olan fiyatın 1399 TL olması), özelliklerin farklı özelliğe yazılması (örneğin: 8 GB ram ve 512 GB harddiske sahip dizüstü bilgisayarın 512 GB ram ve 8 GB harddiske sahip olarak kaydedilmesi) gibi hatalı veriler eklenmiştir. **Çizelge 4.10**'da veri setine belli miktarda hatalı veri girişi olduğu durumda 3 tahmin yönteminden elde edilen model standart hataları görülmektedir. Ayrıca **çizelge 4.10**'da veri setine hatalı veri girişi yapılmadan önceki model standart hataları da karşılaştırma yapabilmek amacıyla "Model standart hatası-ilk" olarak yer almaktadır.

Çizelge 4.10 Veride aykırı değer olduğunda model standart hataları (LogLog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası-ilk	0,13	0,11	0,10
Model standart hatası-aykırı değer olduğu durum	0,72	0,12	0,10

Çizelge 4.10 incelendiğinde, veri setine hatalı veriler girildiğinde EKK yönteminin model standart hata değerinin oldukça yükseldiği görülmüştür. Robust tahmin edicilerden Huber-M ve MM yöntemlerinin ise model standart hatalarının çok fazla değişmediği gözlenmiştir. Bu noktada robust yöntemlerin daha etkin sonuç verdiği görülmekle birlikte yapılan tahminlerde robust yöntemlerin tercih edilmesi önerilmektedir.

4.2 Logaritmik Doğrusal Model (LogDog)

Dizüstü bilgisayarın fiyatını etkileyen 8 faktörden oluşan logaritmik doğrusal regresyon modeli kurulmuştur. Analizler için kurulan doğrusal logaritmik hedonik regresyon modeli aşağıda verilmiştir:

$$\ln P_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_8 X_{i8} + \varepsilon_i, \quad i = 1, \dots, n. \quad (4.5)$$

Kurulan hedonik regresyon modelinden En Küçük Kareler (EKK), robust yöntemlerden Huber-M ve MM yöntemi ile elde edilen parametre tahminleri, standart hatalar ve elde edilen t değerleri **çizelge 4.11**'de yer almaktadır. Ayrıca elde edilen p değerleri de incelenmiş olup bulguları desteklediği görülmüştür.

Çizelge 4.11 Tahmin sonuçları (LogDog)

	EKK			Robust (Huber-M)			Robust (MM)		
	Tahmin	St Hata	t- değeri	Tahmin	St Hata	t- değeri	Tahmin	St Hata	t- değeri
α	9,44	0,04	213,34	9,58	0,04	230,60	9,38	0,16	220,27
X_1	0,08	0,01	13,46	0,08	0,01	15,80	0,09	0,01	11,44
X_2	0,04	0,00	66,67	0,04	0,00	69,69	0,04	0,01	47,78
X_3	0,00	0,00	-2,15	0,00	0,00	-2,01	0,00	0,01	11,49
X_4	-0,08	0,00	-26,45	-0,09	0,00	-30,36	-0,08	0,05	-21,46
X_5	0,68	0,01	76,24	0,64	0,01	76,88	0,53	0,01	68,38
X_6	0,31	0,01	31,83	0,26	0,01	28,65	0,20	0,01	26,70
X_7	0,01	0,01	0,70	-0,03	0,01	-3,11	-0,09	0,01	0,92
X_8	0,06	0,01	8,20	0,07	0,01	9,88	0,09	0,01	11,49

Tahmin yöntemlerinin etkinliklerini karşılaştırmak amacıyla her bir tahmin yönteminin model standart hataları ise **çizelge 4.12**'de görülmektedir. **Çizelge 4.12**'e göre standart hatalar birbirine oldukça yakın olmak ile birlikte en düşük standart hata değeri MM tahmin yöntemine aittir.

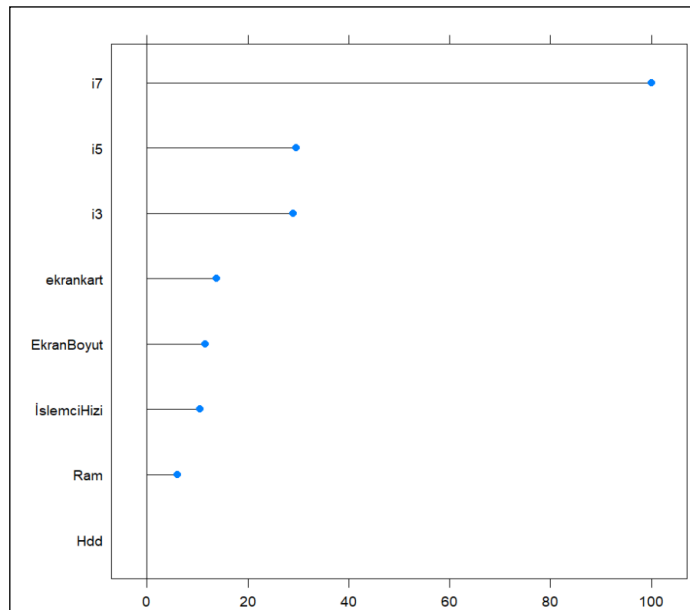
Çizelge 4.12 Tahmin yöntemlerinin model standart hataları (LogDog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası	0,14	0,12	0,11

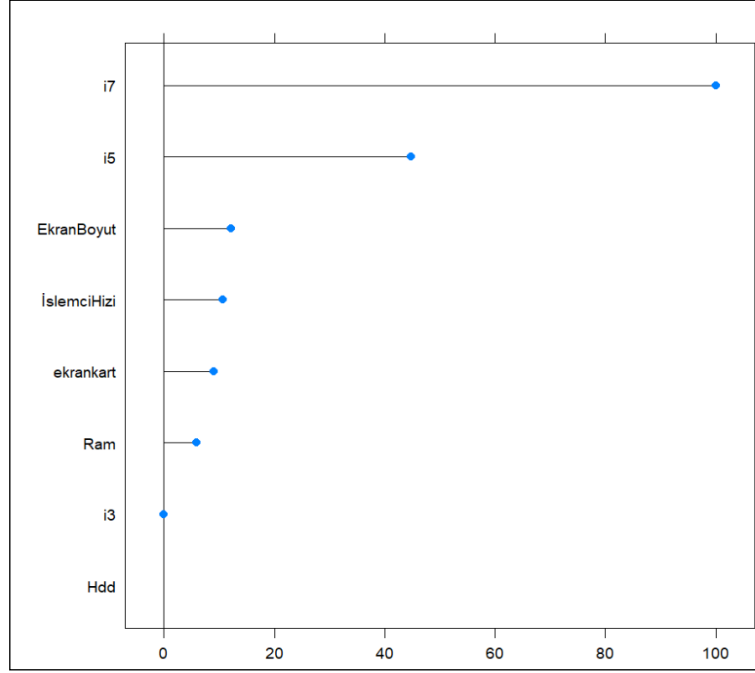
4.2.1 Değişkenlerin önem düzeyi

Dizüstü bilgisayar fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla Ridge, Lasso ve Elastiknet yöntemlerinden faydalanılmıştır. Her üç yöntem için yapılan analizler sonucunda fiyat değişkenini en çok etkileyen 3 değişken ön plana çıkarılmıştır.

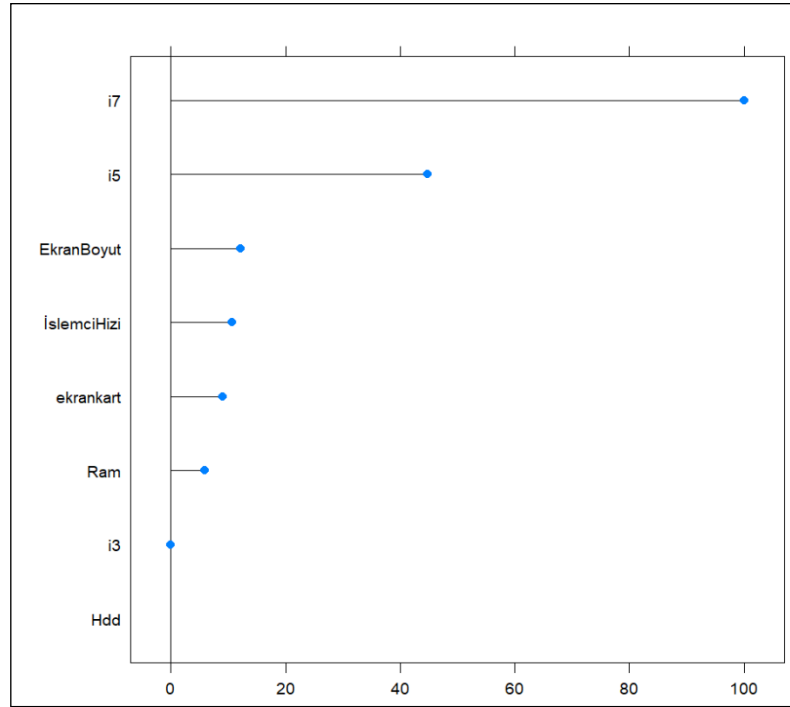
Ridge, Lasso ve Elastik net yöntemleri ile elde edilen sonuçlar **şekil 4.13-4.15**'te görülmektedir. Sonuçlara göre; X_5 (İşlemci- i7), X_6 (İşlemci- i5), X_7 (İşlemci- i3) ve X_4 (ekran boyutu) değişkenlerinin önem düzeyi bakımından ön plana çıktığı görülmektedir.



Şekil 4.13 Ridge önemli değişkenler (LogDog)

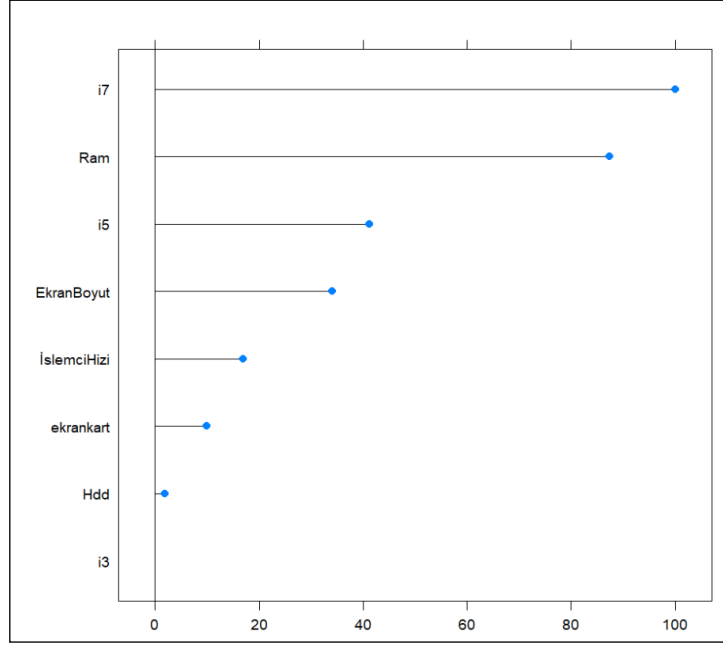


Şekil 4.14 Lasso önemli değişkenler (LogDog)

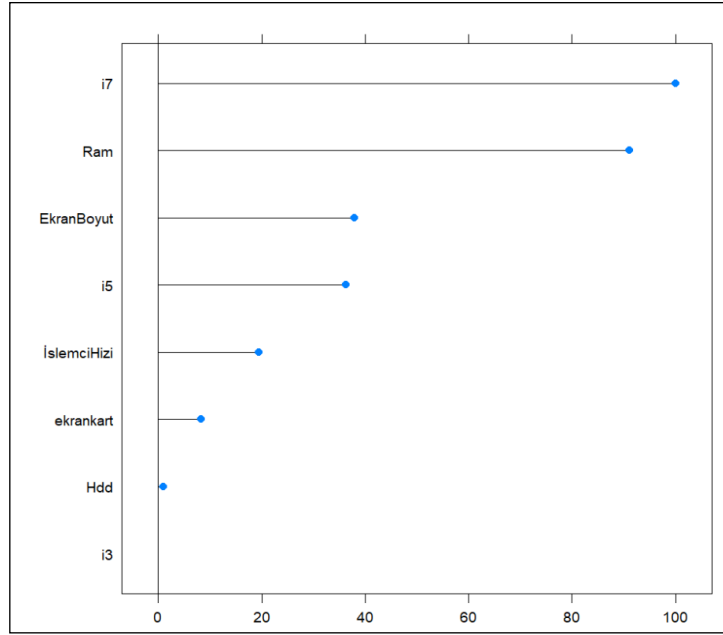


Şekil 4.15 Elastiknet önemli değişkenler (LogDog)

EKK ve robust yöntemler kapsamında X_5 (İşlemci- i7), X_6 (İşlemci- i5), X_2 (Ram) ve X_4 (ekran boyutu) değişkenlerinin önem düzeyi bakımından ön plana çıktığı görülmektedir (Şekil 4.16-4.17).



Şekil 4.16 EKK önemli değişkenler (LogDog)



Şekil 4.17 Robust önemli değişkenler (LogDog)

Logaritmik doğrusal olarak kurulan hedonik regresyon modelinde yer alan bağımsız değişkenlerin önem düzeyi birden fazla yöntem ile incelenmiş olup en önemli 3 değişken uygulanan yöntem bazında **çizelge 4.13**'te görülmektedir. **Çizelge 4.13**'te ayrıca kullanılan yöntemlerin performansını ölçen MAE, RMSE ve R^2 gibi göstergeler de hesaplanmıştır.

Analiz sonuçlarına göre kullanılan yöntemlerin hepsinin hemen hemen aynı performansta olduğu gözlenmiştir. En önemli değişkenler bütünsel olarak incelendiğinde ise X_5 (İşlemci- i7), X_6 (İşlemci- i5), X_7 (İşlemci- i3), X_4 (ekran boyutu) ve X_2 (Ram) değişkenlerinin diğer bağımsız değişkenlere nazaran daha önemli olduğu görülmüştür (**Çizelge 4.13**).

Çizelge 4.13 Değişken önem düzeyi sonuçları (LogDog)

	En Önemli 3 bağımsız değişken	MAE	RMSE	R²
EKK	(X_5, X_2, X_6)	0,11	0,14	0,93
Robust	(X_5, X_2, X_4)	0,11	0,14	0,93
Ridge	(X_5, X_6, X_7)	0,11	0,15	0,92
Lasso	(X_5, X_6, X_4)	0,11	0,14	0,93
Elastik net	(X_5, X_6, X_4)	0,11	0,14	0,93

4.2.2 Korelasyon analizi

Logaritmik doğrusal olarak kurmuş olduğumuz hedonik regresyon modelinde bağımsız değişkenler arasındaki ilişki düzeyi ölçülmüştür. Öncelikle değişkenler arasındaki potansiyel çoklu bağlantıyı tespit etmek amacıyla değişkenlerin Varyans Şişirme Faktörü (VIF) ölçülmüştür.

Çizelge 4.14 Bağımsız değişkenlerin VIF değerleri (LogDog)

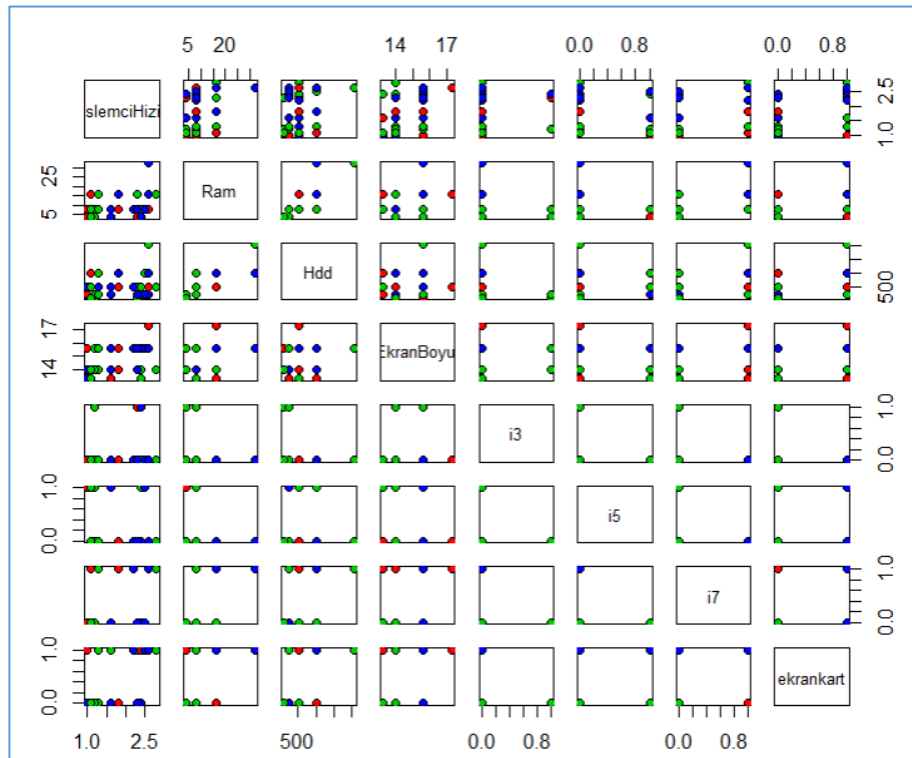
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
VIF	2,47	4,00	3,99	2,40	3,76	2,75	2,68	2,76

Çizelge 4.14'e göre X_2 (Ram) ve X_3 (Harddisk) değişkenlerinin VIF değerlerinin yüksek olduğu görülmüştür. Bu değerler iki değişken arasında önemli düzeyde bir ilişki olduğu şüphesini uyandırmıştır. Bu sebeple pearson korelasyon katsayısı yöntemi ile değişkenlerin korelasyon matrisi oluşturulmuştur (**Çizelge 4.15**).

Çizelge 4.15 Bağımsız değişkenlere ait korelasyon matrisi (LogDog)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1,00	0,33	0,22	0,62	0,35	-0,16	-0,30	0,64
X_2	0,33	1,00	0,82	0,14	0,55	-0,37	-0,30	0,36
X_3	0,22	0,82	1,00	-0,06	0,56	-0,33	-0,28	0,17
X_4	0,62	0,14	-0,06	1,00	0,16	0,11	-0,25	0,59
X_5	0,35	0,55	0,56	0,16	1,00	-0,44	-0,60	0,22
X_6	-0,16	-0,37	-0,33	0,11	-0,44	1,00	-0,18	-0,32
X_7	-0,30	-0,30	-0,28	-0,25	-0,60	-0,18	1,00	0,00
X_8	0,64	0,36	0,17	0,59	0,22	-0,32	0,00	1,00

Çizelge 4.15 ve Şekil 4.18’de görüleceği üzere X_2 (Ram) ve X_3 (Harddisk) değişkenleri arasındaki korelasyonun 0,82 yani güçlü ve pozitif yönlü olduğu görülmektedir. Bu nedenle 2 değişkenden birinin modelden çıkarılmasının devam eden analizlerin daha etkin yapılmasına olanak sağlayacaktır. Bu bağlamda Çizelge 4.11’de her iki değişkenin t değerleri incelendiğinde t değeri düşük olan X_3 (Harddisk) değişkeninin modelden çıkarılmasına karar verilmiştir.



Şekil 4.18 Korelasyon grafiği (LogDog)

X_3 (Harddisk) deęişkeninin modelden çıkarılması ile birlikte elde edilen parametre tahminleri **çizelge 4.16**'da yer almaktadır. Kurulan logaritmik doğrusal hedonik regresyon modelinde parametre tahminleri hem klasik EKK yöntemi hem de robust yöntemlerden Huber-M ve MM tahmin yöntemleri ile elde edilmiştir (**Çizelge 4.16**).

Çizelge 4.16 Korelasyon analizi sonrası tahmin sonuçları (LogDog)

	EKK			Robust (Huber-M)			Robust (MM)		
	Tahmin	St Hata	t- değeri	Tahmin	St Hata	t- değeri	Tahmin	St Hata	t- değeri
μ	9,42	0,04	219,22	9,55	0,04	235,16	9,81	0,06	150,89
X_1	0,07	0,01	13,29	0,08	0,01	15,68	0,07	0,01	7,99
X_2	0,04	0,00	101,60	0,04	0,00	106,31	0,04	0,00	50,92
X_4	-0,08	0,00	-26,85	-0,09	0,00	-30,70	-0,09	0,00	-49,30
X_5	0,68	0,01	79,00	0,64	0,01	79,14	0,49	0,04	13,10
X_6	0,31	0,01	31,85	0,26	0,01	28,66	0,09	0,05	1,76
X_7	0,00	0,01	0,38	-0,03	0,01	-3,31	-0,21	0,05	-4,49
X_8	0,06	0,01	8,41	0,07	0,01	9,93	0,07	0,01	11,52

Çoklu bağlantı sorununa yol açan X_3 deęişkeninin modelden çıkarılması ile elde edilen model bazında model standart hataları **çizelge 4.17**'de görülmektedir. Sonuçlar incelendiğinde X_3 deęişkeninin modelden çıkarılmasının belirgin bir fark yaratmadığı sonucuna varılmıştır.

Çizelge 4.17 Korelasyon analizi sonrası model standart hataları (LogDog)

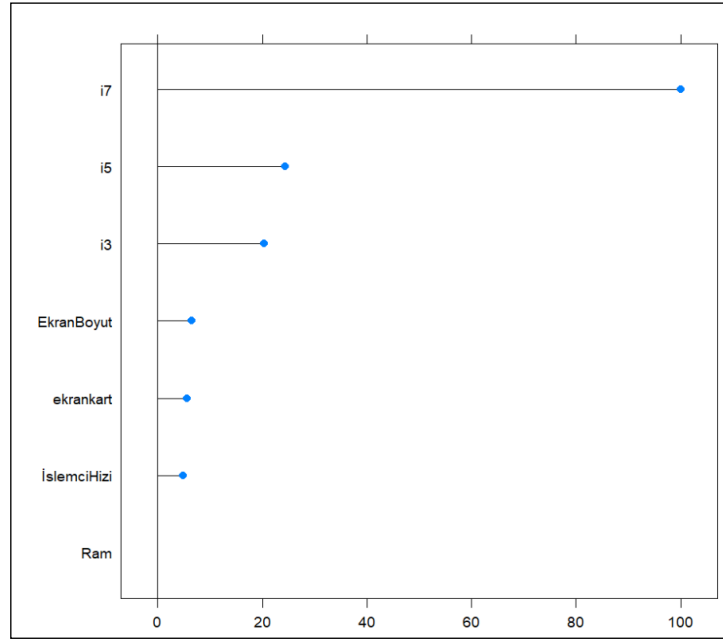
	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası-ilk	0,14	0,12	0,11
Model standart hatası-Korelasyon analizi sonrası	0,14	0,12	0,11

4.2.3 Korelasyon analizi sonrası deęişkenlerin önem düzeyi

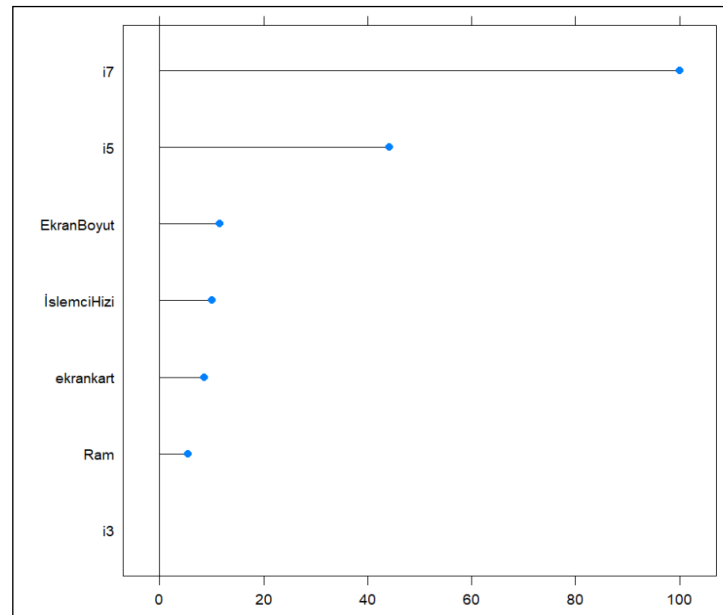
Korelasyon analizi sonrasında, X_3 deęişkeninin çoklu bağlantı problemi nedeniyle model çıkarıldığı durumda da deęişkenlerin önem düzeyi ölçülmüştür. Dizüstü bilgisayar

fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla Ridge, Lasso ve Elastiknet yöntemleri ile birlikte EKK ve robust yöntemlerden faydalanılmıştır.

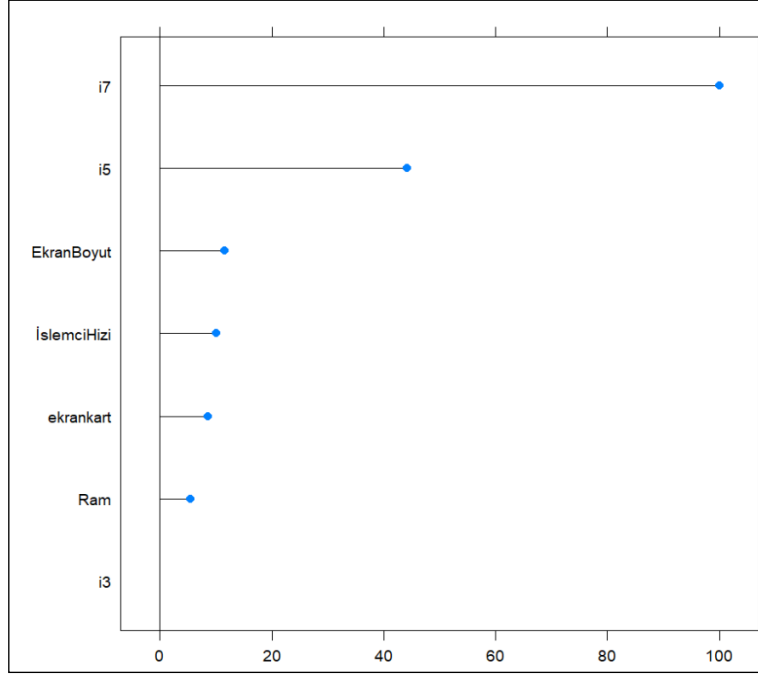
Ridge, Lasso ve Elastiknet yöntemleri ile elde edilen değişken önem düzeyi sonuçları **şekil 4.19-4.21**'de görülmektedir.



Şekil 4.19 Korelasyon analizi sonrası Ridge önemli değişkenler (LogDog)

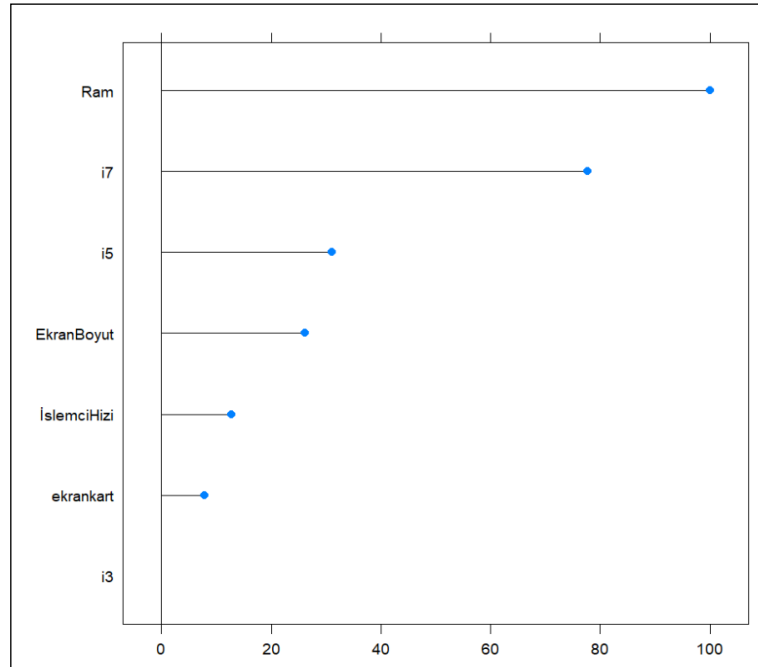


Şekil 4.20 Korelasyon analizi sonrası Lasso önemli değişkenler (LogDog)

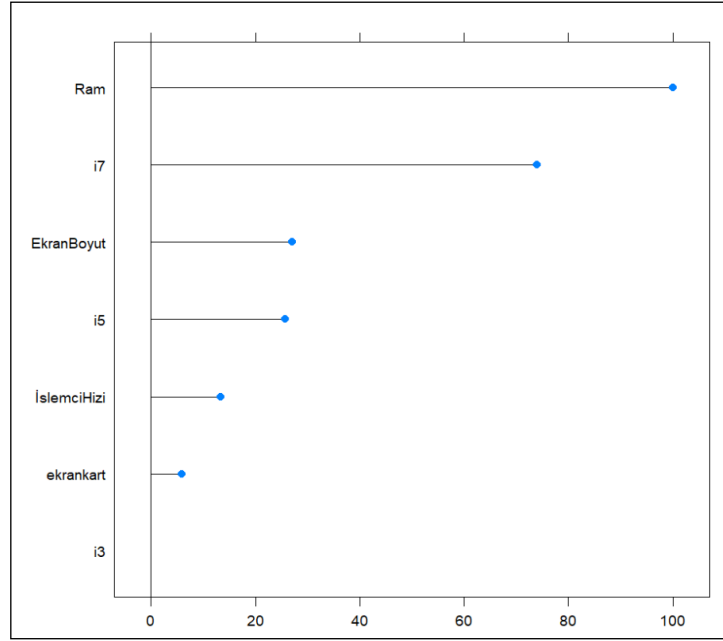


Şekil 4.21 Korelasyon analizi sonrası Elastiknet önemli değişkenler (LogDog)

EKK ve robust yöntemleri ile elde edilen değişken önem düzeyi sonuçları **şekil 4.22-4.23**'de görülmektedir.



Şekil 4.22 Korelasyon analizi sonrası EKK önemli değişkenler (LogDog)



Şekil 4.23 Korelasyon analizi sonrası EKK önemli değişkenler (LogDog)

X_3 değişkenin çoklu bağlantı problemi nedeniyle model çıkarıldığı durumda logaritmik doğrusal hedonik regresyon modeli ile elde edilen tahminler doğrultusunda Ridge, Lasso ve Elastik net yöntemlerine ek olarak klasik EKK ve robust yöntemler ile ölçülen değişkenlerin önem düzeyi ile ilgili hesaplamalar **çizelge 4.18**'de verilmiştir. **Çizelge 4.18**'de ayrıca kullanılan yöntemlerin performansını ölçen MAE, RMSE ve R^2 gibi göstergeler de hesaplanmıştır. Analiz sonuçlarına göre kullanılan yöntemlerin hepsinin hemen hemen aynı performansta olduğu gözlenmiştir. En önemli değişkenler bütünsel olarak incelendiğinde ise X_5 (İşlemci- i7), X_6 (İşlemci- i5), X_7 (İşlemci- i3), X_4 (ekran boyutu) ve X_2 (Ram) değişkenlerinin diğer bağımsız değişkenlere nazaran daha önemli olduğu görülmüştür. **Çizelge 4.18**'de elde edilen sonuçların **çizelge 4.13**'te elde edilen sonuçlar ile oldukça yakın olduğu görülmektedir Bu durum X_3 değişkenin modelden çıkarılmasının sonuçlara belirgin bir etkisinin olmadığını göstermektedir.

Çizelge 4.18 Korelasyon analizi sonrası değişken önem düzeyi sonuçları (LogDog)

	En Önemli 3 bağımsız değişken	MAE	RMSE	R²
EKK	(X_2, X_5, X_6)	0,11	0,14	0,93
Robust	(X_2, X_5, X_4)	0,11	0,14	0,93
Ridge	(X_5, X_6, X_7)	0,11	0,15	0,93
Lasso	(X_5, X_6, X_4)	0,11	0,14	0,93
Elastik net	(X_5, X_6, X_4)	0,11	0,14	0,93

4.2.4 Veride aykırı değer olması durumu

Çizelge 4.19’da veri setine belli miktarda hatalı veri girişi olduğu durumda 3 tahmin yönteminden elde edilen model standart hataları görülmektedir. Ayrıca çizelge 4.19’da veri setine hatalı veri girişi yapılmadan önceki model standart hataları da karşılaştırma yapabilmek amacıyla “Model standart hatası-ilk” olarak yer almaktadır.

Çizelge 4.19 Veride aykırı değer olduğunda model standart hataları (LogDog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası-ilk	0,14	0,12	0,11
Model standart hatası-aykırı değer olduğu durum	0,40	0,13	0,12

Çizelge 4.19 incelendiğinde, veri setine hatalı veriler girildiğinde EKK yönteminin model standart hata değerinin yükseldiği görülmüştür. Robust tahmin edicilerden Huber-M ve MM yöntemlerinin ise model standart hatalarının çok fazla değişmediği gözlenmiştir. Bu noktada robust yöntemlerin daha etkin sonuç verdiği görülmekle birlikte yapılan tahminlerde robust yöntemlerin tercih edilmesi önerilmektedir.

4.3 Tam Doğrusal Model (DogDog)

Dizüstü bilgisayarın fiyatını etkileyen 8 faktörden oluşan tam doğrusal regresyon modeli kurulmuştur. Analizler için kurulan tam doğrusal hedonik regresyon modeli aşağıda verilmiştir:

$$P_i = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_8 X_{i8} + \varepsilon_i, \quad i = 1, \dots, n. \quad (4.6)$$

Kurulan hedonik regresyon modelinden En Küçük Kareler (EKK), robust yöntemlerden Huber-M ve MM yöntemi ile elde edilen parametre tahminleri, standart hatalar ve elde edilen t değerleri **çizelge 20**'de yer almaktadır. Ayrıca elde edilen p değerleri de incelenmiş olup bulguları desteklediği görülmüştür.

Çizelge 4.20 Tahmin sonuçları (DogDog)

	EKK			Robust (Huber-M)			Robust (MM)		
	Tahmin	St Hata	t- değeri	Tahmin	St Hata	t- değeri	Tahmin	St Hata	t- değeri
μ	12516,38	538,39	23,25	11792,67	469,77	25,10	10728,69	281,60	38,10
X_1	1069,84	68,49	15,62	853,11	59,76	14,28	368,36	44,66	8,25
X_2	727,59	7,44	97,85	626,20	6,49	96,51	413,71	6,45	64,11
X_3	-1,30	0,14	-9,45	0,10	0,12	0,80	2,09	0,15	14,02
X_4	-955,02	39,07	-24,44	-847,61	34,09	-24,86	-624,81	20,27	-30,83
X_5	5669,04	108,64	52,18	5420,30	94,79	57,18	5210,64	56,72	91,87
X_6	2709,09	118,35	22,89	2394,44	103,26	23,19	1703,44	62,75	27,15
X_7	2219,96	136,66	16,24	1805,53	119,24	15,14	730,73	61,41	11,90
X_8	14,66	91,80	0,16	203,75	80,10	2,54	450,52	60,12	7,49

Tahmin yöntemlerinin etkinliklerini karşılaştırmak amacıyla her bir tahmin yönteminin model standart hataları ise **çizelge 21**'de görülmektedir. **Çizelge 21**'e en düşük standart hata değeri MM tahmin yöntemine aittir.

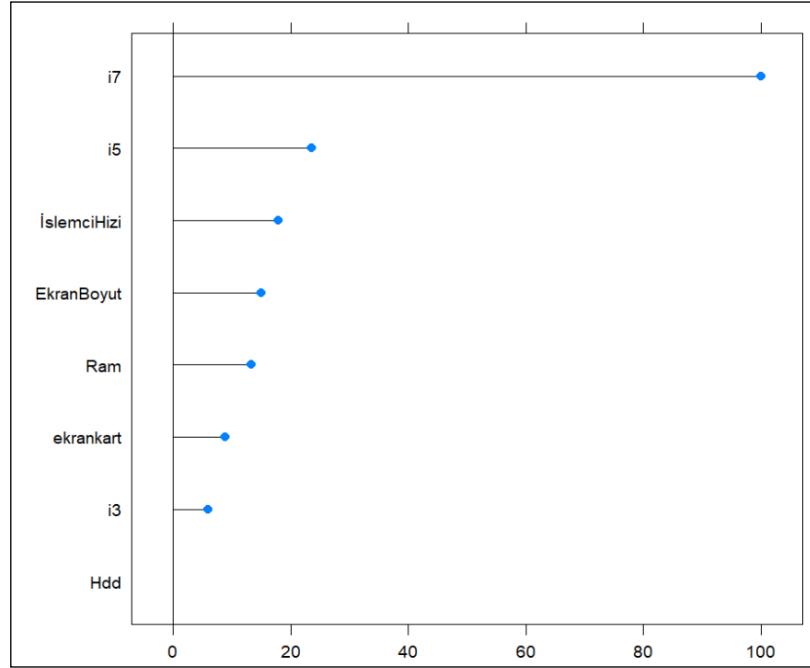
Çizelge 4.21 Tahmin yöntemlerinin model standart hataları (DogDog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası	1715	1304	1050

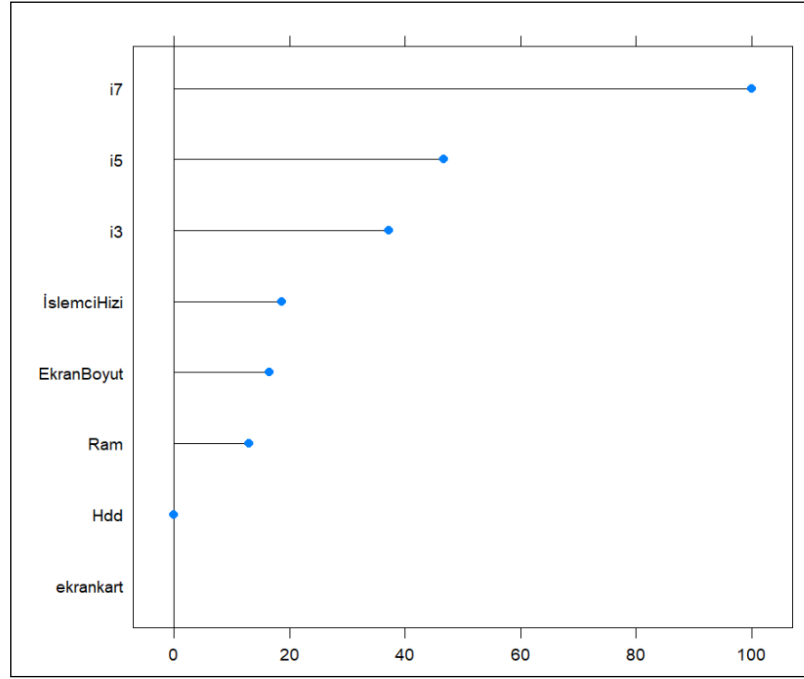
4.3.1 Değişkenlerin önem düzeyi

Dizüstü bilgisayar fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla Ridge, Lasso ve Elastiknet yöntemlerinden faydalanılmıştır. Her üç yöntem için yapılan analizler sonucunda fiyat değişkenini en çok etkileyen 3 değişken ön plana çıkarılmıştır.

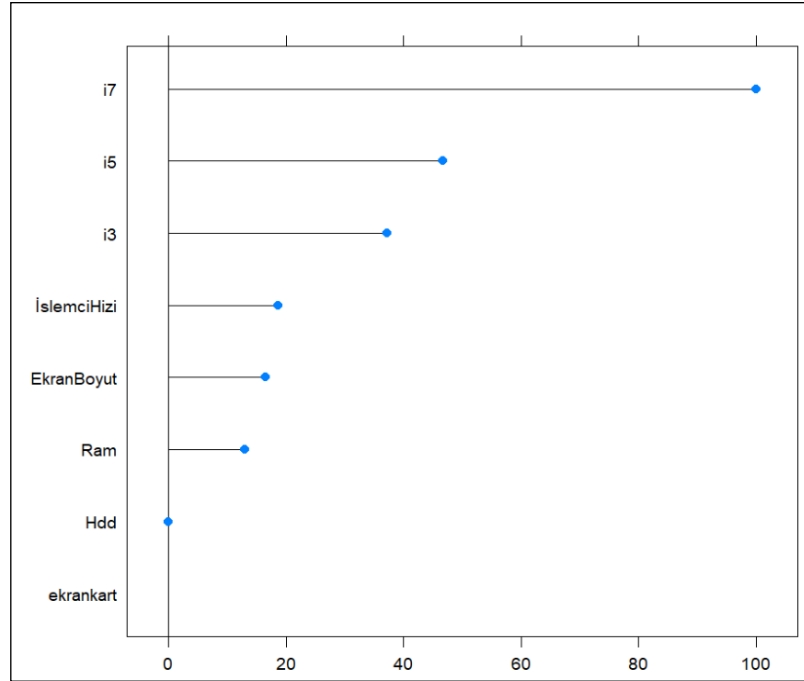
Ridge, Lasso ve Elastik net yöntemleri ile elde edilen sonuçlar **şekil 4.24-4.26**'da görülmektedir. Sonuçlara göre; X_5 (İşlemci- i7), X_6 (İşlemci- i5), X_7 (İşlemci- i3) ve X_1 (işlemci hızı) değişkenlerinin önem düzeyi bakımından ön plana çıktığı görülmektedir.



Şekil 4.24 Ridge önemli değişkenler (DogDog)

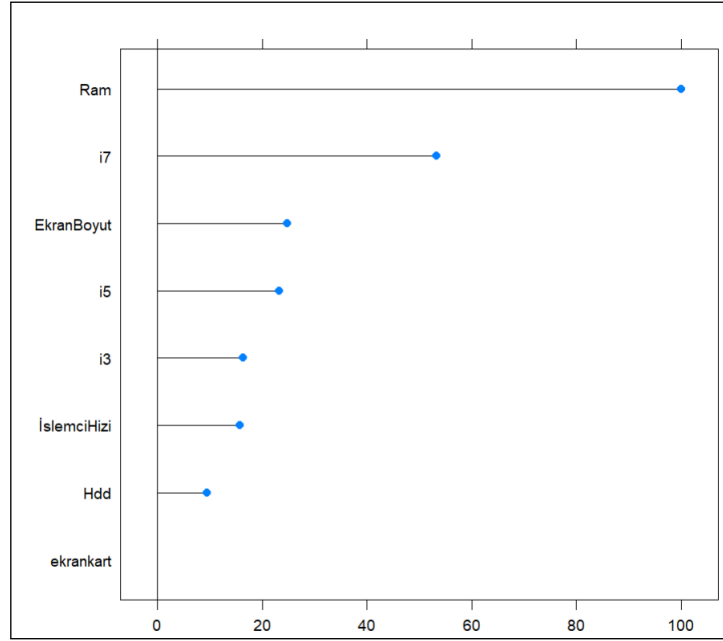


Şekil 4.25 Lasso önemli değişkenler (DogDog)

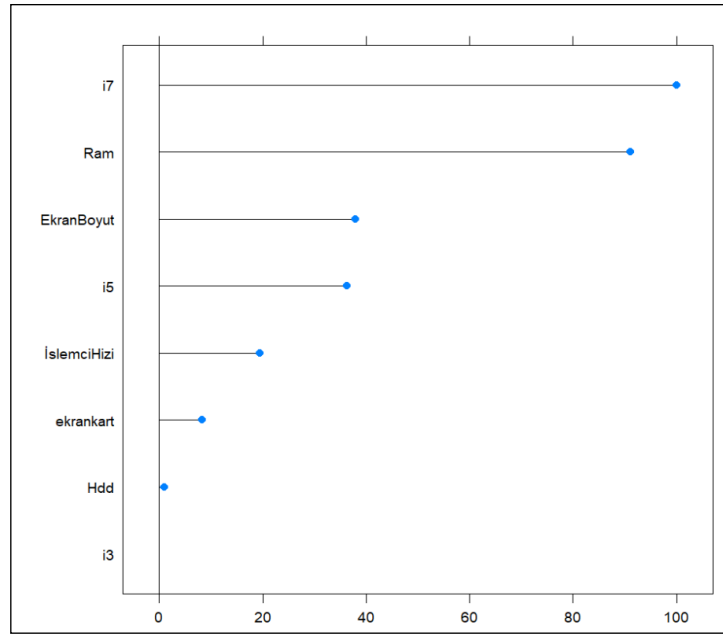


Şekil 4.26 Elastiknet önemli değişkenler (DogDog)

EKK ve robust yöntemler kapsamında X_5 (İşlemci- i7), X_2 (Ram) ve X_4 (ekran boyutu) değişkenlerinin önem düzeyi bakımından ön plana çıktığı görülmektedir (Şekil 4.27-4.28).



Şekil 4.27 EKK önemli değişkenler (DogDog)



Şekil 4.28 Robust önemli değişkenler (DogDog)

Tam doğrusal olarak kurulan hedonik regresyon modelinde yer alan bağımsız değişkenlerin önem düzeyi birden fazla yöntem ile incelenmiş olup en önemli 3 değişken uygulanan yöntem bazında **çizelge 4.22**'de verilmiştir. **Çizelge 4.22**'de ayrıca kullanılan yöntemlerin performansını ölçen MAE, RMSE ve R^2 gibi göstergeler de hesaplanmıştır.

Analiz sonuçlarına göre kullanılan yöntemlerin hepsinin hemen hemen aynı performansta olduğu gözlenmiştir. En önemli değişkenler bütünsel olarak incelendiğinde ise X_5 (İşlemci- i7), X_6 (İşlemci- i5), X_7 (İşlemci- i3), X_4 (ekran boyutu), X_2 (Ram) ve X_1 (işlemci hızı)değişkenlerinin diğer bağımsız değişkenlere nazaran daha önemli olduğu görülmüştür (**Çizelge 4.22**).

Çizelge 4.22 Değişken önem düzeyi sonuçları (DogDog)

	En Önemli 3 bağımsız değişken	MAE	RMSE	R²
EKK	(X_2, X_5, X_4)	1312	1715	0,93
Robust	(X_5, X_2, X_4)	1286	1729	0,93
Ridge	(X_5, X_6, X_1)	1316	1899	0,92
Lasso	(X_5, X_6, X_7)	1309	1715	0,93
Elastik net	(X_5, X_6, X_7)	1309	1715	0,93

4.3.2 Korelasyon analizi

Tam doğrusal olarak kurmuş olduğumuz hedonik regresyon modelinde bağımsız değişkenler arasındaki ilişki düzeyi ölçülmüştür. Öncelikle değişkenler arasındaki potansiyel çoklu bağlantıyı tespit etmek amacıyla değişkenlerin Varyans Şişirme Faktörü (VIF) ölçülmüştür.

Çizelge 4.23 Bağımsız değişkenlerin VIF değerleri (DogDog)

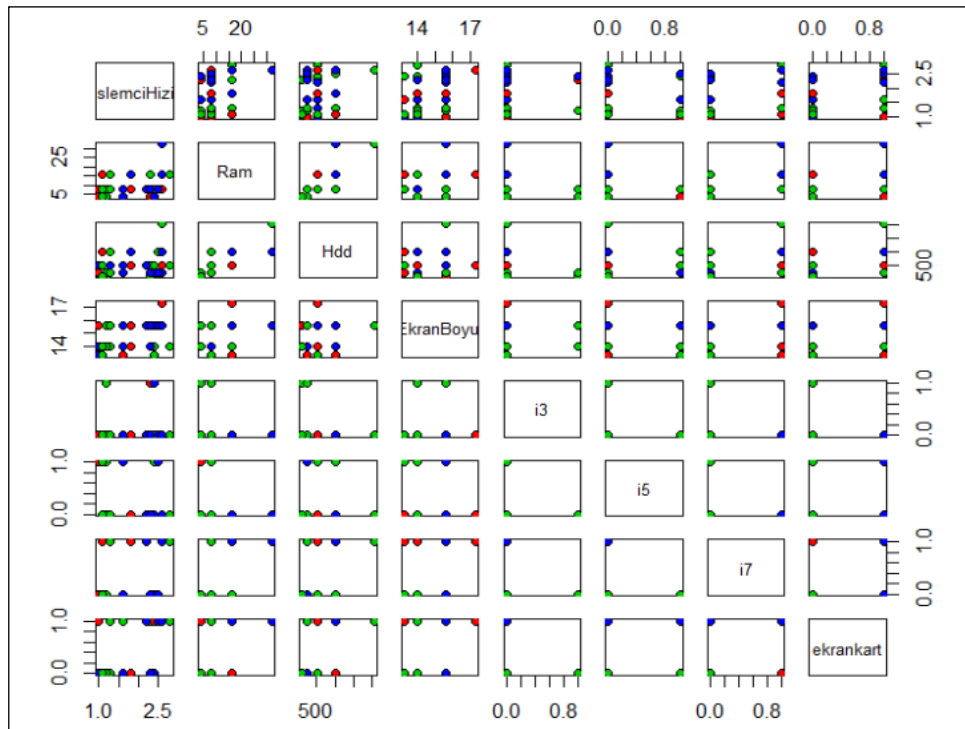
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
VIF	2,57	3,90	3,93	2,38	3,81	2,76	2,56	2,67

Çizelge 4.23'e göre X_2 (Ram) ve X_3 (Harddisk) değişkenlerinin VIF değerlerinin yüksek olduğu görülmüştür. Bu değerler iki değişken arasında önemli düzeyde bir ilişki olduğu şüphesini uyandırmıştır. Bu sebeple pearson korelasyon katsayısı yöntemi ile değişkenlerin korelasyon matrisi oluşturulmuştur (**Çizelge 4.24**).

Çizelge 4.24 Bağımsız değişkenlere ait korelasyon matrisi (DogDog)

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
X ₁	1,00	0,33	0,22	0,62	0,35	-0,30	-0,16	0,64
X ₂	0,33	1,00	0,82	0,14	0,55	-0,30	-0,37	0,36
X ₃	0,22	0,82	1,00	-0,06	0,56	-0,28	-0,33	0,17
X ₄	0,62	0,14	-0,06	1,00	0,16	-0,25	0,11	0,59
X ₅	0,35	0,55	0,56	0,16	1,00	-0,60	-0,44	0,22
X ₆	-0,30	-0,30	-0,28	-0,25	-0,60	1,00	-0,18	0,00
X ₇	-0,16	-0,37	-0,33	0,11	-0,44	-0,18	1,00	-0,32
X ₈	0,64	0,36	0,17	0,59	0,22	0,00	-0,32	1,00

Çizelge 4.24 ve Şekil 4.29’da görüleceği üzere X₂ (Ram) ve X₃ (Harddisk) değişkenleri arasındaki korelasyonun 0,82 yani güçlü ve pozitif yönlü olduğu görülmektedir. Bu nedenle 2 değişkenden birinin modelden çıkarılmasının devam eden analizlerin daha etkin yapılmasına olanak sağlayacaktır. Bu bağlamda çizelge 4.20’de her iki değişkenin t değerleri incelendiğinde t değeri düşük olan X₃ (Harddisk) değişkeninin modelden çıkarılmasına karar verilmiştir.



Şekil 4.29 Korelasyon grafiği (DogDog)

X_3 (Harddisk) deęişkeninin modelden çıkarılması ile birlikte elde edilen parametre tahminleri **çizelge 4.25**'te yer almaktadır. Kurulan tam doğrusal hedonik regresyon modelinde parametre tahminleri hem klasik EKK yöntemi hem de robust yöntemlerden Huber-M ve MM tahmin yöntemleri ile elde edilmiştir (**Çizelge 4.25**).

Çizelge 4.25 Korelasyon analizi sonrası tahmin sonuçları (DogDog)

	EKK			Robust (Huber-M)			Robust (MM)		
	Tahmin	St Hata	t- deęeri	Tahmin	St Hata	t- deęeri	Tahmin	St Hata	t- deęeri
μ	11286,23	528,69	21,35	11877,88	458,52	25,90	12351,62	284,45	43,42
X_1	990,22	68,79	14,40	863,40	59,66	14,47	568,59	53,72	10,58
X_2	673,54	4,81	139,95	631,01	4,17	151,18	522,21	7,53	69,37
X_4	-855,9	38,09	-22,47	-855,57	33,04	-25,90	-783,4	22,33	-35,0
X_5	5369,87	105,18	51,06	5440,16	91,22	59,64	5571,29	56,11	99,29
X_6	2544,22	118,46	21,48	2410,51	102,74	23,46	2045,89	68,15	30,02
X_7	2021,69	136,66	14,79	1824,52	118,52	15,39	1195,24	63,45	18,84
X_8	88,35	92,56	0,96	195,54	80,28	2,44	353,27	70,05	5,04

Çoklu bağlantı sorununa yol açan X_3 deęişkeninin modelden çıkarılması ile elde edilen model bazında model standart hataları **çizelge 4.26**'da görülmektedir. Sonuçlar incelendiğinde X_3 deęişkeninin modelden çıkarılmasının belirgin bir fark yaratmadığı sonucuna varılmıştır.

Çizelge 4.26 Korelasyon analizi sonrası model standart hataları (DogDog)

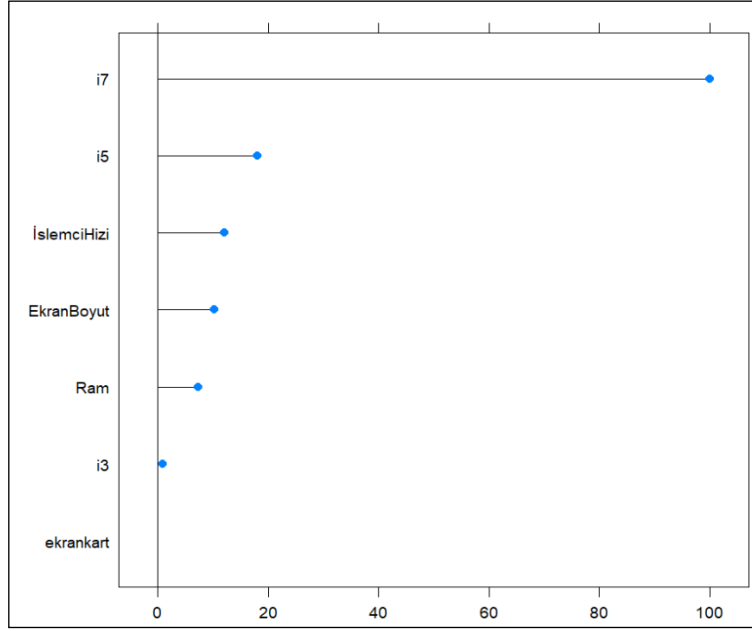
	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası-ilk	1715	1304	1050
Model standart hatası- Korelasyon analizi sonrası	1735	1315	1059

4.3.3 Korelasyon analizi sonrası deęişkenlerin önem düzeyi

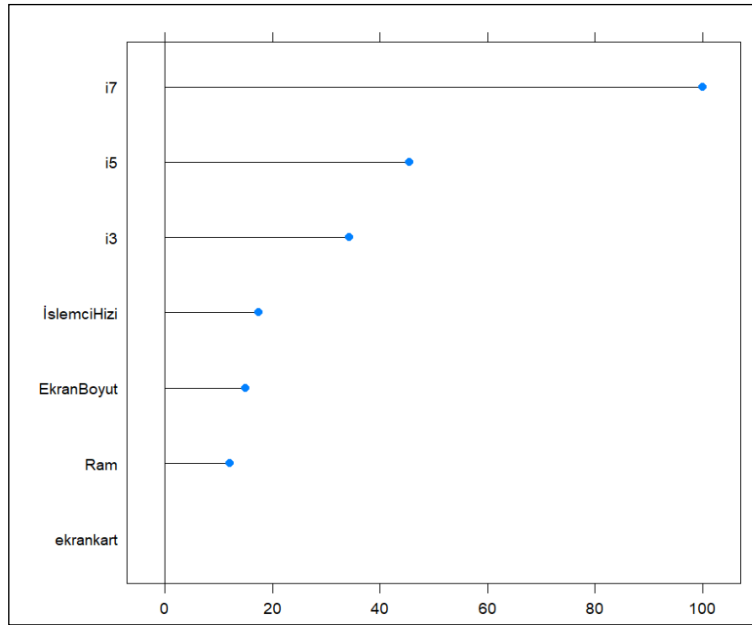
Korelasyon analizi sonrası, X_3 deęişkeninin çoklu bağlantı problemi nedeniyle model çıkarıldığı durumda da deęişkenlerin önem düzeyi ölçülmüştür. Dizüstü bilgisayar

fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla Ridge, Lasso ve Elastiknet yöntemleri ile birlikte EKK ve robust yöntemlerden faydalanılmıştır.

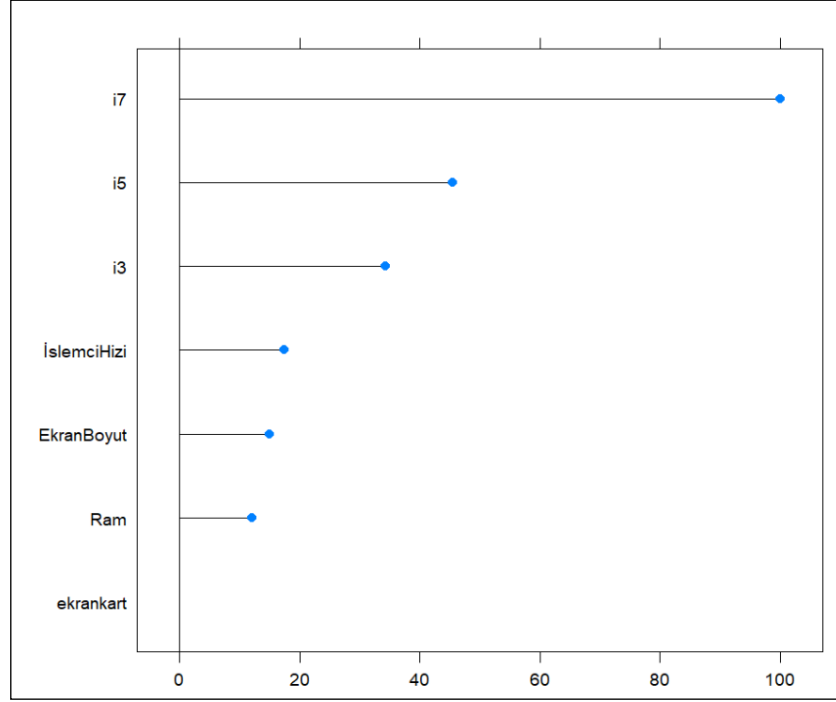
Ridge, Lasso ve Elastiknet yöntemleri ile elde edilen değişken önem düzeyi sonuçları **şekil 4.30-4.32**'de görülmektedir.



Şekil 4.30 Korelasyon analizi sonrası Ridge önemli değişkenler (DogDog)

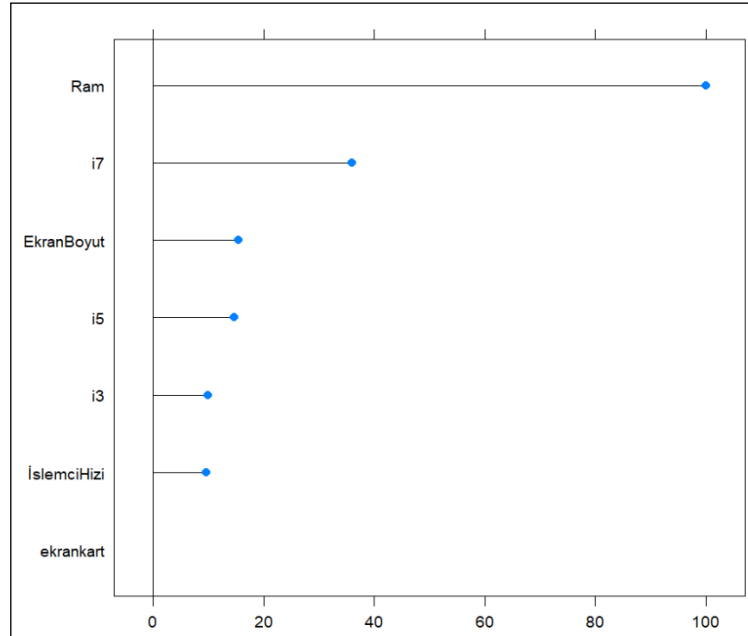


Şekil 4.31 Korelasyon analizi sonrası Lasso önemli değişkenler (DogDog)

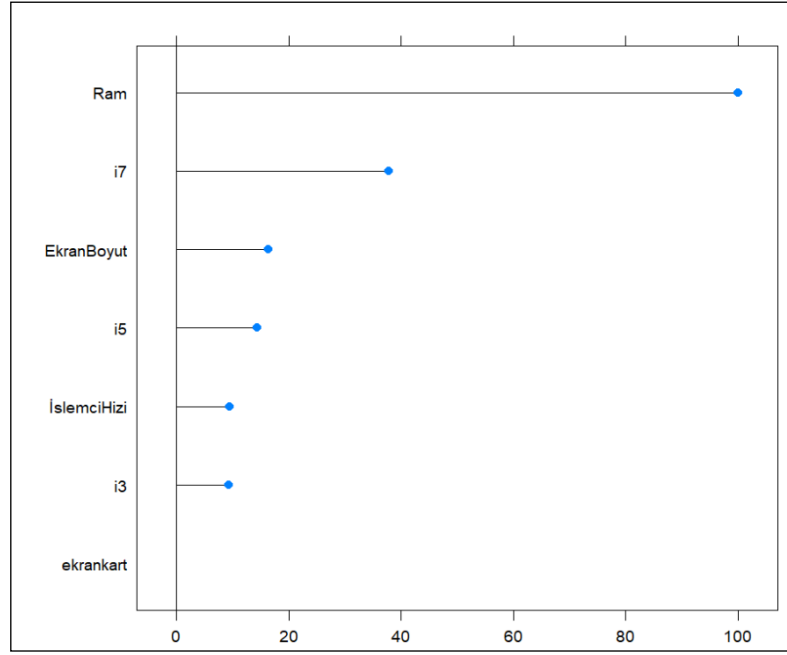


Şekil 4.32 Korelasyon analizi sonrası Elastiknet önemli değişkenler (DogDog)

EKK ve robust yöntemleri ile elde edilen değişken önem düzeyi sonuçları **şekil 4.33-4.34**'de görülmektedir.



Şekil 4.33 Korelasyon analizi sonrası EKK önemli değişkenler (DogDog)



Şekil 4.34 Korelasyon analizi sonrası Robust önemli değişkenler (DogDog)

X_3 değişkeninin çoklu bağlantı problemi nedeniyle model çıkarıldığı durumda tam doğrusal hedonik regresyon modeli ile elde edilen tahminler doğrultusunda Ridge, Lasso ve Elastik net yöntemlerine ek olarak klasik EKK ve robust yöntemler ile ölçülen değişkenlerin önem düzeyi ile ilgili hesaplamalar **çizelge 4.27**'de verilmiştir. **Çizelge 4.27**'de ayrıca kullanılan yöntemlerin performansını ölçen MAE, RMSE ve R^2 gibi göstergeler de hesaplanmıştır. Analiz sonuçlarına göre kullanılan yöntemlerin hepsinin hemen hemen aynı performansta olduğu gözlenmiştir. En önemli değişkenler bütünsel olarak incelendiğinde ise X_5 (İşlemci- i7), X_6 (İşlemci- i5), X_7 (İşlemci- i3), X_4 (ekran boyutu), X_2 (Ram) ve X_1 (İşlemci hızı) değişkenlerinin diğer bağımsız değişkenlere nazaran daha önemli olduğu görülmüştür. **Çizelge 4.27**'de elde edilen sonuçların **çizelge 4.22**'de elde edilen sonuçlar ile oldukça yakın olduğu görülmektedir Bu durum X_3 değişkeninin modelden çıkarılmasının sonuçlara belirgin bir etkisinin olmadığını göstermektedir.

Çizelge 4.27 Korelasyon analizi sonrası değişken önem düzeyi sonuçları (DogDog)

	En Önemli 3 bağımsız değişken	MAE	RMSE	R²
EKK	(X_2, X_5, X_4)	1862	2582	0,85
Robust	(X_2, X_5, X_4)	1837	2678	0,84
Ridge	(X_5, X_6, X_1)	1849	2653	0,84
Lasso	(X_5, X_6, X_7)	1861	2582	0,85
Elastik net	(X_5, X_6, X_7)	1861	2582	0,85

4.3.4 Veride aykırı değer olması durumu

Çizelge 4.28’de veri setine belli miktarda hatalı veri girişi olduğu durumda 3 tahmin yönteminden elde edilen model standart hataları yer almaktadır. Ayrıca çizelge 4.28’de veri setine hatalı veri girişi yapılmadan önceki model standart hataları da karşılaştırma yapabilmek amacıyla “Model standart hatası-ilk” olarak yer almaktadır.

Çizelge 4.28 Veride aykırı değer olduğunda model standart hataları (DogDog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası-ilk	1715	1304	1050
Model standart hatası-aykırı değer olduğu durum	3437	1693	1215

Çizelge 4.28 incelendiğinde, veri setine hatalı veriler girildiğinde EKK yönteminin model standart hata değerinin yükseldiği görülmüştür. Robust tahmin edicilerden Huber-M ve MM yöntemlerinin ise model standart hatalarının çok fazla değişmediği gözlenmiştir. Bu noktada robust yöntemlerin daha etkin sonuç verdiği görülmekle birlikte yapılan tahminlerde robust yöntemlerin tercih edilmesi önerilmektedir.

4.4 Doğrusal Logaritmik Model (DogLog)

Dizüstü bilgisayarın fiyatını etkileyen 8 faktörden oluşan doğrusal logaritmik regresyon modeli kurulmuştur. Analizler için kurulan doğrusal logaritmik hedonik regresyon modeli aşağıda verilmiştir:

$$P_i = \alpha + \beta_1 \ln X_{i1} + \beta_2 \ln X_{i2} + \beta_3 \ln X_{i3} + \dots + \beta_8 \ln X_{i8} + \varepsilon_i, \quad i = 1, \dots, n. \quad (4.7)$$

Kurulan hedonik regresyon modelinden En Küçük Kareler (EKK), robust yöntemlerden Huber-M ve MM yöntemi ile elde edilen parametre tahminleri, standart hatalar ve elde edilen t değerleri **çizelge 4.29**'da yer almaktadır. Ayrıca elde edilen p değerleri de incelenmiş olup bulguları desteklediği görülmüştür.

Çizelge 4.29 Tahmin sonuçları (DogLog)

	EKK			Robust (Huber-M)			Robust (MM)		
	Tahmin	St Hata	t- değeri	Tahmin	St Hata	t- değeri	Tahmin	St Hata	t- değeri
μ	16516,20	2527,00	6,54	12011,57	1640,20	7,32	11352,71	1262,57	8,99
X_1	2003,80	175,20	11,44	804,48	113,68	7,08	106,36	83,01	1,28
X_2	7873,00	158,00	49,83	5321,38	102,54	51,90	3946,31	88,75	44,47
X_3	801,90	143,30	5,60	1669,83	93,00	17,95	1615,60	106,49	15,17
X_4	-11712,2	880,70	-13,30	-9613,41	571,65	-16,82	-8041,67	432,30	-18,60
X_5	3899,60	178,00	21,91	3884,17	115,52	33,62	4228,06	72,73	58,14
X_6	1578,70	183,00	8,63	896,84	118,78	7,55	720,24	71,19	10,12
X_7	3709,70	211,90	17,51	2202,04	137,52	16,01	1103,69	100,48	10,99
X_8	593,60	135,80	4,37	561,75	88,16	6,37	409,30	66,55	6,15

Tahmin yöntemlerin etkinliklerini karşılaştırmak amacıyla her bir tahmin yönteminin model standart hataları ise **çizelge 4.30**'da yer almaktadır. **Çizelge 4.30**'a göre standart hatalar birbirine oldukça yakın olmak ile birlikte en düşük standart hata değeri MM tahmin yöntemine aittir.

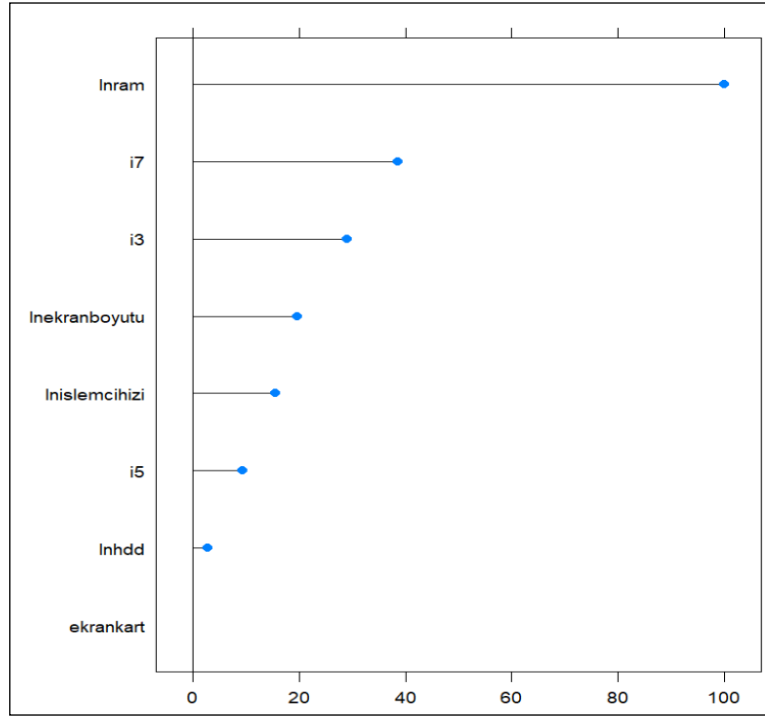
Çizelge 4.30 Tahmin yöntemlerinin model standart hataları (DogLog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası	2573	1527	1135

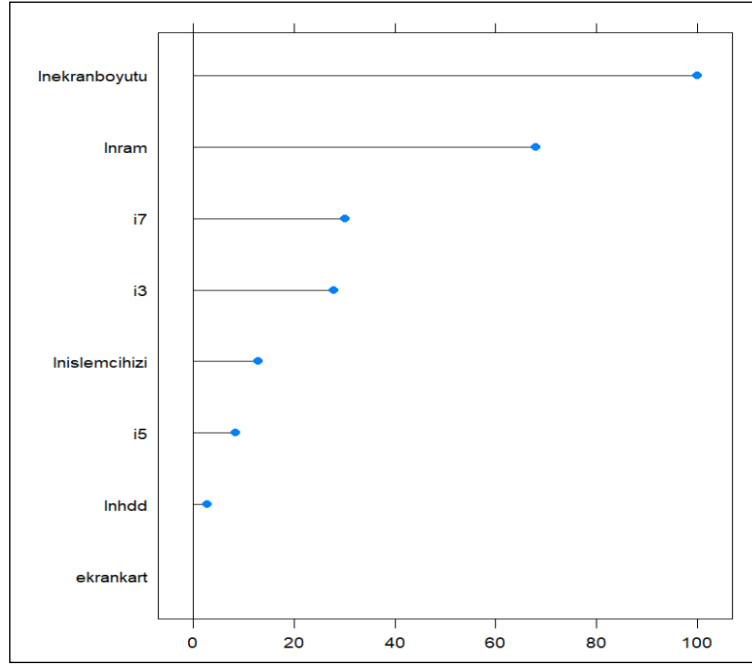
4.4.1 Değişkenlerin önem düzeyi

Dizüstü bilgisayar fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla Ridge, Lasso ve Elastiknet yöntemlerinden faydalanılmıştır. Her üç yöntem için yapılan analizler sonucunda fiyat değişkenini en çok etkileyen 3 değişken ön plana çıkarılmıştır.

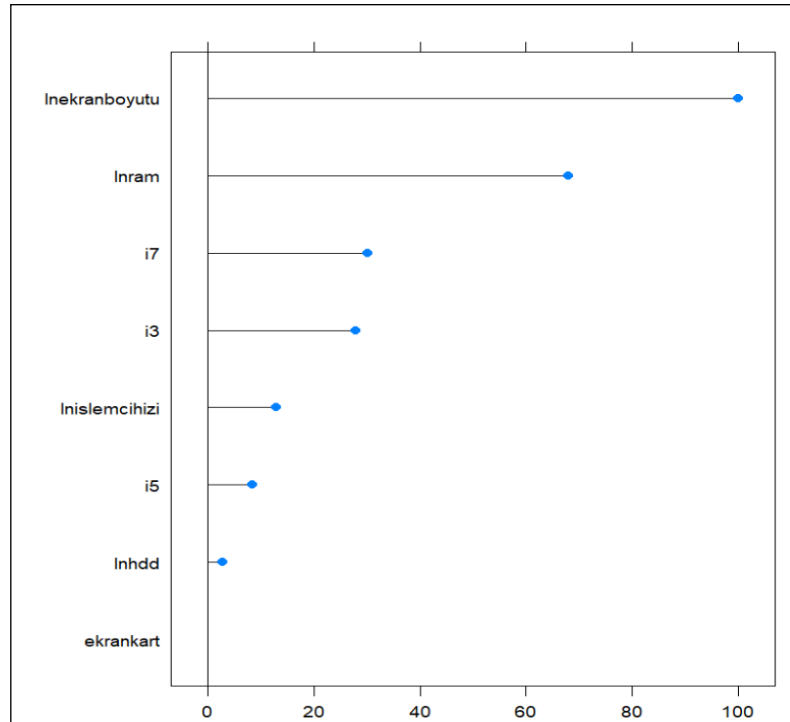
Ridge, Lasso ve Elastik net yöntemleri ile elde edilen sonuçlar **şekil 4.35-4.37**'de görüleceği üzere X_5 (İşlemci- i7), X_7 (İşlemci- i3), X_2 (Ram) ve X_4 (Ekran boyutu) değişkenlerinin önem düzeyi bakımından ön plana çıktığı görülmektedir.



Şekil 4.35 Ridge önemli değişkenler (DogLog)

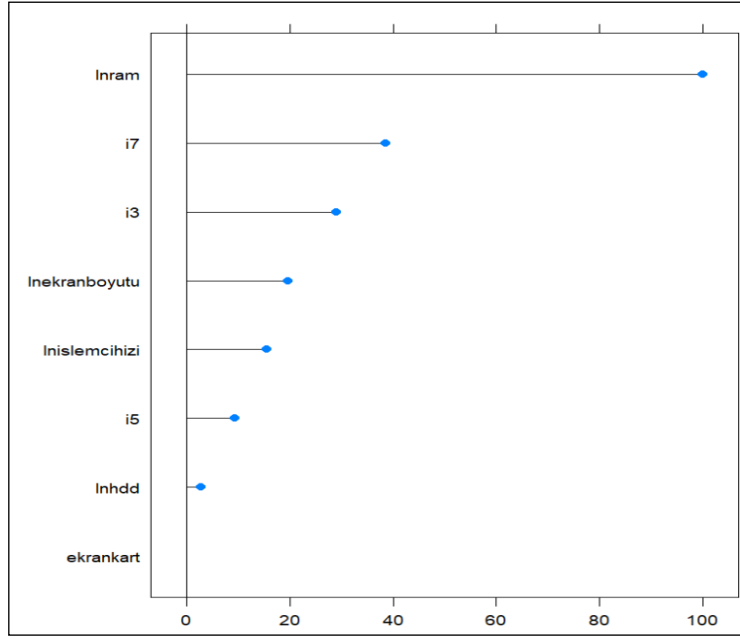


Şekil 4.36 Lasso önemli değişkenler (DogLog)

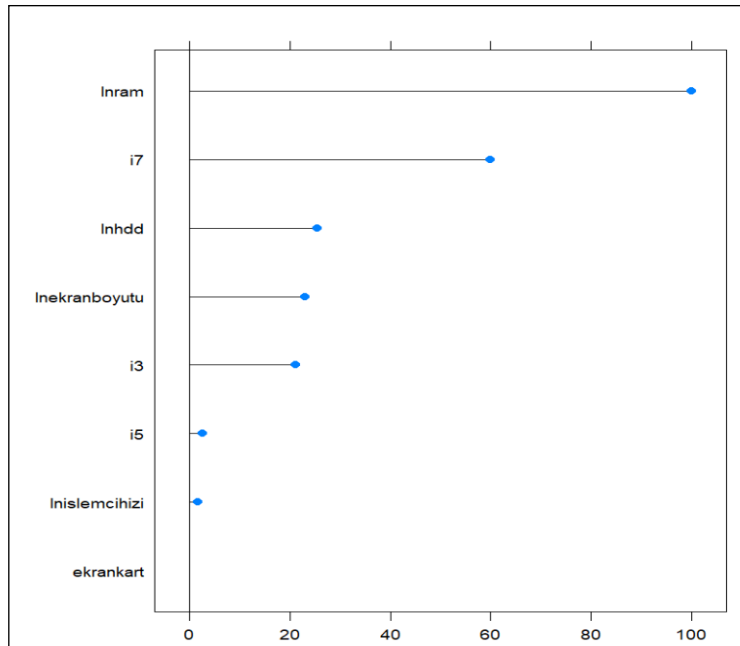


Şekil 4.37 Elastiknet önemli değişkenler (DogLog)

EKK ve robust yöntemler kapsamında X_5 (İşlemci- i7), X_2 (Ram), X_7 (İşlemci- i3) ve X_3 (harddisk) değişkenlerinin önem düzeyi bakımından ön plana çıktığı görülmektedir (Şekil 4.38-4.39).



Şekil 4.38 EKK önemli değişkenler (DogLog)



Şekil 4.39 Robust önemli değişkenler (DogLog)

Doğrusal logaritmik olarak kurulan hedonik regresyon modelinde yer alan bağımsız değişkenlerin önem düzeyi birden fazla yöntem ile incelenmiş olup en önemli 3 değişken uygulanan yöntem bazında **çizelge 4.31**'de verilmiştir. **Çizelge 4.31**'de ayrıca kullanılan yöntemlerin performansını ölçen MAE, RMSE ve R^2 gibi göstergeler de hesaplanmıştır.

Analiz sonuçlarına göre kullanılan yöntemlerin hepsinin hemen hemen aynı performansta olduğu gözlenmiştir. En önemli değişkenler bütünsel olarak incelendiğinde ise X_2 (Ram), X_5 (İşlemci- i7), X_7 (İşlemci- i3), X_4 (ekran boyutu) ve X_3 (harddisk) değişkenlerinin diğer bağımsız değişkenlere nazaran daha önemli olduğu görülmüştür (**Çizelge 4.31**).

Çizelge 4.31 Değişken önem düzeyi sonuçları (DogLog)

	En Önemli 3 bağımsız değişken	MAE	RMSE	R ²
EKK	(X_2, X_5, X_7)	1831	2572	0,85
Robust	(X_2, X_5, X_3)	1674	2780	0,84
Ridge	(X_2, X_5, X_7)	1789	2644	0,84
Lasso	(X_4, X_2, X_5)	1829	2572	0,85
Elastik net	(X_4, X_2, X_5)	1829	2572	0,85

4.4.2 Korelasyon analizi

Doğrusal logaritmik olarak kurmuş olduğumuz hedonik regresyon modelinde bağımsız değişkenler arasındaki ilişki düzeyi ölçülmüştür. Öncelikle değişkenler arasındaki potansiyel çoklu bağlantıyı tespit etmek amacıyla değişkenlerin Varyans Şişirme Faktörü (VIF) ölçülmüştür.

Çizelge 4.32 Bağımsız değişkenlerin VIF değerleri (DogLog)

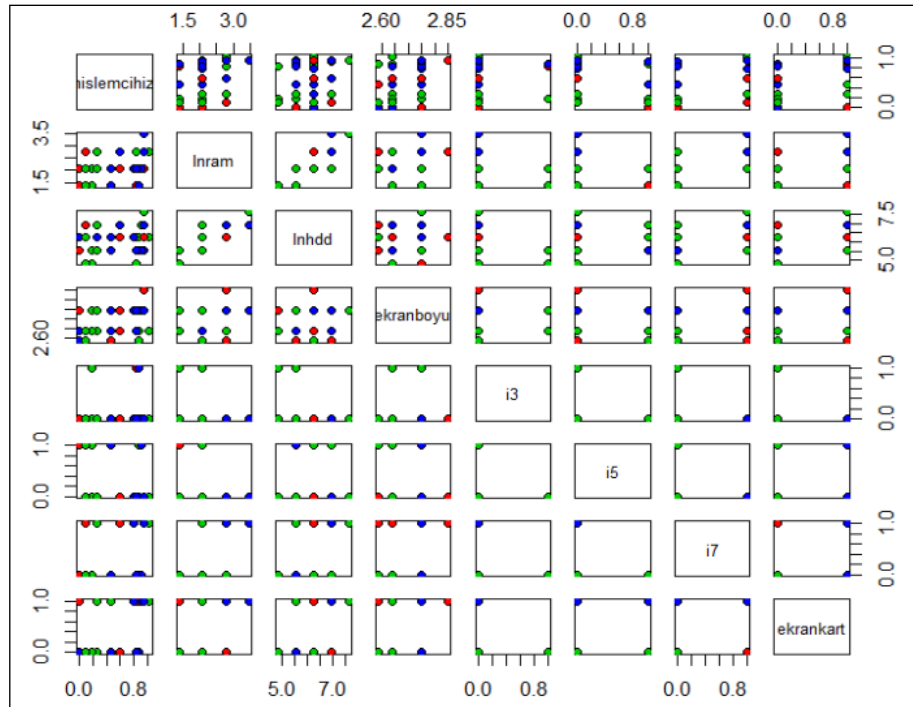
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
VIF	2,20	5,00	5,13	2,40	4,21	2,88	2,52	2,53

Çizelge 4.32'ye göre X_2 (Ram) ve X_3 (Harddisk) değişkenlerinin VIF değerlerinin yüksek olduğu görülmüştür. Bu değerler iki değişken arasında önemli düzeyde bir ilişki olduğu şüphesini uyandırmıştır. Bu sebeple pearson korelasyon katsayısı yöntemi ile değişkenlerin korelasyon matrisi oluşturulmuştur (**Çizelge 4.33**).

Çizelge 4.33 Bağımsız değişkenlere ait korelasyon matrisi (DogLog)

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1	1,00	0,29	0,27	0,60	0,35	-0,31	-0,15	0,59
X_2	0,29	1,00	0,85	0,09	0,65	-0,28	-0,52	0,37
X_3	0,27	0,85	1,00	-0,04	0,68	-0,26	-0,45	0,20
X_4	0,60	0,09	-0,04	1,00	0,15	-0,25	0,11	0,59
X_5	0,35	0,65	0,68	0,15	1,00	-0,60	-0,44	0,22
X_6	-0,31	-0,28	-0,26	-0,25	-0,60	1,00	-0,18	0,00
X_7	-0,15	-0,52	-0,45	0,11	-0,44	-0,18	1,00	-0,32
X_8	0,59	0,37	0,20	0,59	0,22	0,00	-0,32	1,00

Çizelge 4.33 ve Şekil 4.40’da görüleceği üzere X_2 (Ram) ve X_3 (Harddisk) değişkenleri arasındaki korelasyonun 0,85 yani güçlü ve pozitif yönlü olduğu görülmektedir. Bu nedenle 2 değişkenden birinin modelden çıkarılmasının devam eden analizlerin daha etkin yapılmasına olanak sağlayacaktır. Bu bağlamda çizelge 4.29’da her iki değişkenin t değerleri incelendiğinde t değeri düşük olan X_3 (Harddisk) değişkeninin modelden çıkarılmasına karar verilmiştir.



Şekil 4.40 Korelasyon grafiği (DogLog)

X_3 (Harddisk) deęişkeninin modelden çıkarılması ile birlikte elde edilen parametre tahminleri **çizelge 4.34**'te yer almaktadır. Kurulan doğrusal logaritmik hedonik regresyon modelinde parametre tahminleri hem klasik EKK yöntemi hem de robust yöntemlerden Huber-M ve MM tahmin yöntemleri ile elde edilmiştir (**Çizelge 4.34**).

Çizelge 4.34 Korelasyon analizi sonrası tahmin sonuçları (DogLog)

	EKK			Robust (Huber-M)			Robust (MM)		
	Tahmin	St Hata	t-deęeri	Tahmin	St Hata	t-deęeri	Tahmin	St Hata	t-deęeri
μ	22460,10	2302,5	9,76	25156,7	1630,6	15,43	23273,6	1088,6	21,38
X_1	2271,00	169,20	13,42	1583,89	119,84	13,22	567,24	96,77	5,86
X_2	8544,70	103,10	82,85	7031,69	73,04	96,27	5072,51	105,12	48,26
X_4	-12823,4	861,60	-14,88	-12497,8	610,20	-20,48	-10085,0	440,70	-22,88
X_5	4320,00	162,00	26,66	4668,37	114,74	40,69	5033,64	53,97	93,26
X_6	1875,80	175,80	10,67	1509,99	124,54	12,12	1260,94	76,53	16,48
X_7	3960,50	207,90	19,05	2971,46	147,27	20,18	1532,80	88,94	17,23
X_8	457,70	134,20	3,41	276,70	95,04	2,91	235,80	82,09	2,87

Çizelge 4.35 Korelasyon analizi sonrası model standart hataları (DogLog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası-ilk	2573	1527	1135
Model standart hatası-Korelasyon analizi sonrası	2583	1845	1138

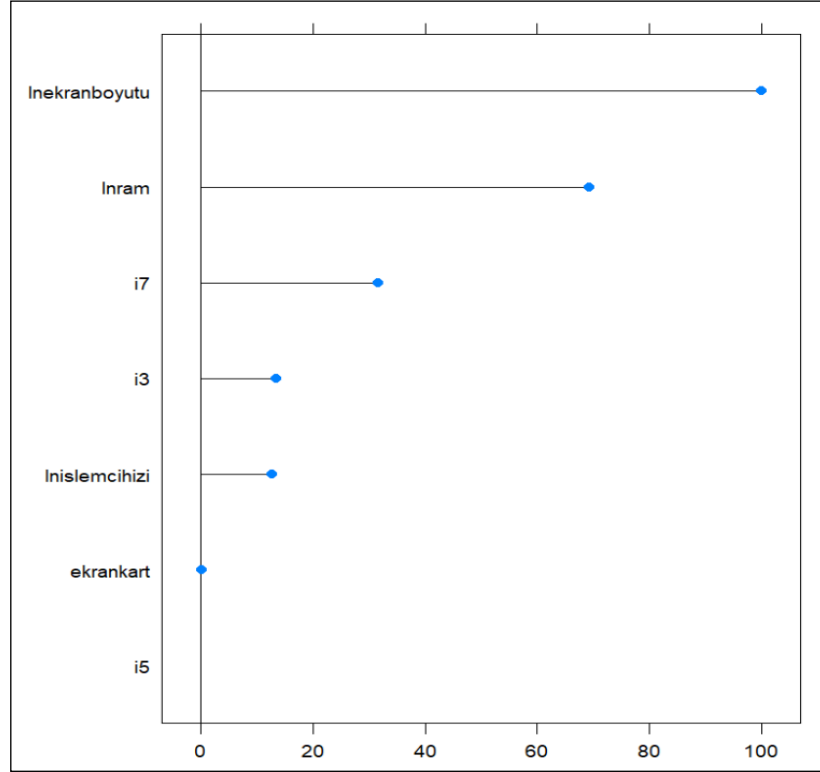
Çoklu bağlantı sorununa yol açan X_3 deęişkeninin modelden çıkarılması ile elde edilen model bazında model standart hataları **çizelge 4.35**'te görülmektedir. Sonuçlar incelendiğinde X_3 deęişkeninin modelden çıkarılmasının belirgin bir fark yaratmadığı sonucuna varılmıştır.

4.4.3 Korelasyon analizi sonrası deęişkenlerin önem düzeyi

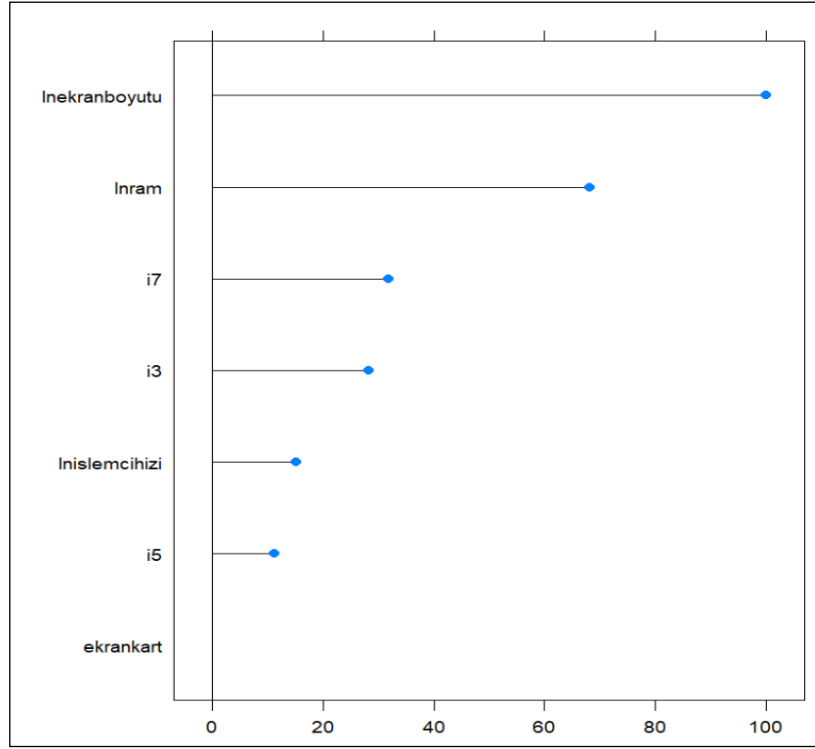
Korelasyon analizi sonrasında, X_3 deęişkeninin çoklu bağlantı problemi nedeniyle model çıkarıldığı durumda da deęişkenlerin önem düzeyi ölçülmüştür. Dizüstü bilgisayar

fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla Ridge, Lasso ve Elastiknet yöntemleri ile birlikte EKK ve robust yöntemlerden faydalanılmıştır.

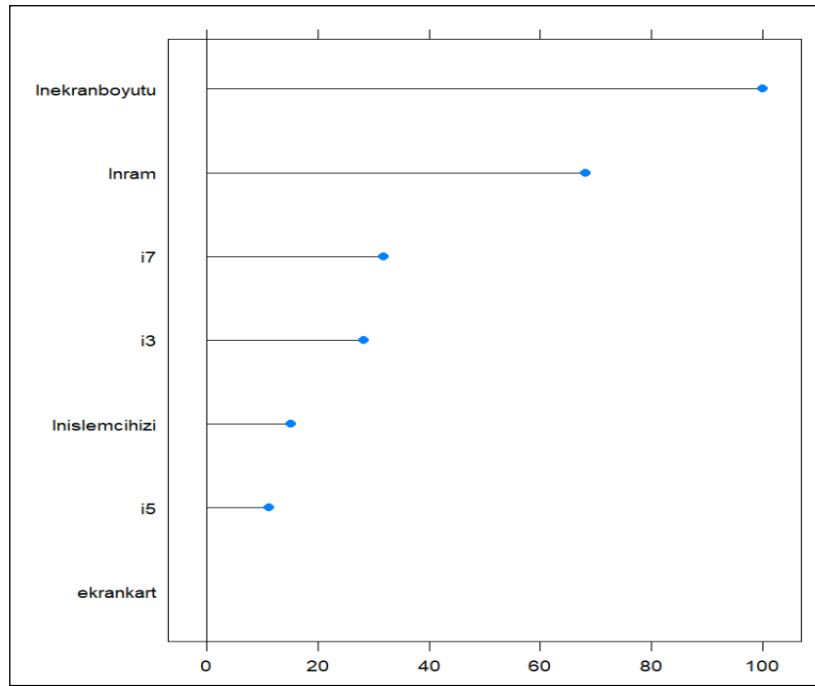
Ridge, Lasso ve Elastiknet yöntemleri ile elde edilen değişken önem düzeyi sonuçları **şekil 4.41-4.43**'te görülmektedir.



Şekil 4.41 Korelasyon analizi sonrası Ridge önemli değişkenler (DogLog)

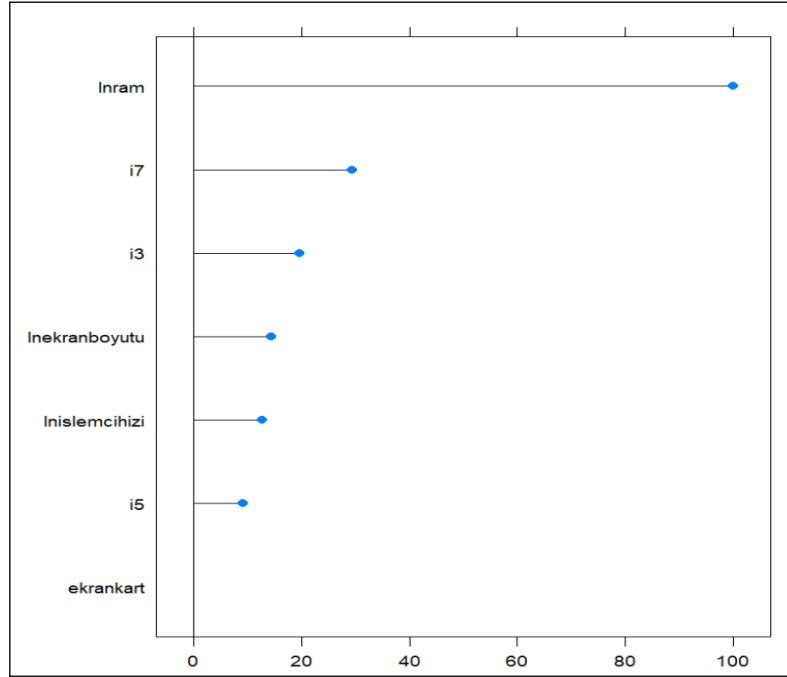


Şekil 4.42 Korelasyon analizi sonrası Lasso önemli değişkenler (DogLog)

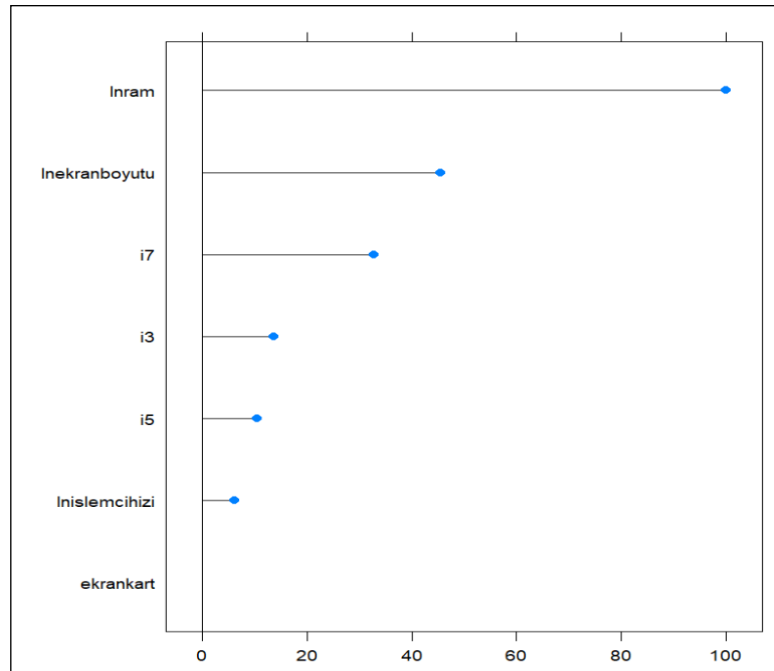


Şekil 4.43 Korelasyon analizi sonrası Elastiknet önemli değişkenler (DogLog)

EKK ve robust yöntemleri ile elde edilen değişken önem düzeyi sonuçları **şekil 4.44-4.45**'te görülmektedir.



Şekil 4.44 Korelasyon analizi sonrası EKK önemli değişkenler (DogLog)



Şekil 4.45 Korelasyon analizi sonrası Robust önemli değişkenler (DogLog)

X_3 değişkeninin çoklu bağlantı problemi nedeniyle model çıkarıldığı durumda doğrusal logaritmik hedonik regresyon modeli ile elde edilen tahminler doğrultusunda Ridge, Lasso ve Elastik net yöntemlerine ek olarak klasik EKK ve robust yöntemler ile ölçülen

değişkenlerin önem düzeyi ile ilgili hesaplamalar **çizelge 4.36**'da verilmiştir. **Çizelge 4.36**'da ayrıca kullanılan yöntemlerin performansını ölçen MAE, RMSE ve R^2 gibi göstergeler de hesaplanmıştır. Analiz sonuçlarına göre kullanılan yöntemlerin hepsinin hemen hemen aynı performansta olduğu gözlenmiştir. En önemli değişkenler bütünsel olarak incelendiğinde ise X_5 (İşlemci- i7), X_7 (İşlemci- i3), X_4 (ekran boyutu) ve X_2 (Ram) ve değişkenlerinin diğer bağımsız değişkenlere nazaran daha önemli olduğu görülmüştür. **Çizelge 4.36**'da elde edilen sonuçların **çizelge 4.31**'de elde edilen sonuçlar ile oldukça yakın olduğu görülmektedir Bu durum X_3 değişkeninin modelden çıkarılmasının sonuçlara belirgin bir etkisinin olmadığını göstermektedir.

Çizelge 4.36 Korelasyon analizi sonrası değişken önem düzeyi sonuçları (DogLog)

	En Önemli 3 bağımsız değişken	MAE	RMSE	R²
EKK	(X_2, X_5, X_7)	1862	2582	0,85
Robust	(X_2, X_4, X_5)	1837	2679	0,84
Ridge	(X_4, X_2, X_5)	1850	2654	0,84
Lasso	(X_4, X_2, X_5)	1862	2583	0,85
Elastik net	(X_4, X_2, X_5)	1862	2583	0,85

4.4.4 Veride aykırı değer olması durumu

Çizelge 4.37'de veri setine belli miktarda hatalı veri girişi olduğu durumda 3 tahmin yönteminden elde edilen model standart hataları yer almaktadır. Ayrıca **çizelge 4.37**'de veri setine hatalı veri girişi yapılmadan önceki model standart hataları da karşılaştırma yapabilmek amacıyla "Model standart hatası-ilk" olarak yer almaktadır.

Çizelge 4.37 Veride aykırı değer olduğunda model standart hataları (DogLog)

	EKK	Robust (Huber-M)	Robust (MM)
Model standart hatası-ilk	2573	1527	1135
Model standart hatası-aykırı değer olduğu durum	3610	1551	1232

Çizelge 4.37 incelendiğinde, veri setine hatalı veriler girildiğinde EKK yönteminin model standart hata değerinin yükseldiği görülmüştür. Robust tahmin edicilerden Huber-M ve MM yöntemlerinin ise model standart hatalarının çok fazla değişmediği gözlenmiştir. Bu noktada robust yöntemlerin daha etkin sonuç verdiği görülmekle birlikte yapılan tahminlerde robust yöntemlerin tercih edilmesi önerilmektedir.

4.5 Karşılaştırma

Bu bölümde dizüstü bilgisayarın fiyatını etkileyen özelliklerin fiyat tahmini için kullanılan yöntemlerin karşılaştırmalı analizleri yer almaktadır. Yapılan karşılaştırmalarda Tam Logaritmik (LogLog), Logaritmik Doğrusal (LogDog), Tam Doğrusal (DogDog) ve Doğrusal Logaritmik (DogLog) olmak üzere 4 adet hedonik regresyon model yapısı kullanılmıştır. En etkin tahmin edici ve model yapısını ön plana çıkarmak amacıyla EKK ve robust (Huber-M ve MM tahmin edicileri) yöntemlerden faydalanılmıştır. Model performansını ölçmek ve modele katkısı en yüksek olan dizüstü bilgisayar özelliklerini gözlemek amacıyla MAE, RMSE ve R^2 göstergeleri hesaplanmıştır.

Model standart hataları incelendiğinde en etkin modelin Tam Logaritmik (LogLog) ve en etkin tahmin edicinin robust MM tahmin edicisi olduğu **çizelge 4.38**'de göre görülmektedir. **Çizelge 4.39**'da ise veride aykırı değer olması durumu incelenmiş ve yine en etkin sonuçların LogLog model yapısında robust (MM tahmin edicisi) yöntemi ile elde edildiği gözlenmiştir.

Çizelge 4.38 Model standart hataları

	LogLog	LogDog	DogDog	DogLog
EKK	0,13	0,14	1715	2573
Robust (Huber-M)	0,11	0,12	1304	1527
Robust (MM)	0,10	0,11	1050	1135

Çizelge 4.39 Veride aykırı değer olduğunda model standart hataları

	LogLog	LogDog	DogDog	DogLog
EKK	0,72	0,40	3437	3610
Robust (Huber-M)	0,12	0,13	1693	1551
Robust (MM)	0,10	0,12	1215	1232

Çizelge 4.40-4.42'ye göre sırasıyla MAE, RMSE ve R^2 göstergeleri incelendiğinde en etkin sonuçların LogLog model yapısında elde edildiği görülmektedir. LogLog model yapısında tahmin edicilerin birbirine yakın performans gösterdiği görülmektedir.

Çizelge 4.40 Tüm modeller için MAE değerleri

	LogLog	LogDog	DogDog	DogLog
EKK	0,10	0,11	1312	1831
Robust	0,10	0,11	1286	1674
Ridge	0,10	0,11	1316	1789
Lasso	0,10	0,11	1309	1829
Elastiknet	0,10	0,11	1309	1829

Çizelge 4.41 Tüm modeller için RMSE değerleri

	LogLog	LogDog	DogDog	DogLog
EKK	0,13	0,14	1715	2572
Robust	0,13	0,14	1729	2780
Ridge	0,14	0,15	1899	2644
Lasso	0,13	0,14	1715	2572
Elastiknet	0,13	0,14	1715	2572

Çizelge 4.42 Tüm modeller için R^2 değerleri

	LogLog	LogDog	DogDog	DogLog
EKK	0,94	0,93	0,93	0,85
Robust	0,94	0,93	0,93	0,84
Ridge	0,94	0,93	0,92	0,84
Lasso	0,94	0,93	0,93	0,85
Elastiknet	0,94	0,93	0,93	0,85

Çizelge 4.43'de her bir model ve tahmin edici yapısında en önemli 3 dizüstü bilgisayar özelliği yer almaktadır. En etkin sonuçların elde edildiği LogLog model yapısında robust yöntem ile en önemli 3 değişkenin X_2 (Ram), X_5 (İşlemci- Intel i7) ve X_6 (İşlemci- Intel i5) olduğu görülmektedir.

Çizelge 4.43 Tüm modeller için en önemli 3 değişken

	LogLog	LogDog	DogDog	DogLog
EKK	(X_2, X_5, X_6)	(X_5, X_2, X_6)	(X_2, X_5, X_4)	(X_2, X_5, X_7)
Robust	(X_2, X_5, X_6)	(X_5, X_2, X_4)	(X_5, X_2, X_4)	(X_2, X_5, X_3)
Ridge	(X_4, X_2, X_5)	(X_5, X_6, X_7)	(X_5, X_6, X_1)	(X_2, X_5, X_7)
Lasso	(X_4, X_5, X_2)	(X_5, X_6, X_4)	(X_5, X_6, X_7)	(X_4, X_2, X_5)
Elastiknet	(X_4, X_5, X_2)	(X_5, X_6, X_4)	(X_5, X_6, X_7)	(X_4, X_2, X_5)

Yapılan analizler ve karşılaştırmalar sonucunda en etkin tahmin edici olarak belirlenen Tam logaritmik (LogLog) robust yöntem için tahmin sonuçları **çizelge 4.44**'te verilmiştir.

Çizelge 4.44 Tam logaritmik robust tahmin edici sonuçları

Değişken	Adı	Tahmini	Standart Hata
X_1	İşlemci hızı	0,15	0,01
X_2	Ram	0,49	0,01
X_3	Hard disk	0,10	0,01
X_4	Ekran boyutu	-1,13	0,05
X_5	İşlemci- Intel i7	0,50	0,01
X_6	İşlemci -Intel i5	0,23	0,01
X_7	İşlemci -Intel i3	0,10	0,01
X_8	Ekran kartı	0,08	0,01

5. SONUÇ

Bu tez çalışmasında büyük veri kullanarak dizüstü bilgisayar fiyatını etkileyen özelliklerin fiyata olan katkısının hedonik regresyon modelleri ile tahmin edilmesi amaçlanmıştır.

Web kazıma yöntemiyle dizüstü bilgisayarlara ait 2020 yılının son çeyreği için büyük veri kaynağından elde edilen teknolojik market verileri üzerine uygulama çalışması yapılmıştır. Uygulama çalışması kapsamında Tam Logaritmik (LogLog), Logaritmik Doğrusal (LogDog), Tam Doğrusal (DogDog) ve Doğrusal Logaritmik (DogLog) olmak üzere 4 farklı hedonik regresyon model yapısı kullanılmıştır. 4 ayrı hedonik regresyon modeli ile dizüstü bilgisayarın fiyatını etkileyen özelliklerin (İşlemci hızı, Ram, Hard disk, Ekran boyutu, İşlemci- Intel i7, İşlemci- Intel i5, İşlemci- Intel i3, Ekran kartı) her birinin fiyata olan katkısı hem EKK hem de robust yöntemler (Huber-M ve MM) ile tahmin edilmiştir. Yapılan analizler ve karşılaştırmalar neticesinde 4 modelin etkinliği karşılaştırılmış ve en etkin tahmini Tam Logaritmik Model (LogLog) yapısında robust tahmin edicilerin verdiği gözlenmiştir.

Dizüstü bilgisayar fiyatını etkileyen faktörlerinin önem düzeyini ölçmek amacıyla Ridge, Lasso ve Elastiknet yöntemlerinden faydalanılmıştır. Bu yöntemlere ek olarak robust ve EKK yöntemleri ile de faktörlerin önem düzeyi incelenmiştir. En etkin sonuçları veren tam logaritmik hedonik regresyon yapısında robust yöntemler ile elde edilen sonuçlara göre; dizüstü bilgisayar özelliklerinden fiyat değişkenini en çok etkileyen özelliklerin ram, işlemci-i7 ve işlemci-i5 olduğu sonucuna varılmıştır.

Tez çalışmasında kurulan hedonik regresyon model yapılarının performanslarını karşılaştırmak amacıyla RMSE, MAE ve R^2 göstergeleri kullanılmıştır. Bu göstergeler ışığında tam logaritmik hedonik regresyon yapısında robust yöntemlerin en etkin olduğu gözlenmiştir.

Kurulan hedonik regresyon modellerde çoklu bağlantı sorunu olup olmadığını tespit etmek amacıyla korelasyon analizleri yapılmıştır. Yapılan korelasyon analizleri sonucunda her model yapısında da ram ve harddisk değişkenleri arasında pozitif yönlü ve güçlü korelasyon gözlenmiştir. Devam eden analizlerde harddisk değişkeni modellerden çıkarılarak tahminler yapılsa da modelden çıkarılmadığı durumlar ile belirgin bir fark görülmemiştir.

Yapılan aykırı değer ve uç değer analizlerinde veri setinde herhangi bir sorunun bulunmadığı tespit edilmiştir. Fakat bu tarz veri setlerinde yaşanabilecek potansiyel hataların olması durumunda hangi modelin daha etkin olacağı da incelenmiştir. Veri setine hatalı ve anlamsız veriler girilerek veri setinde hataların olması durumunda en etkin tahminlerin Tam Logaritmik Model (LogLog) yapısında robust yöntemler ile elde edildiği gözlenmiştir.

Tez çalışmasında yapılan tüm analizler doğrultusunda hedonik regresyon modellerinde robust yöntemlerin etkin sonuçlar verdiği görülmektedir. Veri setlerinde yaşanabilecek hata kaygılarını ortadan kaldıran robust yöntemlerin bu tarz çalışmalarda kullanılması önerilmektedir.

Tez çalışmasının devamında diğer robust yöntemlerin de parametre tahmin süreçlerine dahil edilmesi planlanmaktadır. Ayrıca farklı ürün gruplarına da benzer model alternatifleri ile analizler uygulanarak fiyat tahmini yapılabilir. Gerçekleştirilecek bu analizler Tüketici Fiyat Endeksi kapsamında kalite düzeltmesi çalışmalarında kullanılabilir. Buna ek olarak ve farklı bir çalışma olarak EKK varsayımlarının sağlanmadığı durumlarda ne gibi yöntemler kullanılabileceği ve bu sorunların üstesinden nasıl gelinebileceği konusunda bir çalışma yapılması planlanmaktadır.

KAYNAKLAR

- Aktan, E. 2018. Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu. Bilgi Yönetimi Dergisi, Cilt:1, Sayı:1.
- Altalbany, S. I. 2021. Evaluation of Ridge, Elastic Net and Lasso Regression Methods in Precedence of Multicollinearity Problem: A Simulation Study, Journal of Applied Economics and Business Studies. Volume. 5, Issue 131-142.
- Aldabal, M. 2020. A Comparative Study of Ridge, LASSO and Elastic net Estimators. Carleton University Australian Bureau of Statistics.
- Basoğlu, K. A. ve White, C. E. 2015. Inline XBRL versus XBRL for SEC Reporting. Journal of Emerging Technologies in Accounting,12 (1): 189–199.
- Bourassa, S. C., Cantoni, E., ve Hoesli, M. 2013. Robust Hedonic Price Indexes. International Journal of Housing Markets and Analysis Vol. 9 No. 1, pp. 47-65.
- Box, G.E.P. 1953. Non-normality and tests on variances. Biometrika, 40. 318–335.
- Bulut, H. ve Zaman, T. 2018. Robust Hedonik Modellerin Karşılaştırılması: Beetle Türkiye Piyasa Fiyatını Etkileyen Faktörlerin İncelenmesi. C.Ü. İktisadi ve İdari Bilimler Dergisi, Cilt 19, Sayı 1.
- Dehon, C., Gergaud, O ve Verardi, V. 2006. Robust Hedonic Price Regressions: Unmasking Outliers is not Witchcraft.
- Cox, M. ve Ellsworth, D. 1997. Application-Controlled Demand Paging for Out-of-Core Visualization. 8th IEEE Visualization '97 Conference.
- Diewert, E. 2003. Hedonic Regressions A Consumer Theory Approach. University of Chicago Press.
- Diewert, E., Haan, J. ve Hendriks, R. 2014. Hedonic Regressions and the Decomposition of a House Price Index into Land and Structure Components. Statistics Netherlands.
- Diewert E., Haan J. ve Hendriks, R. 2010. The Decomposition of a House Price index into Land and Structures Components: A Hedonic Regression Approach. Economic Measurement Group Workshop.
- Eurostat 2018, Harmonised Index of Consumer Prices (HICP)- Methodological Manual.
- Friedman, J., Hastie, T. ve Tibshirani, R. 2010. Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1),1-22.

- Fixler, D., Fortuna, C., Greenlees, J. ve Lane, W. 1999. The Use of Hedonic Regressions to Handle Quality Change: The Experience in the U.S. CPI. U.S. Bureau of Labor Statistics.
- Gillian, C. H. ve Davashish, P. 1999. A factor Analytic Study of The Sources of Meaning in Hedonic Consumption. *European Journal of Marketing*, 33,1999, s.274.
- Goes, P. B. 2014. Big Data and IS Research, Management Information Systems Research Center. University of Minnesota, Vol. 38, No. 3.
- Goren, S.C. ve Arslan, O. 2023. A Study On Using Robust Hedonic Regression Implementation, Communications Faculty of Sciences University of Ankara Series A1: Mathematics and Statistics, DOI:10.31801/cfsuasmas.1187009
- Hans, V. D. H. 2004. User Acceptance of Hedonic Information Systems. *MIS Quarterly*, 28:4, s.696.
- Herawati, N., Nisa, K., Setiawan, E. ve Nusyirwan, T. 2018. Regularized multiple regression methods to deal with severe multicollinearity. *International Journal of Statistics and Applications*, 8(4), 167-172.
- Heritier, S., Cantoni, E., Copt, S. ve Victoria-Feser, M.P. 2009. *Robust Methods in Biostatistics*. John Wiley, Chichester.
- Hoerl, A. E. ve Kennard, R. W. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1).
- Hoerl, A.E. ve Kennard, R.W. 2000. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 42, 80-86.
- Huber, P. J. 1973. Robust regression: Asymptotics, conjectures and Monte Carlo. *The Annals of Statistics* 1 (5):799–821. doi:10.1214/aos/1176342503.
- Huber, P.J., 1964. Robust estimation of a location parameter. *Annals of Mathematical Statistics*, Vol. 35 No. 1, pp. 73-101.
- Hulten, C. R. 2003. Price hedonics: a critical review. Federal Reserve Bank of New York, 11 p.
- Hülagu, T., Kızılkaya, E. ve Gencay, A. 2016. A Hedonic House Price Index for Turkey. Central Bank of the Republic of Turkey.
- Jiang L., Phillips, C. B. ve Jun Y. 2014. A New Hedonic Regression For Real Estate Prices Applied To The Singapore Residential Market. Wles Foundation For Research In Economics Yale University.
- Krotov, V. ve Silva, L. 2018. Legality and Ethics of Web Scraping. Twenty-fourth Americas Conference on Information Systems, New Orleans.

- Lee, H. S., Park, K. ve Kim, S.Y. 2003. Estimation of Information Value on the Internet Application of Hedonic Price Model. *Electronic Commerce Research and Applications*, 2, s.74.
- Manoel, N., Felisoni, C. ve Roberto, M. 2009. A Five-Year Hedonic Price Breakdown for Desktop Personal Computer Attributes in Brazil.
- Maronna, R.A., Martin, R.D. ve Yohai, V.J. 2006. *Robust Statistics: Theory and Methods*. Wiley Series in Probability and Statistics.
- McCormack, K., 2013. Diagnostics for hedonic models using an example for cars(hedonic regression. Central Statistics Office, University College Cork, Central Statistics Office, Ireland
- Montgomery, D.C., Peck, E.A. ve Vining G.G. 2001. *Introduction to Linear Regression Analysis*.
- Myers, R. H. 1986. *Classical and modern regression with applications*. 2nd Ed, USA: PWSKENT Publishing Company. <https://lib.ugent.be/catalog/rug01:000851135>
- Selim, S. 2008. Determinants Of House Prices In Turkey: A Hedonic Regression Model. *Doğuş Üniversitesi Dergisi*.
- Snell, J. ve Menaldo, N. 2016. Web Scraping in an Era of Big Data 2.0. Bloomberg BNA. <https://www.bna.com/web-scraping-era-n57982073780/>
- Stigler, M. S. 1981. Gauss and the invention of least squares. *The Annals of Statistics*, Vol. 9, No:3, 465-474.
- Tibshirani, R. 1996. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1), 267{288.}).
- Unwin, T. 1999. Hedonic Price Indexes and Qualities of Wines. *Journal of wine Research*, 10:2, 1999, s.95.
- Vargiu, E. ve Urru, M. 2013. Exploiting web scraping in a collaborative filteringbased approach to web advertising. *Artificial Intelligence Research*, Vol. 2, No. 1.
- Yohai, V.J. 1987. High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15, 642-656.
- Zou, H., ve Hastie, T. 2005. Regularization and variable selection via the Elastic Net. *Journal of The Royal Statistical Society: Series B (Statistical Methodology)*, 67 (2), 301-320.