

T.C.  
YILDIZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

MODEL BAZLI ZAMAN SERİLERİ KÜMELEMESİ

Metin VATANSEVER

DOKTORA TEZİ

Matematik Anabilim Dalı

Matematik Programı

Danışman

Doç. Dr. İbrahim Demir

Ocak, 2020

T.C.  
YILDIZ TEKNİK ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ

**MODEL BAZLI ZAMAN SERİLERİ KÜMELEMESİ**

Metin VATANSEVER tarafından hazırlanan tez çalışması 13.01.2020 tarihinde aşağıdaki jüri üyeleri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalı, Matematik Programı **DOKTORA TEZİ** olarak kabul edilmiştir.

Doç. Dr. İbrahim DEMİR  
Yıldız Teknik Üniversitesi  
Danışman

**Jüri Üyeleri**

Doç. Dr. İbrahim DEMİR, Danışman  
Yıldız Teknik Üniversitesi

Prof. Dr. Ali Hakan BÜYÜKLÜ, Üye  
Yıldız Teknik Üniversitesi

Prof. Dr. Adnan MAZMANOĞLU, Üye  
Toros Üniversitesi

Prof. Dr. Murat SARI, Üye  
Yıldız Teknik Üniversitesi

Prof. Dr. Ali HEPŞEN, Üye  
İstanbul Üniversitesi

---

---

---

---

---

Danışmanım Doç. Dr. İbrahim DEMİR sorumluluğunda tarafımca hazırlanan Model Bazlı Zaman Serileri Kümelemesi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Metin VATANSEVER

İmza



Bu alıřma, Yıldız Teknik niversitesi Bilimsel Arařtırma Proje Koordinatrlę' nn 2015-01-05-DOP02 numaralı projesi ile desteklenmiřtir.



*Aileme*

## TEŐEKKÜR

---

“Model Bazlı Zaman Serileri Kmelemesi” bařlıklı doktora tez alıřmamın her ařamasında öneri, bilgi, tecrbe ve katkılarıyla beni ynlendiren saygıdeęer danıřman hocam Do. Dr. İbrahim DEMİR’e teőekkrlerimi sunarım.

Tez İzleme Komitemdeki saygıdeęer hocam Prof. Dr. Adnan Mazmanoęlu’na yaptıęı deęerli katkı ve ynlendirmeleri iin teőekkr ederim.

Zaman serileri, grsel veri madencilięi ve kmeleme analizi konuları ile beni tanıřtıran, eęitim hayatım boyunca bana ilham vererek yaptıęım akademik alıřmalar iin bana vizyon kazandıran deęerli hocam Prof. Dr. Ali Hakan Bykl’ye teőekkrlerimi sunarım.

Tez alıřmasında kullanılan veri kmesinin hazırlanmasında ve temini konusunda beni destekleyen Prof. Ali Hepřen’e ve REIDIN.com řirketine katkılarından dolayı teőekkrlerimi bir bor bilirim.

Bařta bu zorlu eęitim sreci olmak zere, hayatımın her anında yanımda olan, maddi ve manevi desteklerini benden esirgemeyen deęerli annem Hlbiye Vatansever ve babam Ahmet Vatansever’e sonsuz teőekkrlerimi sunarım.

Metin VATANSEVER

<b>SİMGE LİSTESİ</b>	<b>VIII</b>
<b>KISALTMA LİSTESİ</b>	<b>X</b>
<b>ŞEKİL LİSTESİ</b>	<b>XII</b>
<b>TABLO LİSTESİ</b>	<b>XIV</b>
<b>ÖZET</b>	<b>XVI</b>
<b>ABSTRACT</b>	<b>XVIII</b>
<b>1 GİRİŞ</b>	<b>1</b>
1.1 Literatür Özeti	1
1.2 Tezin Amacı	7
1.3 Orjinal Katkısı	8
1.4 Hipotez	8
<b>2 ZAMAN SERİLERİ KÜMELEMESİ</b>	<b>9</b>
2.1 Zaman Serileri Kümelemesi	9
2.2 Kümeleme Süreci	10
2.3 Uzaklık ve Benzerlik Ölçüleri	10
2.4 Zaman Serileri Kümeleme Algortimaları	14
2.5 Zaman Serileri Küme Doğrulama	25
<b>3 ZAMAN SERİLERİ MODELLEMESİ</b>	<b>37</b>
3.1 Otoregresif Modelleri	39
3.2 Hareketli Ortalamalar Modelleri	40
3.3 Otoregresif Hareketli Ortalamalar Karma Modelleri	41
3.4 Durağan Olmayan Otoregresif Hareketli Ortalamalar Modelleri	41
3.5 Durağanlık Analizi	42
3.6 Model Belirleme Kriterleri	46
3.7 Model Artıkları Durağanlık Testleri	48
<b>4 MODEL BAZLI ZAMAN SERİLERİ KÜMELEME YAKLAŞIMLARI</b>	<b>50</b>

<b>5 UYGULAMALAR</b>	<b>53</b>
5.1 Veri Kümesinin Açıklaması .....	53
5.2 Otoregresif Zaman Serisi Modellerinin Geliştirilmesi .....	55
5.3 Uzaklık Benzerlik Ölçüsünün ve Kümeleme Algoritmasının Belirlenmesi ..	57
5.4 Küme Sayısının Belirlenmesi .....	66
5.5 Otoregresif ve Bulanık C-Ortalamlar Model Öngörü Performanslarının Karşılaştırılması .....	74
<b>6 SONUÇ VE ÖNERİLER</b>	<b>85</b>
<b>KAYNAKÇA</b>	<b>87</b>
<b>A VERİ KÜMESİ İSTATİSTİKLERİ</b>	<b>92</b>
<b>B DURAĞANLIK SINAMASI</b>	<b>98</b>
B.1 Durağanlık Testi R Kodları .....	98
B.2 Konut Satış Fiyat Endeksleri ADF, PP ve KPSS Birim Kök Test Sonuçları....	104
B.3 Konut Satış Fiyat Endeksleri ADF, PP ve KPSS Birim Kök Test Sonuçları....	109
<b>C OTOMATİK AR MODELLERİ BELİRLEME R KODLARI</b>	<b>114</b>
C.1 AR Model Parametre Belirleme R Kodları .....	114
<b>D AR TAHMİNLENEN MODEL PARAMETRELERİ</b>	<b>116</b>
<b>E MATLAB Kümeleme ve VCV Kodları</b>	<b>121</b>
E.1 Hiyerarşik Kümeleme ve VCV Kodları .....	121
E.2 K-Ortalamlar Kümeleme ve VCV Kodları .....	122
E.3 SOM Kümeleme ve VCV Kodları .....	122
E.4 Bulanık C-Ortalamlar Kümeleme ve VCV Kodları .....	123
<b>F OM VE OMBBK ÖNGÖRÜ PERFORMANSLARI</b>	<b>124</b>
<b>TEZDEN ÜRETİLMİŞ YAYINLAR</b>	<b>129</b>



## SİMGE LİSTESİ

---

E	Beklenen Değer
s	Benzerlik Ölçüsü, Standart Hata
n	Birim Sayısı
m	Bulanıklık Katsayısı
$\varepsilon$	Tolerans Terimi
$\Delta$	Geri Fark Alma İşlemi
q	Hareketli Ortalama Derecesi
$\theta$	Hareketli Ortalamalar Model Parametresi
t	İterasyon, Zaman Anı
w	Kazanan Nöron
h	Komşuluk Fonksiyonu
C	Küme
$\bar{x}$	Küme Ortalaması
k	Küme Sayısı, Gecikme Sayısı
i,j	Matris İndisleri
$\lambda$	Minkowski Parametre Değeri
$\mu$	Ortalama Değer, Regresyon Modeli Sabit Terimi
$\emptyset$	Otoregresif Model Parametresi
$\alpha$	Öğrenme Katsayısı
F	Öngörü Fonksiyonu
B	Öteleme Operatörü
m	Referans Vektörleri
c	Sabit Terim
$\sigma$	Standart Sapma
p	Toplam Değişken Sayısı, Otoregresif Derecesi, Konut Metre Kare Değeri
N	Toplam Gözlem Sayısı
R	Topolojik Komşuluk
d	Uzaklık Ölçüsü, Fark Alma İşlemi
u	Üyelik Elemanı
U	Üyelik Matrisi
v	Vektör

x Veri Birimi  
X Veri Matrisi  
y Zaman Serisi



## KISALTMA LİSTESİ

---

AIC	Akaike Bilgi Kriteri
ADF	Augmented Dickey Fuller
SI	Ayrışma Endeksi
Q	Box-Pierce Testi
FCM	Bulanık C-Ortalamalar Kümeleme Yöntemi
CH	Calinski-Harabasz Endeksi
DB	Davies-Bouldin Endeksi
D	Dunn Endeksi
ARIMA	Durağan Olmayan Otoregresif Hareketli Ortalamalar Karma Modeli
AR	Düzeltilmiş R Endeksi, Otoregresif Modeli
OECD	Ekonomik İşbirliği ve Kalkınma Örgütü
FM	Fowlkes-Mellows Endeksi
VCV	Görsel Küme Doğrulama
MA	Hareketli Ortalamalar Modeli
H	Hartigan Endeksi
HI	Homojenlik Endeksi
C	Hubert-Levin Endeksi
J	Jaccard Endeksi
RMSSTD	Kök-Ortalama-Kare Standart Sapma Endeksi
KL	Krzanowski-Lai Endeksi
KPSS	Kwiatkowski-Phillips-Schmidt Shin testi
LPC	Linear Predictive Coding
LB	Ljung-Box Testi
ARMA	Otoregresif Hareketli Ortalamalar Karma Modeli
OM	Otoregresif Model
OMBBK	Otoregresif Model Bazlı Bulanık Kümeleme
PP	Phillips-Perron Testi

R	Rand Endeksi
RS	R-Kareler Endeksi
BIC	Schwarz Bayesci Bilgi Kriteri
S	Silhouette Endeksi



## ŞEKİL LİSTESİ

Şeki 1.1	Zaman Serileri Kümeleme Yöntemleri: Ham Veri, Özellik ve Model Bazlı	3
Şekil 2.1	Kümeleme Süreci .....	10
Şekil 2.2	Hiyerarşik Kümeleme Diyagramı.....	15
Şekil 2.3	Küme Doğrulama Yöntemleri.....	26
Şekil 2.4	Kümeleme Öncesi Ve Sonrası Matris Grafikleri .....	36
Şekil 3.1	Box-Jenkins Aşamaları Akış Diyagramı .....	38
Şekil 4.1	Model Bazlı Zaman Serileri Kümeleme Adımları .....	50
Şekil 5.1	Öklid Uzaklığına Dayalı Tekil Bağlantılı VCV Sonuçları .....	61
Şekil 5.2	Chebychev Uzaklığına Dayalı Tam Bağlantılı VCV Sonuçları .....	61
Şekil 5.3	Cityblock Uzaklığına Dayalı Ortalama Bağlantılı VCV Sonuçları .....	62
Şekil 5.4	Chebychev Uzaklığına Dayalı Ağırlıklı Ortalama Bağlantılı VCV Sonuçları .....	62
Şekil 5.5	Chebychev Uzaklığına Dayalı Merkezi Bağlantılı VCV Sonuçları .....	63
Şekil 5.6	Mahalanobis Uzaklığına Dayalı Medyan Bağlantılı VCV Sonuçları .....	63
Şekil 5.7	Korelasyon Benzerliğine Dayalı Ward VCV Sonuçları .....	64
Şekil 5.8	Chebychev Uzaklığına Dayalı K-Ortalamlar VCV Sonuçları .....	64
Şekil 5.9	Öklid Uzaklığına Dayalı Bulanık C-Ortalamlar VCV Sonuçları.....	65
Şekil 5.10	Minkowski Uzaklığına Dayalı Özdüzenleyici Haritalar VCV Sonuçları...65	
Şekil 5.11	Bulanık C-ortalamlar görsel doğrulama sonuçları .....	68
Şekil 5.12	3 Küme İçin Model Bazlı Bulanık C-Ortalamlar VCV Sonuçları .....	69
Şekil 5.13	3 Küme İçin Ham Bazlı Bulanık C-Ortalamlar VCV Sonuçları .....	70
Şekil 5.14	Üç Küme İçin İki Boyutlu Temel Bileşenler Analizi .....	71
Şekil 5.15	Üç Küme İçin Üç Boyutlu Temel Bileşenler Analizi .....	71

<b>Şekil 5.16</b> Küme 1 İçin Küme Profilleri .....	75
<b>Şekil 5.17</b> Küme 2 İçin Küme Profilleri .....	76
<b>Şekil 5.18</b> Küme 3 İçin Küme Profilleri .....	76
<b>Şekil 5.19</b> Kocaeli, Gölcük, Merkez Mahallesi: Gerçek, OM ve OMBBK Tahmini Öngörü Değerleri.....	.84



## TABLO LİSTESİ

<b>Tablo 5.1</b> Farklı Kümeleme Algortimaları İçin En Uygun Uzaklık Benzerlik Ölçüleri .....	59
<b>Tablo 5.2</b> En Uygun Küme Sayısı .....	67
<b>Tablo 5.3</b> Üç Küme İçin Ankara İli Kümeleme Sonuçları .....	72
<b>Tablo 5.4</b> Üç Küme İçin Bursa İli Kümeleme Sonuçları .....	72
<b>Tablo 5.5</b> Üç Küme İçin İstanbul İli Kümeleme Sonuçları .....	73
<b>Tablo 5.6</b> Üç Küme İçin İzmir İli Kümeleme Sonuçları .....	73
<b>Tablo 5.7</b> Üç Küme İçin Kocaeli İli Kümeleme Sonuçları .....	73
<b>Tablo 5.8</b> Kümelere Göre Standart Sapma Değerleri .....	74
<b>Tablo 5.9</b> Bulanık C-Ortalamalar Parametre Tahmin Değerleri .....	74
<b>Tablo 5.10</b> Kocaeli, Gölcük, Merkez Mahallesi: OM ve OMBBK Tahmini Öngörü Değerleri .....	83
<b>Tablo 5.11</b> Kocaeli, Gölcük, Merkez Mahallesi: OM ve OMBBK Tahmini Öngörü Performans Değerlendirme .....	83
<b>Tablo A.1</b> Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 1 .....	92
<b>Tablo A.2</b> Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 2 .....	93
<b>Tablo A.3</b> Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 3 .....	94
<b>Tablo A.4</b> Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 4 .....	95
<b>Tablo A.5</b> Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 5 .....	96
<b>Tablo A.6</b> Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 6 .....	97
<b>Tablo B.1</b> Sıfırıncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 1 .....	104
<b>Tablo B.2</b> Sıfırıncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 2 .....	105
<b>Tablo B.3</b> Sıfırıncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 3 .....	106

<b>Tablo B.4</b> Sıfırncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 4 .....	107
<b>Tablo B.5</b> Sıfırncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 5 .....	108
<b>Tablo B.6</b> Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 6 .....	109
<b>Tablo B.7</b> Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 7 .....	110
<b>Tablo B.8</b> Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 8 .....	111
<b>Tablo B.9</b> Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 9 .....	112
<b>Tablo B.10</b> Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 10 .....	113
<b>Tablo D.1</b> AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 1 .....	116
<b>Tablo D.2</b> AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 2 .....	117
<b>Tablo D.3</b> AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 3 .....	118
<b>Tablo D.4</b> AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 4 .....	119
<b>Tablo D.5</b> AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 5 .....	120
<b>Table G.1</b> Ocak-Aralık 2017 Dönemi Öngörü Performansları 1 .....	124
<b>Table G.2</b> Ocak-Aralık 2017 Dönemi Öngörü Performansları 2 .....	125
<b>Table G.3</b> Ocak-Aralık 2017 Dönemi Öngörü Performansları 3 .....	126
<b>Table G.4</b> Ocak-Aralık 2017 Dönemi Öngörü Performansları 4 .....	127
<b>Table G.5</b> Ocak-Aralık 2017 Dönemi Öngörü Performansları 5 .....	128



## MODEL BAZLI ZAMAN SERİLERİ KÜMELEMESİ

Metin VATANSEVER

Matematik Anabilim Dalı

Doktora Tezi

Danışman: Doç. Dr. İbrahim Demir

Bu tezin amacı, verileri birbirine benzer eğilimler gösteren zaman serileri şeklinde kümelemek için model bazlı bir zaman serisi kümeleme yaklaşımı oluşturmak ve her bir küme için kümeleri temsil eden tek bir model üretmektir. Böylece zaman serileri için kurulan uygun modellerin öngörülmesi ve bunun için gerekli işlemlerin daha etkin bir şekilde yapılabilmesi hedeflenmektedir.

Çalışmada otoregresif zaman serileri kümeleme yöntemi, tek, tam, grup ortalama, ağırlıklı ortalama, merkez, medyan, Ward hiyerarşik, K-ortalamalar, SOM, bunalık C-ortalamalar kümeleme algoritmaları ve Öklid, Minkowski, City Block, Chebyshev, Cos, Korelasyon ve Cepstral bazlı uzaklık fonksiyonlarıyla birlikte kullanılmıştır. Otoregresif model bazlı zaman serileri kümeleme yöntemi, Türkiye' nin beş büyük ilinin (İstanbul, Ankara, İzmir, Bursa and Kocaeli) 196 adet mahallesine ait aylık gayrimenkul satış fiyatları endeks verilerine uygulanmıştır. Model parametre tahmini ve öngörü performanslarını karşılaştırmak için her bir mahalleye ilişkin endeks verileri, 2010 - 2016 yılları arası eğitim, 2017 yılı test dönemi olarak seçilmiştir.

Otoregresif model parametreleri tahmin edilen zaman serileri üzerinden elde deneysel bulgulara göre, Öklid, Minkowski, City Block, Chebyshev, Cos, korelasyon ve cepstral uzaklık benzerlik fonksiyonları kullanılan tek, tam, grup ortalama, ağırlıklı ortalama, merkez, medyan, Ward hiyerarşik, K-ortalamalar, SOM; bulanık C-ortalamalar kümeleme yöntemlerine göre daha başarılı sonuçlar vermiştir. Bunun yanında, bireysel otoregresif zaman serileri model sonuçlarına kıyasla bulanık c-ortalamalar kümeleme sonuçlarının daha başarılı olduğu görülmüştür.

Literatürde, Türkiye'deki beş büyük ilin mahalleri için gayrimenkul satış fiyat değişimlerini kümeleyen ve öngöründe bulunan çalışmalar bulunmamaktadır. Bu anlamda bu tez çalışması mahalle bazlı gayrimenkul piyasaları için yapılan ilk kümeleme ve öngörü çalışmasıdır.

**Anahtar Kelimeler:** Model bazlı zaman serileri kümelemesi, Model doğrulama, Otoregresif model, Bulanık C-ortalamalar kümeleme yöntemi, Zaman serileriyle öngörü

## MODEL-BASED TIME SERIES CLUSTERING

Metin VATANSEVER

Department of Mathematics

Doctor of Philosophy Thesis

Advisor: Assoc. Prof. Dr. İbrahim DEMİR

The purpose of this dissertation is build an autoregressive a model-based time series clustering approach to cluster data into time series with similar trends and to produce a single model that represents the clusters for each cluster. Thus, it is aimed to predict the appropriate models created for the time series and to perform the necessary operations more effectively.

In this study, autoregressive time series clustering method, single, complete, group, weighted, centroid, median, Ward hierarchical, K-means, SOM and fuzzy C-means algorithms with Euclidean, Minkowski, City Block, Chebyshev, Cos, Correlation and Cepstral-based distance functions are used as clustering method. The autoregressive model-based time series clustering approach are used for monthly house sale price indices for 196 districts of 5 major cities (İstanbul, Ankara, İzmir, Bursa and Kocaeli) of Turkey for the test period: 2010M1-2016M12. In order to test the performance of forecasting, we use the validation period: 2017M1-2017M2.

The finding on the fitted autoregressive model parameters of each time series shows that fuzzy C-means clustering algorithm with Euclidean distance function

give more successful results in clustering compared to single, complete, group, weighted, centroid, median, Ward hierarchical, K-means and SOM clustering methods with Euclidean, Minkowski, City Block, Chebyshev, Cosine, Correlation and Cepstral-based distance functions. Moreover, the results of autoregressive model-based fuzzy C-means time series clustering method provides less forecasting functions with much better forecasting and similarity compared to individual autoregressive models.

In the literature, there is no previous study that clusters and forecasts house sale price indices of 196 districts of 5 major cities of Turkey. At this point, it is the first cluster and forecasting study for neighborhood-based housing markets.

**Keywords:** Model-based time series clustering, Cluster validation, Autoregressive model, Fuzzy C-means clustering, Time series forecasting

### 1.1 Literatür Özeti

Bilimin çeşitli alanlarında sıklıkla karşılaşılan veri formlarından biri olan zaman serileri, temel olarak düzenli zaman aralıkları boyunca toplanan  $y_1, y_2, \dots, y_n$  şeklinde veri noktalarıdır. Uygulamalarda benzer özellik ve eğilim sergileyen zaman serilerinin belirlenmesine, ortak stratejilerin geliştirilmesine ve öngörülerin yapılabilmesine ihtiyaç duyulmaktadır [1]-[5].

Zaman serileri kümeleme analizlerinin genel amacı, doğal grupları kesin olarak bilinmeyen zaman serilerini, önceden belirlenmiş özellik ve eğilimlere (mevsimsellik, eğilim vs.) göre gruplamaktır. Bu haliyle zaman serileri kümeleme analizleri bir denetimsiz (unsupervised) sınıflandırma yöntemidir. Zaman serileri kümeleme analizlerinin sunduğu faydaları: (i) çok sayıda zaman serisi için eş anlı olarak daha az sayıda model ile öngörü çalışmalarının yapılabilmesi (ii) veri sayısının az olduğu zaman serileri için öngörü çalışmalarının yapılabilmesi (iii) özelde finansal zaman serileri için, benzer karakteristiğe sahip olan zaman serileri belirlenerek, ortak (örneğin bölgesel) stratejilerin geliştirilebilmesi şeklinde özetlenebilir [1]-[5].

Amaca, veri tipine ve veri büyüklüğüne göre farklı kümeleme yöntemleri bulunmaktadır. Literatürde bulunan çeşitli kümeleme algoritmaları ve yazılım araçları çoğunlukla zamana göre değişim göstermeyen kesit verileri için geliştirilmiştir. Han ve arkadaşları [6]'na göre zaman serisi olmayan veri setleri için kümeleme yöntemleri beş sınıfa ayrılmaktadır: Aşamalı (hierarchical), model bazlı, yoğunluk bazlı, grid bazlı ve bölmeli (partitioning) yöntemler. Ancak önemi, farkındalığı ve boyutu sürekli artan zaman serisi verileri, bilimin farklı alanlarında giderek artan oranlarda yer bulur hale gelmekte, amaca yönelik olarak da sınıflandırılma ihtiyacı duymaktadır. Ayrıca zaman serilerinin yüksek boyutlu, karmaşık, dinamik oluşu ve veri toplanmasına ilişkin olarak sistematik hatalar içerebilmesinde dolayı Han ve arkadaşları [6]'in bahsettiği klasik kümeleme

algoritmaları zaman serileri kümeleme çalışmaları için doğrudan uygulanmayabilmektedir. Bu bağlamda zaman serilerinin kesit verilerinden farklı kümeleme yöntemleriyle incelenmesi gerekmektedir.

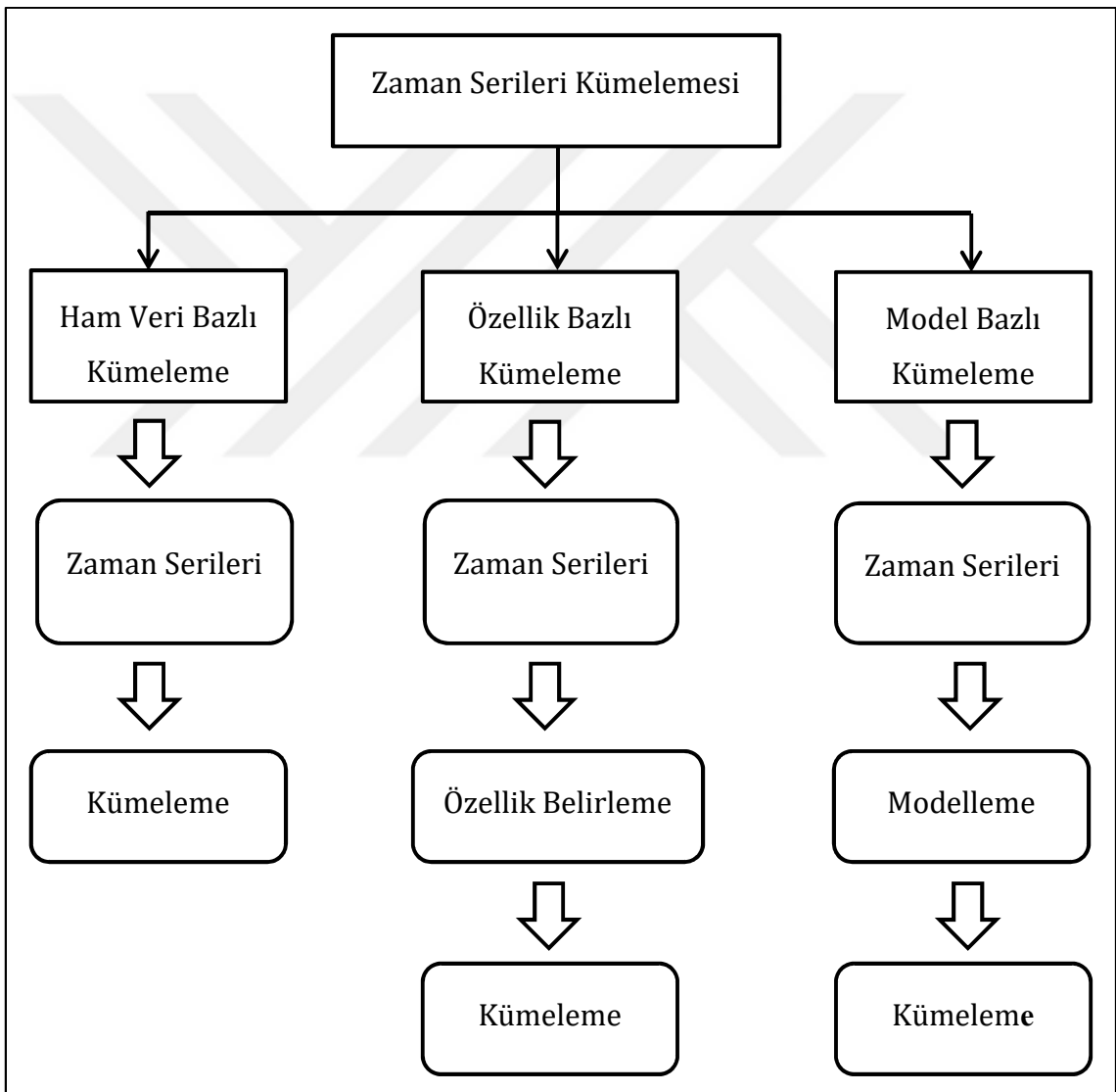
Zaman serileri çeşitli dönüştürme yöntemleri ile zamandan bağımsız olan kesit verilerine dönüştürülebilirler. Bu dönüştürme zaman serilerini temsil eden belirli özelliklerin tespiti ya da her zaman serisini tahminleyen modellerin parametre değerlerinin tespitiyle olabilmektedir. Böylece Han ve arkadaşları [6]' nin sınıflandırdığı kümeleme yöntemleri bu dönüştürülmüş, zamandan bağımsız, kesit verilerine doğrudan uygulanabilirler. Bu bağlamda, Liao [4]' ya göre zaman serileri kümeleme yöntemleri üç ana başlık altında incelenebilmektedir: Ham veri bazlı, özellik bazlı ve model bazlı. Her üç kümeleme yaklaşımı için de geleneksel uzaklık/benzerlik ölçüleri ve kümeleme algoritmaları kullanılabilir.

Ham veri bazlı zaman serileri kümeleme yöntemi, kesit verileri için geliştirilen kümeleme algoritma ve yaklaşımlarının zaman serisi verilerine doğrudan uygulanmasından oluşmaktadır. Burada, zaman serilerinin, uzaklık/benzerlik ölçüleriyle kümeleme algoritmalarının kullanabileceği uygun formata getirilmesi yeterlidir. Bu yöntem büyük miktarda zaman serileri için hesaplama gücü ve eksik değerlerle çalışamama sorunu yaratabilmektedir; ayrıca kümeleme sonuçları hatalı ve aykırı değerlere karşı hassasiyet gösterebilmektedir [4][5].

Özellik bazlı zaman serileri kümeleme yöntemi, zaman serilerini temsil eden özelliklerin (eğilim, mevsimsellik, standart sapma, ortalama, sadeleştirme vb.) belirlenerek bu belirlenen özellikler üzerinden yapılan kümeleme çalışmalarını içermektedir. Buradaki temel nokta zaman düzleminde değişkenlik gösteren zaman serilerine bir dönüştürme işlemi uygulayarak bir boyut indirilmesi yapmaktır. Bu haliyle özellik bazlı zaman serileri kümeleme yöntemleri, ham veri bazlı kümeleme yöntemlerine kıyasla hesaplama kolaylığı sağlamaktadır. Diğer taraftan bu yöntemde, zaman serilerini doğru bir şekilde temsil eden özelliklerin belirlenebilme sorunu bulunmaktadır. Farklı özelliklere göre yapılan zaman serileri kümeleme çalışmaları farklı sonuçlar verebilmektedir [4][5].

Her bir zaman serisinin kendi içinde model kalıplarının tahminlenmesi ve tahminlenen bu modellerin parametrelerinin zaman serilerini temsil eden

vektörler şeklinde ifade edilerek kümeleneşine model bazlı zaman serileri kümeleme yöntemi denilmektedir. Buradaki ana fikir, benzer zaman serilerinin belirli bir olasılık dağılımı altında benzer modellerle üretildiğı düşüncesidir. Model bazlı zaman serileri kümeleme yöntemi, ham veri bazlı kümeleme yöntemlerinin aksine hatalı, eksik ve aykırı değerlere karşı daha az duyarlıdır. Ayrıca özellik bazlı zaman serileri kümeleme yönteminde olduğı gibi zaman serilerini temsil ettiğini düşündüğümüz özelliklerin kestirim sorunu da yoktur [4][5]. Bu sebeplerden dolayı bu tez çalışmasında, model bazlı zaman serileri kümeleme yöntemi zaman serileri sınıflandırma problemi için bir yöntem olarak seçilmiştir.



**Şekil 1.1** Zaman Serileri Kümeleme Yöntemleri: Ham Veri, Özellik ve Model Bazlı

Literatürde yer alan ham veri bazlı zaman serileri kümeleme çalışmalarına örnek olarak şu çalışmaları sıralayabiliriz: Kakizawa, Shumway ve Taniguchi, Gauss dağılımına uymayan çok değişkenli zaman serilerinin kümelenebilmesi için genelleştirilmiş bir spektral uzaklık ölçüsü önermişlerdir [7]. Ayrıca, küme başarılarını ölçmek adına, ayrışma ölçüsü olarak, Kullback-Leibler ve Chernoff bilgi ölçülerini geliştirmişlerdir. Gerçek veri seti olarak deprem ve maden patlama verilerinden oluşan tarihsel çok değişkenli zaman serileri veri kümesine hiyerarşik kümeleme algoritmalarını uygulayarak bir kümeleme ve sınıflandırma çalışması yapmışlardır. Golay ve arkadaşları [8], beyin aktivitelerinin haritalanması için tek değişkenli zaman serilerini kullanarak kümeleme çalışmaları gerçekleştirmişlerdir. Kümeleme çalışmalarında, bulanık mantık temelli bir kümeleme algoritmasını Öklid uzaklığı ve ikili çapraz korelasyon benzerlik ölçüleri ile birlikte kullanmışlardır. Küme başarıları, küme içi varyans ölçümü kullanarak değerlendirilmiştir. Van ve Van [9], tarihsel enerji tüketim davranışlarını özetleyebilmek ve anlamlandırabilmek için görsel teknikler üzerinde çalışmışlardır. Görsellikle ele alınan sorun, zaman serilerinin, birden fazla zaman ölçeğinde (günler, haftalar, mevsimler) örüntülerin ve eğilimlerin nasıl belirleneceği ve ifade edileceğidir. Sundukları çözüm, benzer davranış özelliği gösteren zaman serilerini gruplamak ve gruplanmış zaman serilerinin ortalama örüntülerini takvimin ilgili günleri için grafikleştirmektir [9]. Liao ve arkadaşları [10] tek değişkenli zaman serileri için, klasik Öklid uzaklığına dayalı K-ortalamlar kümeleme yöntemini kullanmak yerine, hesaplama kolaylığı açısından 400 kat daha etkin olan biçim bazlı (shape-based) bir kümeleme yöntemi geliştirmişlerdir. Benzer şekilde, Möller ve arkadaşları [11], klasik kümeleme algoritmalarına alternatif olarak bulanık mantık temelli bir kümeleme algoritması önermişlerdir. Aynı yazarlar ayrıca, 2003 yılına kadar zaman serilerinde sıklıkla kullanılan Öklid uzaklık ve korelasyon benzerlik ölçülerine alternatif olarak, kısa dönemli zaman serilerinin daha iyi kümelenebilmesine olanak tanıyan kısa zaman serileri uzaklığı isimli bir uzaklık ölçüsü de geliştirmişlerdir [11]. Ratanamahatana ve Niennattrakul [12], video, resim ve ses kayıtlarından oluşan çok değişkenli multimedya zaman serilerinin kümelenebilmesini yapmışlardır. Çalışmalarında, klasik kümeleme algoritmalarından K-ortalamlar ve K-medoids algoritmalarını



kullanmışlardır. Onlar, diğer arařtırmacılardan farklı olarak alıřmalarında klid uzaklık lüsü yerine dinamik zaman bükmesi uzaklık lüsünü kullanmışlardır. Hepřen ve Vatansever [13], il bazlı getiri endeks verilerini kullanarak ham veri bazlı bir sınıflandırma alıřması gerekleřtirmişlerdir. Arařtırmacılar alıřmalarında tek, tam, ortalama, merkezi, medyan ağırlıklı ortalamalar baėlantılı ve Ward hiyerarřik kümeleme yöntemlerini klid uzaklığıyla beraber kullanmışlardır. Ayrıca alıřmalarında Dunn ve Silhouette sayısal küme doėrulama endekslerini görsel yöntemlerle beraber kullanarak küme sonuçlarının güvenilirliğini deėerlendirmişlerdir.

Özellik bazlı zaman serileri kümeleme alıřmalarına örnek olarak řu alıřmaları sırasıyla verebiliriz: Shaw ve King [14], fizik problemlerinde karřılařılan doėrusal olmayan zaman serilerini sınıflandırmak için hiyerarřik kümeleme yöntemlerini kullanmışlardır. Kümelemeye bařlamadan önce zaman serileri zaman düzleminde temel bileřenler analizi yöntemi kullanılarak ayrılmıř ve bu haliyle zaman serileri zamandan baėımsız bir düzlemde temsil edilerek bir boyut indirgemesi gerekleřtirilmiřtir. Owsley ve arkadaşları [15], yaptıkları titreřim verilerini sınıflandırma alıřmaları için yapay sinir aėları algoritmalarından özdüzenleyici haritaları (Self Organizing Maps - SOM) kullanmışlardır. Özdüzenleyici haritaların bulduėu özelliklerle temsil edilen titreřim verilerini Saklı Markov Modeli ile sınıflandırmışlardır. Owsley ve arkadaşları [15] alıřmalarında küme ii varyansı sınıflandırma işlemlerinin bařarısını lmek için ele almışlardır. Vlachos ve arkadaşları [16], zaman serileri verilerine wavelet dönüşümü uygulayarak, geliřtirdikleri K-ortalamalar kümeleme algoritmasıyla sınıflandırma yapmışlardır. Buldukları sonuçlarla, arařtırmacılar, klasik K-ortalamalar ve ham veri bazlı kümeleme yöntemlerine göre küme ii varyansı daha düşük küme grupları tespit etmişlerdir. Rasanen ve Kolehmainen [17], saatlik elektrik tüketim davranıřlarını segmentlere ayırarak incelemiřlerdir. Rasanen ve Kolehmainen [17] segmentasyon alıřmalarında veri madenciliėi sorunlarının üstesinden gelmek ve kümeleme alıřmalarının, doėruluktan ödün vermeden, hesaplama hızını arttırmak adına zaman serilerinin global özelliklerini simgeleyen ortalama, standart sapma, arpıklık, basıklık ve periyodik özelliklerini kullanmışlardır. Kümeleme bařarılarını Davies-Bouldin sayısal küme doėrulama sonuçlarıyla

değerlendirmişlerdir [17]. Benzer şekilde Fulcher ve Jones [18] büyük hacimli zaman serilerini gündeme almışlardır. Araştırmacılar çalışmalarında laboratuvar ortamında üretilen simülasyon zaman serilerini kullanmışlardır. Zaman serilerini karakterize edebilmek için dağılım, korelasyon, entropi, durağanlık, gibi pek çok sayıda özellik belirlemişlerdir. Bu belirlenen özelliklerin en çok bilgi vaat edenleri, doğrusal ileri özellik seçim yöntemiyle seçilmiştir. Zaman serileri, bu seçilen özelliklerle, en yakın komşuluk yöntemiyle sınıflandırılmıştır. Araştırmacılar doğru zaman serisi özelliklerinin seçilmesi ile büyük hacimli ve aynı uzunlukta olmayan zaman serilerinin, zaman etkinliği açısından verimli ve küme ayrışması açısından başarılı bir şekilde, sınıflandırılacaklarını ortaya koymuşlardır [18].

Biernacki ve arkadaşları [19] yaptıkları zaman serileri kümelemesi için Gauss karışım modellerini (Gaussian Mixture Models) ele almışlardır. Ramoni ve arkadaşları [20] çok değişkenli sensor zaman serilerinin kümelenmesi üzerine çalışmışlardır. Çalışmalarında öncelikle, çok değişkenli zaman serilerini Markov Zincirleri (Markov Chains) yardımıyla modellemişlerdir. Markov Zincirlerinin ürettiği geçiş olasılıkları hiyerarşik kümeleme algoritmalarıyla sınıflandırılarak, zaman serilerinin kümelenmesini gerçekleştirmişlerdir. Bu araştırmacılar Markov zinciri ile çok değişkenli model bazlı kümelemenin, benzerlik anlamında, tesadüfi belirlenen kümelere göre daha başarılı sonuçlar verdiğini gözlemlemişlerdir. Tran ve Wagner [21], ses doğrulama çalışmasında, Gauss modelleri ile yaptıkları bulanık C-ortalamalar kümelemeyle, yanlış sınıflamadan kaynaklı doğan hataları, klasik normalleştirme yöntemlerine göre küçültmüşlerdir. Xiong ve Yeung [22], yapay zaman serileri kümeleme çalışmalarında, ARIMA (Autoregressive Integrated Moving) yöntemlerine alternatif olarak karma ARMA (Autoregressive Moving) modellerini önermişlerdir. Araştırmacılar, maksimum olabilirlik yöntemiyle kümeleme çalışmalarını gerçekleştirmişlerdir [22]. Lin ve arkadaşları [23] gerçek zaman serileri verilerine dalgacık dönüşüm modellerini (wavelet transform models) uygulayarak K-ortalamalar kümeleme yöntemiyle sınıflandırma yapmışlardır. Corduas ve Piccolo [24], 2008 yılında ilk defa, otoregresif zaman serisi modellerine dayalı uzaklık ölçümlerinin istatistiksel özelliklerini incelemişlerdir. Ek olarak, Runkler ve Seeding [25] otoregresif model bazlı zaman serileri kümelemesi üzerine çalışmışlardır. Araştırmacılar zaman serileri

kümeleme çalışmalarıyla iki noktayı hedeflemişlerdir: (i) benzer tarihsel davranış özellikleri gösteren zaman serilerini sınıflandırmak ve (ii) sınıfları temsil eden az sayıda jenerik modellerle çok sayıda zaman serisi için genellenebilir başarılı öngörüler yapmak. Veri kümesi olarak, bir elektronik üreticisinin tarihsel ürün satış bilgileri ve bilgisayar ortamında üretilen yapay zaman serileri kullanılmıştır. Çalışmadaki temel nokta, her bir bağımsız zaman serisi için otoregresif model katsayı ve parametrelerini tahminlemek ve bu tahminlenen parametreleri bulanık C-ortalama kümeleme algoritması ile kümelemektir. Söz konusu araştırmacılar, otoregresif model bazlı bulanık C-ortalama yöntemi ile hem sınıflandırma anlamında hem de öngörü anlamında güvenilir sonuçlar elde edilebilmektedir [25]. Benzer şekilde, Güler ve Karahasan [26], otoregresif model bazlı bulanık zaman serileri kümeleme yaklaşımıyla, Türkiye' nin OECD ülkelerinden yaptığı, ülke bazlı ithalat rakamlarını öngörmeyi denemişlerdir. Ghassempour ve arkadaşları [27], sağlık sektöründe çok boyutlu veri kümesi için gizli Markov modellerini kümelemede etkin olarak kullanmışlardır. Kümelemede K-medoid' i ve küme doğrulama ölçütü olarak da Dunn endeksini kullanmışlardır. Güler ve Akkuş [28], hava kirliliğinin tahmini için yaptıkları araştırmalarında, bulanık K-medoid kümeleme algoritmasına dayalı bir bulanık zaman serisi (FTS) modeli önermişlerdir. Önerdikleri yöntemin, özellikle çok sayıda aykırı değer içeren zaman serilerinin öngörülmesinde başarılı olduğu görülmektedir [28].

## **1.2 Tezin Amacı**

Bu tezin amacı, Türkiye'nin beş büyük ilinin mahallelerinin gayrimenkul satış fiyat hareketleri için, model bazlı zaman serileri kümeleme yöntemiyle değer bazlı kümelemeden ziyade eğilim bazlı bir kümeleme yapmaktır. Çalışmada hedeflenen diğer bir nokta ise kümeleme analizi sonucu elde edilen az sayıda jenerik modellerle çok sayıda zaman serisi için genellenebilir başarılı öngörüler yapmaktır.

### 1.3 Orjinal Katkısı

Model bazlı zaman serileri kümeleme çalışmasıyla Türkiye'nin beş büyük ilinin (Anakara, Bursa, İstanbul, İzmir ve Kocaeli) mahalle bazlı gayrimenkul satış fiyat eğilimlerini kümeleyen ve öngörü yapan literatürdeki ilk çalışmadır.

Kümeleme analizlerinde, veri dağılımına bağlı olarak performanları değişebilen çeşitli kümeleme algortimaları, uzaklık/benzerlik ve küme doğrulama ölçümleri, bulunmaktadır. Zaman serisi kümelemesi üzerine literatürdeki çalışmalar incelendiğinde, bu çeşitli ölçümleri tek bir çalışmada, kümeleme kalitesini arttırıcı yönde inceleyen ve performanslarını karşılaştıran çalışmalara rastlanmamaktadır.

Bu çalışmasında, kümeleme sürecinde çeşitli uzaklık/benzerlik ölçümleri, kümeleme algoritmaları ve küme doğrulama ölçümleri karşılaştırmalı olarak incelenmektedir. Ayrıca çalışmada veri görselleştirme algoritmaları, insan görme duyu sisteminin de devreye sokulmasıyla, kümeleme sürecinde karşılaşılan sorunların tespiti ve giderilmesinde sayısal yöntemlere yardımcı olarak kullanılmaktadır. Bu bağlamda yapılan ilk çalışma olma özelliği ile de literatüre katkıda bulunmaktadır.

### 1.4 Hipotez

Otoregresif model bazlı zaman serileri kümeleme yaklaşımı ile değer bazlı kümelemeden ziyade eğilim bazlı bir kümeleme daha etkin sonuçlar verir.

Model bazlı zaman serileri kümeleme yaklaşımıyla elde edilen küme merkezlerinin, içinde buldukları kümeleri temsil eden otoregresif bir model olduğu varsayımı altında, küme elemanları olan zaman serileri için, az sayıda öngörü fonksiyonuyla çok sayıda zaman serisi için, genellenebilir, başarılı öngörüler yapılabilir.

## 2.1 Zaman Serileri Kümelemesi

Bilimin çeşitli alanlarında sıklıkla karşılaşılan veri formlarından biri olan zaman serileri temel olarak düzenli zaman aralıkları boyunca toplanan  $y_1, y_2, \dots, y_n$  şeklinde veri noktalarıdır. Uygulamalarda benzer özellik ve eğilim sergileyen zaman serilerinin belirlenmesine, ortak stratejilerin geliştirilmesine ve öngörülerin yapılabilmesine ihtiyaç duyulmaktadır [1]-[5].

Zaman serileri kümelerinin genel amacı, doğal grupları kesin olarak bilinmeyen zaman serilerini, önceden belirlenmiş özellik ve eğilimlere (mevsimsellik, eğilim vs.) göre gruplamaktır. Kümeleme işlemi sonucunda benzer özellik, mevsimsellik ve eğilimi bulunan zaman serileri aynı küme içerisinde denetimsiz olarak yer alırlar. Bu haliyle zaman serileri kümeleme analizleri bir denetimsiz (unsupervised) sınıflandırma yöntemidir.

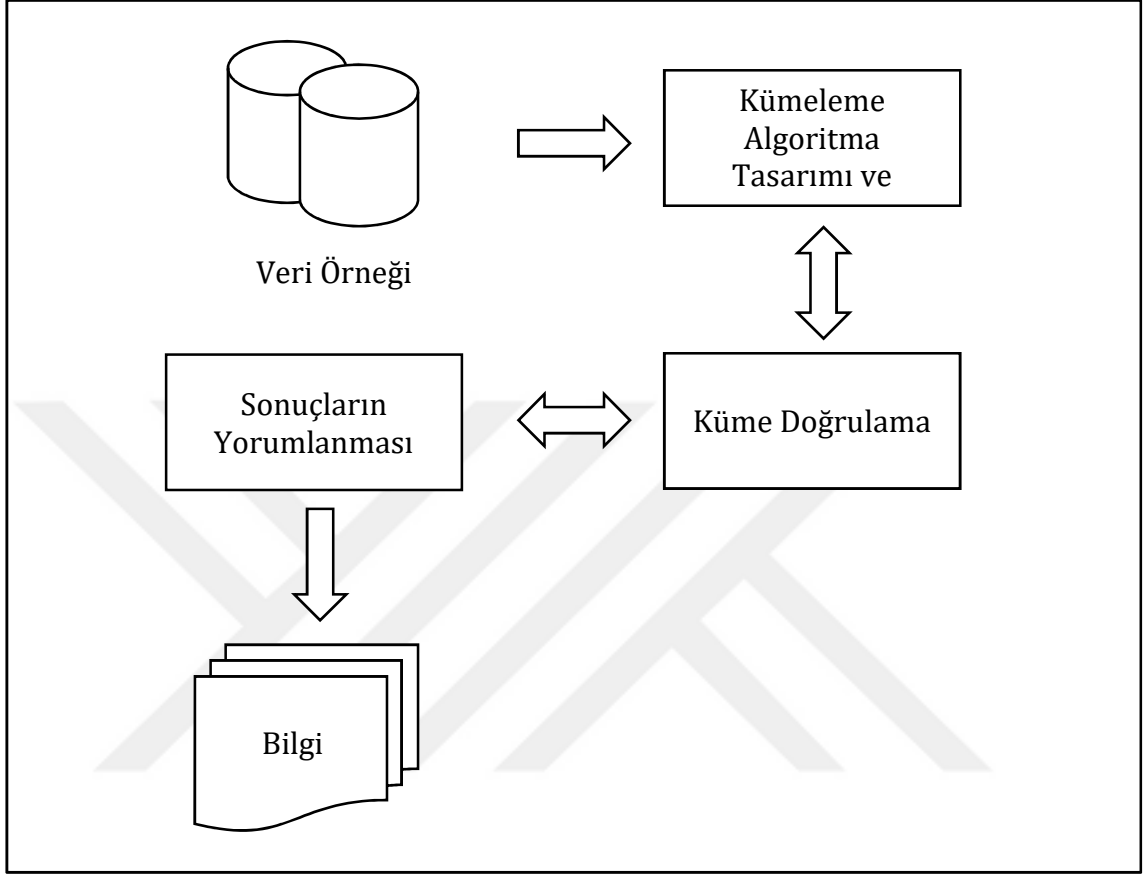
Doğal küme yapılarının tespiti veri inceleme adımının önemli bir parçası olabilmektedir. Zaman serileri kümelemesinin aşağıdaki gibi özel amaçları da bulunabilmektedir [1][5]:

1. Aykırı değerlerin tespiti,
2. Benzer eğilim ve korelasyonlu zaman serilerinin tespiti,
3. Veri indirgemesi,
4. Model uydurmanın kolaylaştırılması,
5. Tahmin,
6. Çok sayıda zaman serisi için eş anlı olarak daha az sayıda model ile öngörü,
7. Veri sayısının az olduğu zaman serileri için öngörü,
8. Benzer karakteristiğe sahip olan zaman serileri belirlenerek, ortak stratejilerin geliştirilebilmesi.

Bu bölümde, kümeleme ve zaman serileri kümelemeleriyle ilgili temel bilgiler verilecektir.

## 2.2 Kümeleme Süreci

Kümeleme analizi algoritmasa seçimi, küme doğrulama ve sonuçların yorumlanması olmak üzere üç adımdan oluşmaktadır [2][3].



Şekil 2.1 Kümeleme Süreci [2]

Şekil 2.1, kümeleme sürecine ilişkin adımları akış diyagramı şeklinde göstermektedir. Unutulmamalıdır ki, kümeleme analizi tek döngülü bir süreç değil, tatmine edici sonuçlara erişilene kadar devam eden bir süreçtir [2][3].

## 2.3 Uzaklık ve Benzerlik Ölçüleri

Kümeleme işlemi, birimlerin birbirilerine olan benzerlikleri ya da uzaklıkları kullanılarak gerçekleştirilmektedir. Temel olarak,  $x_i$  ve  $x_j$  iki veri gözlem vektörü arasındaki benzemezliği sayısal değerle ifade eden ölçüye uzaklık ölçüsü; diğer taraftan benzerliği sayısal değerle ifade eden ölçüye benzerlik ölçüsü denilmektedir. Uzaklık ve benzerlik ölçülerine yakınlık ölçüleri de denilmektedir.

Söz konusu ölçeklerin üst sınırı bulunmamakta, ölçeğe bağlı olarak değişmektedir [29][30].

Uzaklık ölçülerinde mutlak yönlü pozitif bir artış birimler arasındaki benzerliğin azalışını, azalış da birimler arasındaki benzerliğin artışını göstermektedir. Benzerlik ölçülerinde ise mutlak yönlü pozitif bir artış birimler arasındaki benzerliğin artışını, azalış da birimler arasındaki benzerliğini azalışını ifade eder [3][29].

Literatürde, değişkenlerin kesikli ya da sürekli olmalarına göre çeşitli uzaklık ve benzerlik ölçüleri bulunmaktadır. Bu çalışmada, sürekli değişkenlerle çalışılacağı için sürekli değişkenler için kullanılan uzaklık ve benzerlik ölçüleri incelenecektir [1].

### 2.3.1 Uzaklık Ölçüleri

#### 2.3.1.1 Öklid Uzaklığı

Her  $x_i$  ve  $x_j$  gözlem vektörü için Öklid uzaklığı

$$d(x_i, x_j) = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.1)$$

ile hesaplanır.

#### 2.3.1.2 Minkowski Uzaklığı

Her  $x_i$  ve  $x_j$  gözlem vektörü için Minkowski uzaklığı

$$d(x_i, x_j) = \left\{ \lambda \sum_{k=1}^p |x_{ik} - x_{jk}|^\lambda \right\}^{\frac{1}{\lambda}}, \quad \lambda \geq 0 \quad (2.2)$$

şeklinde hesaplanır.

Farklı  $\lambda$  değerleri için farklı uzaklık ölçüleri elde edilir. Örneğin,  $\lambda = 1$  için Manhattan uzaklık ölçüsü,  $\lambda = 2$  için Öklid uzaklık ölçüsü elde edilir.

### 2.3.1.3 City Block Uzaklığı

City Block uzaklığına, “Manhattan uzaklığı” da denir.  $\lambda = 1$  için Minkowski, City Block uzaklığına dönüşür.

### 2.3.1.4 Chebyshev Uzaklığı

Her  $x_i$  ve  $x_j$  gözlem vektörü için Chebyshev uzaklığı

$$d(x_i, x_j) = \max_k |x_{ik} - x_{jk}| \quad (2.3)$$

ile verilir.

### 2.3.1.5 Cepstral Bazlı Uzaklık

Cepstral bazlı uzaklık, literatürde linear predictive coding (LPC) olarak da bilinir. Buradaki temel mantık iki otoregresif (AR) zaman serisi arasındaki mesafeyi tespit etmektir.

$X_t, Y_t \in \mathcal{L}$ , AR sürecinde tanımlanan iki zaman serisi ve  $p$  gecikme sayısı olmak üzere,  $X_t$  ve  $Y_t$  sırasıyla

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t \quad (2.4)$$

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (2.5)$$

olarak verilir.

AR katsayıları

$$\phi_x = (\phi_{1,x}, \phi_{2,x}, \dots, \phi_{i,x}, \dots); \quad \phi_y = (\phi_{1,y}, \phi_{2,y}, \dots, \phi_{i,y}, \dots) \quad (2.6)$$

ile bulunur.

Cepstral bazlı uzaklığın hesaplanabilmesi için girdi olarak kullanılan Cepstral katsayısı, AR süreçli bir zaman serisi için

$$c_n = \begin{cases} -\phi_1 & n = 1 \\ -\phi_n - \sum_{m=1}^{n-1} \left(1 - \frac{m}{n}\right) \phi_m c_{n-m} & 1 < n \leq p \\ -\sum_{m=1}^p \left(1 - \frac{m}{n}\right) \phi_m c_{n-m} & p < n \end{cases} \quad (2.7)$$

ile tanımlanır.



İki AR modeli arasındaki Cepstral bazlı uzaklık, Cepstral katsayıları arasındaki uzaklığın belirlenmesiyle bulunur. Cepstral katsayıları arasındaki uzaklığı belirlemek için yukarıda bahsedilen Öklid, Minkowski, City Block ve Chebyshev uzaklık ölçümlerinden herhangi bir tanesi tercih edilebilir.

$X_t$  ve  $Y_t$ , AR süreçli zaman serileri için Öklid uzaklığı kullanılarak hesaplanan Cepstral bazlı uzaklık

$$d_{\text{ceps}} = \{\sum_{t=1}^T (c_t^X - c_t^Y)^2\}^{1/2} \quad (2.8)$$

ifadesiyle verilir.

### 2.3.2 Benzerlik Ölçüleri

#### 2.3.2.1 Açısal Benzerlik Ölçüsü (Cosine Similarity Measure)

Açısal benzerlik ölçüsü iki gözlem vektörü arasındaki açının kosinüsüdür. Her  $x_i$  ve  $x_j$  gözlem vektörü için açısal benzerlik ölçüsü

$$s_{ij} = \frac{\sum_{i=1}^p x_i x_j}{\sqrt{\sum_{i=1}^p x_i^2} \sqrt{\sum_{i=1}^p x_j^2}} \quad (2.9)$$

olarak verilir.

#### 2.3.2.2 Korelasyon Benzerlik Ölçüsü (Correlation Similarity Measure)

Korelasyon benzerlik ölçüsü, iki gözlem vektörünün birbirine olan benzerlik derecesini gösterir. Korelasyon benzerlik ölçüsü  $[-1, +1]$  arasında değer alır. Her  $x_i$  ve  $x_j$  gözlem vektörü için korelasyon benzerlik ölçüsü aşağıdaki gibi hesaplanır.

$$s_{ij} = \frac{\sum_{i=1}^p (x_i - \mu_x)(y_i - \mu_y)}{s_x s_y}, \quad \mu_x = \frac{1}{p} \sum_{i=1}^p x_i, \quad s_x = \sqrt{\sum_{i=1}^p (x_i - \mu_x)^2} \quad (2.10)$$

### 2.3.3 Uzaklık ve Benzerlik Ölçümlerini Birbirine Dönüşümü

Çeşitli dönüşüm yöntemleriyle benzerlik ölçüleri, uzaklık ölçülerine dönüştürülebilir [3][23].

$$d_{ij} = 1 - s_{ij} \quad (2.11)$$

veya c sabit terimiyle

$$d_{ij} = c - s_{ij} \quad (2.12)$$

$$d_{ij} = \sqrt{s_{ii} - 2s_{ij} + s_{jj}} \quad (2.13)$$

eğer  $s_{ii} = s_{jj} = 1$  ise

$$d_{ij} = \sqrt{2(1 - s_{ij})} \quad (2.14)$$

Tersine, uzaklık ölçüleri benzerlik ölçülerine de

$$s_{ij} = (1 + d_{ij})^{-1} \quad (2.15)$$

şeklinde dönüştürülebilir [3][29].

## 2.4 Zaman Serileri Kümeleme Algoritmaları

Uygulamanın amacına, verinin dağılımına ve büyüklüğüne bağlı olarak farklı kümeleme yöntemleri bulunmaktadır. Değişik kaynaklarda farklı kümeleme yöntemleri farklı şekillerde sınıflandırılabilir. Kümeleme yöntemleri aşamalı (hiyerarşik) ve aşamalı olmayan kümeleme yöntemleri olarak ikiye ayrılmaktadır. Bu alt bölümde, yaygın olarak kullanılan bu algoritmalara yer verilecektir.

### 2.4.1 Hiyerarşik Kümeleme Yöntemleri (Hierarchical Cluster Methods)

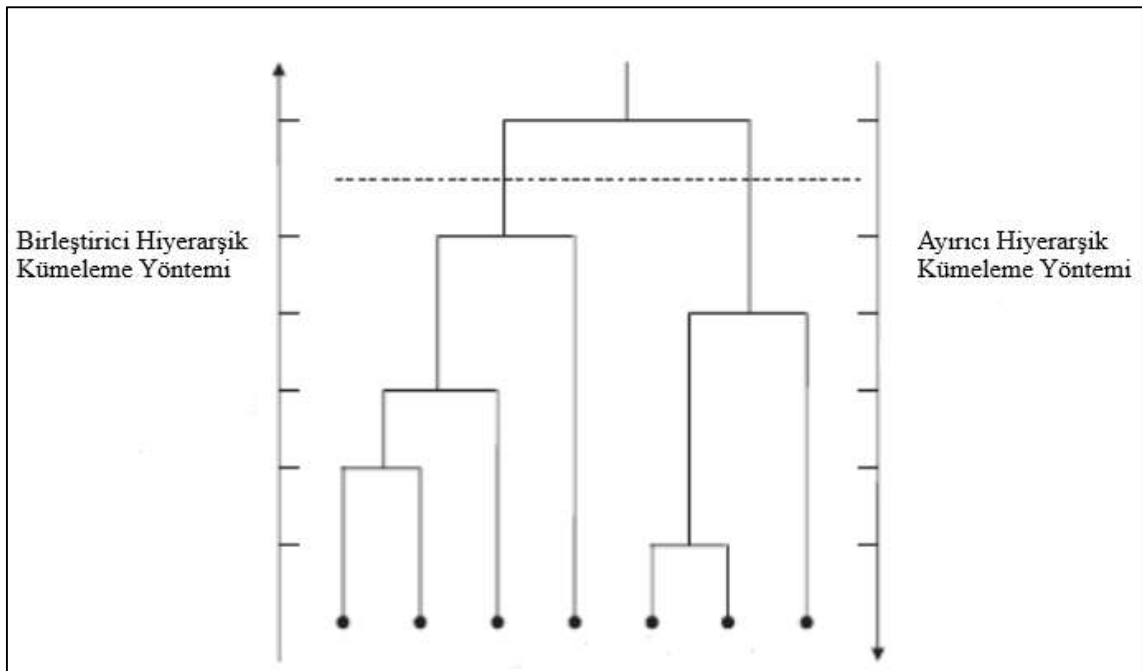
Hiyerarşik kümeleme yöntemleri, veri setinde yer alan birimlerin birbirlerine olan uzaklık ve benzerliklerini dikkate alarak, birimleri bir küme ağacı üzerinde aşamalı olarak bir araya getirir. Dendogram adı verilen bu ağaç yapısı hiyerarşik kümeleme sürecini temsil eder. Dendogram veri birimlerinin adım adım nasıl gruplandığını gösterir. Bu temsili gösterim, potansiyel küme yapılarını keşfetmemize ve kümeleme sürecini görselleştirmemize de olanak tanır. Dendogramların farklı uzaklık ve benzerlik seviyelerinde kesilmesiyle küme sonuçları elde edilir [3][4].

Hiyerarşik yapının aşağıdan yukarıya veya yukarıdan aşağıya kurulmasına göre birleştirici (agglomerative) veya ayırıcı (divisive) olmak üzere iki tipte hiyerarşik kümeleme yöntemi bulunmaktadır [3][6]. Birleştirici hiyerarşik kümeleme yöntemleri, başlangıçta tüm birimleri tek başlarına ayrı birer küme olarak

düşünerek, n birimi aşamalı olarak tek bir küme kalana kadar birleştirir. Ayrırcı hiyerarşik kümeleme yöntemleri ise birleştirici hiyerarşik kümeleme yöntemlerinin aksine, başlangıçta tüm birimleri tek bir küme olarak düşünerek, n birimi aşamalı olarak n tane küme elde edilene kadar ayırır [6]. Şekil 2.2, sekiz gözlem için birleştirici ve ayrırcı hiyerarşik kümeleme yöntemleri için çizilen bir dendogram örneğini göstermektedir. Okların yönüne göre dendogram üzerinden birleştirici ve ayrırcı hiyerarşik kümeleme yönteminin kümeleme adımları takip edilebilir.

Birleştirici hiyerarşik kümeleme yöntemini genel olarak aşağıdaki gibi özetlenebilir [3]:

1. Veri kümesindeki n tane gözlem küme olarak düşünülür. Her n küme için uzaklık/benzerlik matrisi hesaplanır.
2. Kümeler arasındaki minimum uzaklık veya maksimum benzerlik belirlenir. En yakın kümeler birleştirilir.
3. Yeni belirlenen küme ile diğer kümeler arasındaki uzaklık ve benzerlik matrisi hesaplanır.
4. 2. ve 3. adım tek bir küme kalıncaya dek tekrarlanır.



Şekil 2.2 Hiyerarşik Kümeleme Diyagramı [3]

Hiyerarşik kümeleme yöntemleri eşit uzunlukta zaman serilerine uygulanabildiği gibi eşit uzunlukta olmayan zaman serilerine doğru uzaklık benzerlik ölçümünün seçilmesi suretiyle de uygulanabilmektedir [4].

Birleştirici hiyerarşik kümeleme yöntemleri bölmeli hiyerarşik kümeleme yöntemlerine göre hesaplama işlemleri açısından daha etkin algoritmalarıdır. Bundan dolayı birleştirici hiyerarşik kümeleme yöntemleri en sık kullanılan hiyerarşik kümeleme yöntemleridir [3]. Bu tezde en bilinen birleştirici hiyerarşik kümeleme yöntemlerinden tek bağlantılı (single link) tam bağlantılı (complete link), grup ortalama bağlantılı (group average linkage), ağırlıklı ortalama bağlantılı (weighted average linkage), medyan bağlantılı (median linkage), merkezi bağlantılı (centroid linkage) ve Ward yöntemlerinden bahsedilecektir.

#### **2.4.1.1 Tek Bağlantılı (Single Linkage) Hiyerarşik Kümeleme Yöntemi**

Bilinen en basit hiyerarşik kümeleme yöntemidir. Tek bağlantılı kümeleme yöntemi, minimum uzaklıktaki birimleri veya kümeleri hiyerarşik olarak birleştirir. Minimum uzaklıklar dikkate aldığı için tek bağlantılı kümeleme yöntemine en yakın komşuluk yöntemi de denilmektedir [3][31].

$I$ . kümenin daha önce oluşan  $i$ . ve  $j$ . kümelerden hangisi ile birleşeceğini belirlemek için  $I$ . küme ile  $i$ . ve  $j$ . kümelerin uzaklıklarına bakılır. Bu uzaklıklardan en küçük olanı ile birleştirme işlemi yapılarak  $I$ . küme elde edilir.  $I$ . kümenin  $i$ . ve  $j$ . kümeye olan uzaklık fonksiyonu [3][31]

$$D(C_l, (C_i, C_j)) = \min(D(C_l, C_i), D(C_l, C_j)) \quad (2.16)$$

ile bulunur.

Tek bağlantılı kümeleme yöntemi eliptik olmayan küme yapılarını belirlemede başarılıdır [3][31].

#### **2.4.1.2 Tam Bağlantılı (Complete Linkage) Hiyerarşik Kümeleme Yöntemi**

Tam bağlantılı kümeleme yöntemi, tek bağlantılı kümeleme yönteminin aksine, veri birimlerini veya kümelerine, aralarındaki maksimum uzaklığa göre, hiyerarşik olarak birleştirir. Maksimum uzaklıklar dikkate alındığı için tam bağlantılı kümeleme yöntemine en uzak komşuluk yöntemi de denilmektedir [3][31].

$I$ . kümenin daha önce oluşan  $i$ . ve  $j$ . kümelerden hangisi ile birleşeceğini belirlemek için  $I$ . küme ile  $i$ . ve  $j$ . kümelerin uzaklıklarına bakılır. Bu uzaklıklardan en büyük olanı ile birleştirme işlemi yapılarak  $I$ . küme elde edilir.  $I$ . kümenin  $i$ . ve  $j$ . kümeye olan uzaklık fonksiyonu [3][31]

$$D(C_l, (C_i, C_j)) = \max(D(C_l, C_i), d(C_l, C_j)) \quad (2.17)$$

biçiminde bulunur.

Tam bağlantılı kümeleme yöntemi, gürültü ve uç değerlere (outliers) karşı tek bağlantılı kümeleme yöntemine göre daha az hassas olmasının yanında küçük ve kompakt küme yapılarının belirlenmesinde daha başarılıdır [3][31].

#### 2.4.1.3 Grup Ortalama Bağlantılı (Group Average Linkage) Hiyerarşik Kümeleme Yöntemi

Grup ortalama bağlantılı kümeleme yöntemi, kümeleri oluşturan birimlerin farklı kümelerdeki birimlere olan uzaklıklarının ortalamasını kullanarak, birim veya kümeleri hiyerarşik olarak birleştirir. Elde edilen yeni  $I$ . küme ile birleştirilen eski küme arasındaki uzaklık [3][31]

$$D(C_l, (C_i, C_j)) = \frac{1}{2}(D(C_l, C_i) + d(C_l, C_j)) \quad (2.18)$$

biçiminde bulunur.

#### 2.4.1.4 Ağırlıklı Ortalama Bağlantılı (Weighted Average Linkage) Hiyerarşik Kümeleme Yöntemi

Ağırlıklı bağlantılı kümeleme yöntemi, grup ortalama bağlantılı kümeleme yöntemine benzemektedir. Her iki yöntem de ortalama uzaklık fonksiyonunu kullanmaktadır. Aradaki tek fark ağırlıklı ortalama bağlantılı kümeleme yöntemi kümeleri oluşturan birim sayısına göre ortalama uzaklık fonksiyonunu ağırlıklandırmaktadır. Elde edilen  $I$ . küme ile  $i$ . ve  $j$ . kümeler arasındaki ağırlıklı ortalama bağlantılı uzaklık [3][31]

$$D(C_l, (C_i, C_j)) = \frac{n_i}{n_i+n_j}D(C_l, C_i) + \frac{n_j}{n_i+n_j}D(C_l, C_j) \quad (2.19)$$

fonksiyonu ile hesaplanır.

#### 2.4.1.5 Merkez Bağlantılı (The Centroid Linkage) Hiyerarşik Kümeleme Yöntemi

Merkezi bağlantılı kümeleme yöntemi, küme merkezleri arasındaki uzaklığı dikkate alarak küme veya birimleri birleştirir. En genel olarak küme merkezleri, küme ortalamaları olarak ifade edilir. Elde edilen  $I$ . küme ile  $i$ . ve  $j$ . kümeler arasındaki merkezi bağlantılı mesafe [3][31]

$$D(C_l, (C_i, C_j)) = \frac{n_i}{n_i+n_j} D(C_l, C_i) + \frac{n_j}{n_i+n_j} D(C_l, C_j) - \frac{n_i n_j}{(n_i+n_j)^2} D(C_i, C_j) \quad (2.20)$$

ile bulunur.

#### 2.4.1.6 Medyan Bağlantılı (Median Linkage) Hiyerarşik Kümeleme Yöntemi

Medyan bağlantılı kümeleme yöntemi merkezi bağlantılı kümeleme yöntemine benzemektedir. Aradaki tek fark, merkezi bağlantılı kümeleme yöntemi kümeleri oluşturan birimlerin ortalamasını küme merkezi olarak kabul ederken, medyan bağlantılı kümeleme yöntemi kümeleri oluşturan birimlerin medyan değerlerini küme merkezi olarak kabul etmektedir. Elde edilen  $I$ . küme ile  $i$ . ve  $j$ . kümeler arasındaki medyan bağlantılı uzaklık [3][31]

$$D(C_l, (C_i, C_j)) = \frac{1}{2} D(C_l, C_i) + \frac{1}{2} D(C_l, C_j) - \frac{1}{4} D(C_i, C_j) \quad (2.21)$$

olarak bulunur.

#### 2.4.1.7 Ward (Ward's Method) Hiyerarşik Kümeleme Yöntemi

Ward kümeleme yöntemi birleştirilen iki küme arasındaki hata kareleri minimize etmeyi amaç edinen bir yöntemdir. Elde edilen  $I$ . küme ile  $i$ . ve  $j$ . kümeler arasındaki Ward's bağlantılı uzaklık [3][31]

$$D(C_l, (C_i, C_j)) = \frac{n_i+n_l}{n_i+n_j+n_l} D(C_l, C_i) + \frac{n_j+n_l}{n_i+n_j+n_l} D(C_l, C_j) - \frac{n_l}{(n_i+n_j)^2} D(C_i, C_j) \quad (2.22)$$

ile hesaplanır.

## 2.4.2 Bölmeli Kümeleme Yöntemleri (Partition Clustering Method)

Bölmeli kümeleme yöntemleri, belirli bir kriter fonksiyonun optimizasyonuna dayanılarak kümeleme işlemini gerçekleştirmektedirler. Kümeler tarafsız bir bölme kriterine göre oluşturuldukları için aynı kümedeki birimler birbirine benzerken, farklı kümedeki birimlerden farklıdırlar. Hiyerarşik kümeleme yöntemlerine kıyasla bölmeli kümeleme yöntemleri büyük veri setlerine de uygulanabilmektedir [3][31].

Bölmeli kümeleme yöntemleriyle elde edilen  $k$  adet küme aşağıdaki koşulları sağlar:

1. Her bir küme en azından bir adet veri birimi içerir.
2. Her bir veri birimi yalnızca bir kümeye atanabilir.

Bulanık C- ortalamalar kümeleme yönteminde bu ikinci koşul sağlanmaz.

### 2.4.2.1 K-Ortalamlar (K-Means) Kümeleme Yöntemi

Bölmeli yöntemler arasında en yaygın kullanılan, K-ortalamlar kümeleme yöntemidir. Temel olarak, K-ortalamlar kümeleme yöntemi,  $p$  değişkenli veri birimlerini, küme içi kareler toplamını minimize edecek şekilde  $K$  adet kümeye ayırmayı amaçlar. Birimlerin kümelere ayrılması iteratif bir şekilde gerçekleşir. Veri birimleri her iterasyonda farklı kümelere atanarak en uygun çözüm permutasyonel bir şekilde gerçekleşir [3][29][31].

K-ortalamlar kümeleme yöntemi genel olarak aşağıdaki adımlardan oluşmaktadır [3][29][31]:

1.  $k$  adet küme sayısı belirlenir.
2.  $k$  adet kümenin başlangıç küme merkezleri rastgele olarak belirlenir.
3. Veri birimlerinin küme merkezlerine olan uzaklıkları belirlenen uzaklık veya benzerlik ölçülerine göre hesaplanır.
4. Birimler belirlenen uzaklık veya benzerlik ölçülerine göre en yakın oldukları küme merkezlerinin içinde bulunduğu kümelere atanır.
5. Küme merkezleri küme birimlerinin ortalamaları alınarak güncellenir.
6. Birimlerin, 4. ve 5. adımdaki küme merkezlerine olan uzaklıkları karşılaştırılır.

7. Uzaklık kabul edilebilir oranda azalmış ise 5. adıma dönülür.
8. Eğer büyük ve anlamlı bir değişiklik olmamışsa iterasyon sona erdirilir.

K-ortalamlar kümeleme yönteminde, genel olarak, iterasyonun sonlandırılmasında hata kareler toplamı bir optimizasyon yaklaşımı olarak kullanılır. Bu yaklaşımla, belirlenen uzaklık veya benzerlik ölçüsüne göre her birimin en yakın olduğu küme merkezlerine olan kareli toplam uzaklıkları (SSE: Sum of Square Error) minimize edilmeye çalışılır. SSE' lerin küçük olması küme merkezlerinin kümeleri ne derece temsil ettiğinin bir göstergesi olarak kullanılır. SSE' ler küçüldükçe küme merkezlerinin kümeleri temsil etme başarıları artmaktadır [1].

Kümeleme sürecinde, Öklid uzaklığı kullanılırsa SSE

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} D(c_i, x) = \sum_{i=1}^k \sum_{x \in C_i} (c_i - x)^2 \quad (2.23)$$

$$c_i = \frac{1}{n_i} \sum_{x \in C_i} x \quad (2.24)$$

şeklinde ifade edilir. Burada  $c_i$  i. kümeyi,  $x$  i. kümedeki bir birimi ve  $n_i$  i. kümedeki birim sayısını  $c_i$  i. kümenin ortalamasını ve  $k$  küme sayısını gösterir.

**İspat:**

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} D(c_i, x) = \sum_{i=1}^k \sum_{x \in C_i} (c_i - x)^2 \quad (2.25)$$

$$\frac{\partial}{\partial c_k} SSE = \frac{\partial}{\partial c_k} \sum_{i=1}^k \sum_{x \in C_i} (c_i - x)^2$$

$$= \sum_{i=1}^k \sum_{x \in C_i} \frac{\partial}{\partial c_k} (c_i - x)^2$$

$$= \sum_{x \in C_k} 2 * (c_k - x_k) = 0$$

$$\sum_{x \in C_k} 2 * (c_k - x_k) = 0 \Rightarrow \sum_{x \in C_k} x_k \Rightarrow c_k = \frac{1}{n_k} \sum_{x \in C_k} x_k$$

Özellikle, büyük miktarda zaman serileri kümelemelerinde hesaplama basitliğinden dolayı K-ortalamlar kümeleme yöntemi tercih edilmektedir. Ancak yüksek boyutlu veri setlerinde, iteratif hesaplama yaklaşımından ötürü K-ortalamlar



kümeleme yöntemi etkinliğini kaybedebilmektedirler. Ayrıca K-ortalamlar kümeleme yönteminin başarısı, büyük oranda rastgele belirlenen küme merkezlerinin seçimine bağlıdır. Diğer bir nokta ise K-ortalamlar kümeleme yöntemi, verilen k adet küme sayısına göre kümeleme işlemi yapmaktadır. Eğer, seçilen k adet küme sayısı doğal küme sayısını temsil edecek şekilde seçilmediyse, kümelerin kalitesi etkilenecektir [2]-[4].

#### **2.4.2.2 Özdüzenleyici Haritalar (Self Organizing Maps - SOM) Kümeleme Yöntemi**

İstatistik tabanlı kümeleme algoritmalarına alternatif olarak yapay sinir ağı algoritmaları da kümeleme problemlerinde kullanılabilir. İlk defa, Kohonen [32] tarafından 1982 yılında geliştirilen özdüzenleyici haritalar (Self Organizing Maps-SOM) kümeleme çalışmalarında en yaygın kullanılan yapay sinir ağı algoritmasıdır.

SOM denetimsiz öğrenme mantığı ile çalışan bir algoritmadır. Bundan dolayı denetimli öğrenme (supervised learning) algoritmalarının aksine, SOM yapay sinir ağlarında, olayları öğrenmek için bir öğretmene veya ağın üretmesi gereken çıktıların ağa öğretilmesine gerek yoktur. Bu özelliğinden dolayı SOM ağları kümeleme analizlerinde kullanılabilir [2][3][32].

SOM ağları vasıtasıyla, veri setindeki birimler benzer özelliklerine göre hem kümelenebilir hem de görsel olarak haritalandırılabilirler. Bu özelliğinden dolayı, SOM ağları, klasik K-ortalamlar kümeleme yöntemi ile çok boyutlu ölçekleme yöntemlerinin her ikisinin işlevlerini aynı anda yerine getirebilmektedir. SOM ağları hem verilerin kümelenebilmesinde hem de görselleştirme aracı olarak çeşitli uygulamalarda tercih edilmektedir [2][3][32][33].

Basit bir ifade ile SOM ağları tek katmanlı bir yapay sinir ağı olup giriş ve çıkış nöronlarından oluşmaktadır. Veri setindeki değişken sayısı giriş nöronlarının sayısını belirlerken, çıkış nöronları ise bulunmak istenen kümeleri temsil etmektedir. Çıkış nöronları arasında doğrudan bir bağlantı yoktur [32][33].

Diğer yapay sinir ağlarına kıyasla SOM ağlarında, çıkış katmanındaki nöronların dizilimi analiz sonuçlarını doğrudan etkilemektedir. Bu nöron dizilimleri doğrusal, dikdörtgen, altıgen veya küp şeklinde olabilir. En çok tercih edilen nöron

dizilimi karesel ve altıgen şeklindedir. Burada bahsedilen dizilim topolojik komşuluk açısından önemlidir [32][33]. Şekil 2.5, temsili olarak iki boyutlu düzlemdeki altıgen ve karesel nöron dizilimlerini göstermektedir.

Giriş nöronları ile her bir çıkış nöronu arasındaki bağlantıyı referans vektörleri (code-book vectors) sağlar. Bu vektörler bir katsayılar matrisinin sütunlarını temsili olarak göstermektedir. SOM yapay sinir ağları eğitilirken topolojik komşuluk yapısı referans vektörlerinin güncellenmesinde kullanılır.

SOM kümeleme yöntemi genel olarak aşağıdaki adımlardan oluşmaktadır [2][3]:

1. SOM çıkış , topolojik yapısının ve başlangıç referans vektörlerinin  $m_i(0)$ ,  $i = 1, \dots, K$  belirlenmesi.
2. X veri deseni için, topolojik komşuluk için j kazanan nöronun J komşuluğundaki nöronları bulunur.

$$J = \arg_i \min \{ \|x - m_j\| \} \quad (2.26)$$

3. t. iterasyon için referans vektörleri aşağıdaki kurala göre güncellenir:

$$m_i(t+1) = m_i(t) + h_{wi}(t)[x - m_i(t)] \quad (2.27)$$

$h_{wi}(t)$  komşuluk fonksiyonu olup

$$h_{wi}(t) = \alpha(t) \exp\left(\frac{-\|r_w - r_i\|^2}{2\sigma^2(t)}\right) \quad (2.28)$$

ile ifade edilir.

Burada,  $\alpha$  sürekli olarak azalan öğrenme katsayısını ve r çıkış katmanındaki topolojik nöron dizilimindeki ilgili nöronların konumunu ifade eder.  $\sigma$  ise sürekli olarak azalan komşuluk alanının genişliğini göstermektedir.

$$h_{wi}(t) = \begin{cases} \alpha(t), & \text{eğer } w \text{ nöronu kazanan } J \text{ nöronunun komşuluğuna aitse} \\ 0, & \text{aksi takdirde} \end{cases}$$

4. Nöron pozisyonlarında değişim görülemeyene kadar 2. ve 3. adımlar tekrarlanır.

Algoritma adımlarından da anlaşılacağı üzere ilk önce referans vektörlerine bir değer atanır. Bu atama işlemi çoğunlukla rastgeledir. Öğrenme döngüsüne başlamadan önce öğrenme katsayısı ( $\alpha$ ) ve topolojik komşuluk değişkenine ( $R$ ) yüksek bir değer atanır.  $\alpha$ , 0 ile 1 arasında değer alabilir ve başlangıç değerinin 1'e yakın olması tercih edilir [1]-[3].

SOM ağları yarışmacı ağlardır. Kazanan nöron ve komşu nöronların referans vektörleri sürekli güncellenir. Öğrenme sürecinde kazanan nöronu bulmak için minimum Öklid uzaklığının kullanılabilmesi gibi kümeleme analizlerinde kullanılan diğer uzaklık ve benzerlik ölçümleri de kullanılabilir [1]-[3].

Eğitim süreci boyunca yani döngü devam ettiği müddetçe referans vektörlerinin güncellenmesi devam eder. Güncellenen birimlerin referans vektörleri giriş desenine her iterasyonda bir miktar daha yaklaşmaktadır. Genel olarak,  $h_{wi}(t)$  komşuluk fonksiyonu tek tepeli bir fonksiyon olup kazanan birimin bulunduğu yerin çevresinde simetrik ve kazananla uzaklaştıkça sürekli azalan bir yapıdadır [1]-[3].

Referans vektörlerinin değişim hızı öğrenme oranı denilen  $\alpha(t)$  ile belirlenir ve bu oran zamanla azalarak en sonunda 0 değerine yakınsar. Etkileşime dahil edilecek birimler, komşuluk fonksiyonu denilen  $h_{wi}(t)$  ile belirlenir. Etkileşime dahil edilen bu birimlerin sayısı da zamanla azalır ve eğitim işleminin sonuna doğru sadece kazanan birim etkileşime girer [1]-[3].

Referans vektörlerinin hareketiyle giriş deseni ve referans vektörleri arasındaki uzaklık fonksiyonu sürekli azalarak referans vektörleri giriş desenine benzer hale gelir. Kazanan nöronun sonraki iterasyonlarda kazanma olasılığı artmaktadır. Sadece kazanan nöron değil bu nörona komşu diğer nöronların da kazananla birlikte etkileşime girmesi sebebiyle birbirine benzer desenlerin uzaysal kümelenmesi sağlanmaktadır. Sonuç olarak,  $p$  boyutlu bir veri setinde, benzer özellikler sergileyen veri birimleri özdüzenleyici haritaların iki boyutlu çıkış uzayında komşu olarak konumlandırılmaktadır [1]-[3].

### 2.4.2.3 Bulanık C-Ortalamlar (Fuzzy C-Means - FCM) Kümeleme Yöntemi

Bulanık C-ortalamlar kümeleme yöntemi (FCM) bulanık mantığa dayanan kümeleme algoritmalarının en popülerlerinden bir tanesidir. Bulanık C-ortalamlar kümeleme yöntemi ilk defa Dunn [34] tarafından ortaya atılmış, daha sonra Bezdek [35] tarafından geliştirilmiştir.

Bulanık C-ortalamlar kümeleme yönteminde, veri birimleri iki veya daha fazla kümeye ait olabilmektedir. Buradaki temel mantık, veri birimleri [0,1] aralığında değişen üyelik değerlerine göre kümelere aittir. Buna göre birimlerin ait olduğu tüm kümelerdeki üyelik değerlerinin toplamı 1 olmaktadır. Birimler, küme merkezine alacağı en büyük üyelik değerine göre kümelere atanmaktadır. Buna göre veri birimi hangi küme merkezine yakınsa o kümeye ait olma üyelik değeri, diğer kümelere ait olma üyelik değerinden daha büyüktür [2][3].

Veri birimleri  $x_j \in R^d, j = 1, \dots, N$  ve  $K$  adet  $C = \{C_1, C_2, \dots, C_K\}$  küme merkezleri için FCM amaç fonksiyonu

$$f(X; U, V) = \sum_{i=1}^K \sum_{j=1}^N (u_{ij})^m \|x_j - v_i\|^2 \quad (2.29)$$

şeklindedir.

Burada, küme merkezi (prototip) vektörü

$$V = [v_1, v_2, \dots, v_k], v_i \in R^d \quad (2.30)$$

$U$  üyelik matrisi, [0,1] aralığında değerler alabilen  $u_{ij}$  elemanlarından oluşur.

$j$ . birim ve  $i$ . küme için  $u_{ij}$  üyelik elemanları aşağıdaki özellikleri sağlar.

$$\sum_{i=1}^K u_{ij} = 1 \quad (2.31)$$

$$0 < \sum_{i=1}^N u_{ij} < N \quad (2.32)$$

$\|.\|$ , literatürde genel olarak Öklid uzaklığını ifade eder. Ancak kümeleme analizlerinde kullanılan diğer uzaklık/benzerlik ölçümleri de FCM' de kullanılabilir.

$m, m \geq 1$  özelliğinde olup bulanık ağırlıklandırma katsayısıdır.

FCM algoritmasının çalışma mantığı, temel olarak, aşağıdaki gibi tanımlanmıştır [2][3]:

1.  $t = 0$  iterasyonu için,  $K$  küme sayısı,  $m$  bulanıklık katsayısı,  $\varepsilon$  işlem sonlanma kriteri ve  $U$  üyelik matris elemanları rasgele olarak belirlenir.
2. Denklem (2.34) kullanılarak küme merkezleri hesaplanır.

$$v_i = \frac{\sum_{j=1}^N (u_{ij})^m x_j}{\sum_{j=1}^N (u_{ij})^m}, 1 \leq i \leq K \quad (2.33)$$

3. Aşağıdaki denklem kullanılarak üyelik değerleri güncellenir.

$$u_{ij} = \sum_{k=1}^K \frac{(1/(\|x_j - v_i\|^2))^{1/(m-1)}}{(1/(\|x_j - v_k\|^2))^{1/(m-1)}}, i = 1, \dots, K \text{ ve } j = 1, \dots, N \quad (2.34)$$

4. Belirli bir iterasyon sonunda veya denklem (2.36) koşulundaki  $\varepsilon$  kriter değerinden küçük bir değere erişilinceye kadar 2. ve 3. adım tekrarlanır,

$$\|U^{t+1} - U^{(t)}\| < \varepsilon. \quad (2.35)$$

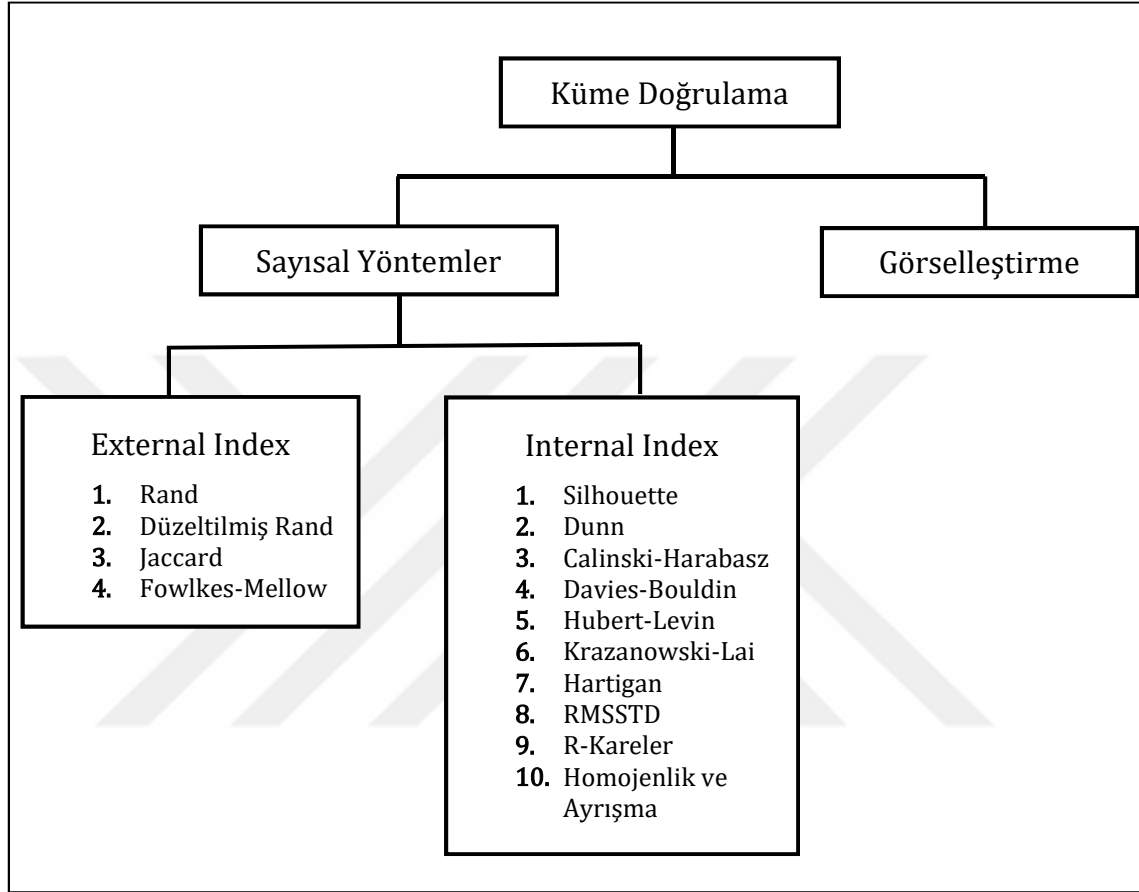
FCM, rastgele belirlenen başlangıç değerlerine bağlı olarak iteratif optimum sonuca ulaşmaya çalışan bir algoritmadır. Algoritma sonunda küme merkezleri veri setini temsil eden yerlerde konumlanırlar. Dolayısıyla, doğru küme merkezlerinin tespiti bu başlangıç değerlerinin seçimine bağlıdır [36][37].

## 2.5 Zaman Serileri Küme Doğrulama (Time Series Clustering Validity)

Bir veri setinde anlamlı küme yapıları bulunmasa bile kümeleme algoritmaları doğası gereği istenen küme sayısı doğrultusunda veri setinden küme bulacaktır. Ancak kümeleme algoritmalarıyla bulunan bu kümeler veri setinde her zaman küme olacağı anlamına gelmemektedir. Aynı şekilde veri setinde anlamlı küme yapıları olsa bile istenen küme sayısı doğrultusunda bulunan kümelerin bu gerçek kümeleri ne derece temsil ettiği de bilinmemektedir [1]-[3]. Bundan dolayı kümeleme analizlerinde elde edilen kümeleme sonuçlarının değerlendirilmesi kritik önem taşımaktadır. Literatürde küme sonuçlarının değerlendirilmesine yönelik çeşitli küme doğruluk yöntemleri (cluster validity) bulunmaktadır

[5][38][39]. Aghabozorgi ve arkadaşlarına [5] göre, küme doğrulama yöntemleri, temel olarak, sayısal ve sayısal olmayan görsel yöntemler diye ikiye ayrılır.

Şekil 2.3, zaman serileri kümeleme analizlerinde kullanılan sayısal ve sayısal olmayan küme doğrulama endekslerini özetlemektedir.



Şekil 2.3 Küme Doğrulama Yöntemleri

### 2.5.1 Sayısal Küme Doğrulama Ölçümleri (Scala Cluster Validity Measures)

Sayısal küme doğrulama ölçümleri, küme sonuçlarının doğruluğunu sayısal değerlerle ifade eden yöntemlerdir. Bu sayısal yöntemler harici, ve dahili küme doğrulama endeksleri olmak üzere ikiye ayrılırlar [2][3][5]. Bu endeksler arasındaki temel fark, dış bilgilerin küme doğrulama çalışmaları için kullanılıp kullanılmadığıdır. Örneğin entropi küme sınıfları, bilinen veri seti için küme çalışmalarının sonuçlarını değerlendiren dışsal bir küme doğrulama endeksidir.

Her ne kadar literatürde küme doğrulama yöntemleri için geliştirilen çeşitli endeksler olsa da üzerinde fikir birliğine varılmış ve evrensel olarak kabul görmüş

tek bir küme doğrulama endeksi bulunmamaktadır. Bu sebeple, kümeleme sürecinin başarılı bir şekilde sonuçlandırılabilmesi için karşılaştırmalı olarak, birçok küme doğrulama endeksi beraber kullanılmalıdır [1][5][13][40].

### 2.5.1.1 Harici Küme Doğrulama Endeksleri (External Cluster Validity Indices)

Harici küme doğrulama endeksleri, kümeleme sonuçlarını bir dış bilgi yardımıyla değerlendiren endekslerdir. Bu dış bilgi genellikle kümelerin orijinal sınıflarıdır. Buradaki temel mantık, kümeleme süreci ile elde edilen küme sınıflarının orijinal küme sınıflarıyla ne derece ölçüldüğünün ortaya konmasıdır. Bu alt başlıkta en sık kullanılan Rand, Düzeltmiş Rand, Jaccard ve Fowlkes-Mellow harici küme doğrulama endeksleridir [2][3][5].

X veri matrisi üzerinde, C doğal küme yapılarını ve P kümeleme algoritmalarının bulunduğu küme yapıların ifade etmek üzere:  $C = \{C_1, \dots, C_m\}$  ve  $P = \{P_1, \dots, P_s\}$ .  $\{x_v, x_u\}$ , X veri matrisinde tanımlanan veri birimi çiftini gösterebilir. Bu durumda dışsal küme doğrulama endeks hesaplamaları için aşağıdaki ifadeler kullanılır [38][39].

SS: Aynı C küme sınıfından gelen ikili veri birimi aynı P küme sınıfına dahil.

SD: Aynı C küme sınıfından gelen ikili veri birimi farklı P küme sınıflarına dahil.

DS: Farklı C küme sınıflarından gelen ikili veri birimi aynı P küme sınıfına dahil.

DD: Farklı C küme sınıflarından gelen ikili veri birimi farklı P küme sınıflarına dahil.

a, b, c ve d; SS, SD, DS ve DD durumlarına düşen ikili veri birimlerini ifade etsin. Bu tanımlar doğrultusunda dışsal küme doğrulama endeksleri aşağıdaki gibi hesaplanır. Endeksler için belirli bir alt ve üst sınır bulunmamaktadır. Verinin yapısına göre endeks değerleri farklı maksimum ve minimum değerler alabilmektedir. Karşılaştırma anlamında en yüksek dışsal endeks değerleri C ve P küme sınıfları arasındaki maksimum benzerliği ifade ederken; minimum endeks değerli ise C ve P küme sınıfları arasındaki minimum benzerliği ifade etmektedir. Diğer bir ifadeyle endeks değerlerindeki artış kümeleme başarısındaki artışı ifade etmektedir [38][39].

### 1. Rand Endeksi (Rand Index)

$$R = (a + d)/(a + b + c + d) \quad (2.36)$$

### 2. Düzeltilmiş Rand Endeksi (Adjusted Rand Index)

$$AR = \frac{\binom{n}{2}(a+d)-[(a+b)(a+c)+(c+d)(b+d)]}{\binom{n}{2}^2-[(a+b)(a+c)+(c+d)(b+d)]} \quad (2.37)$$

### 3. Jaccard Endeksi

$$J = a/(a + b + c) \quad (2.38)$$

### 4. Fowlkes-Mellows Endeksi (Adjusted Rand Index)

$$FM = \sqrt{\frac{a}{a+b} \cdot \frac{a}{a+c}} \quad (2.39)$$

#### 2.5.1.2 Dahili Küme Doğrulama Endeksleri (Internal Cluster Validity Indices)

Dahili küme doğrulama endeksleri, herhangi bir ön bilgiye gereksinim duymaksızın duymaksızın veri setinde bulunan özellikler ve bilgiler kullanılarak kümeleme sonuçlarını değerlendiren endekslerdir. Örneğin, içsel olarak belirlenen herhangi bir noktaya olan yakınlık veya uzaklık küme sonuçlarını değerlendirme kriteri olarak kullanılabilir [2][3][5].

En sık kullanılan dahili küme doğrulama endeksleri olan Silhouette, Davies-Bouldin, Calinski- Harabasz, Dunn, Hubert-Levin (C-index), Krzanowski-Lai, Hartigan, Ortalama Kareler Standart Sapma (Root-Mean-Square Standard Deviation), R-Kareler (R-Squared) , Kısmi R Kareler (Semi-partial R- Squared), Homojenlik (Homogeneity) ve Ayrışma (Separation) yöntemleri karşımıza çıkar.

#### 1. Silhouette Endeksi

Silhouette endeksi, veri birimlerinin bulunduğu kümelere uygunluğunu belirlemek amacıyla Kaufman ve Rousseeuw [41] tarafından geliştirilmiştir.

$a(x_i)$ , i. veri biriminin bulunduğu kümedeki diğer küme elemanlarına olan ortalama uzaklığı/benzerliğini ve  $b(x_i)$ , i. veri biriminin diğer kümelerde bulunan küme



elemanlarına olan ortalama uzaklıkların minimumunu gösterebilir. Bu durumda  $i$ . veri birimi için Silhouette değeri

$$s(x_i) = \frac{b(x_i) - a(x_i)}{\max\{a(i), b(i)\}} \quad (2.40)$$

ile hesaplanır.

Silhouette değeri -1 ile +1 arasında olup -1'e yaklaştıkça birimlerin yanlış kümeleneceği artmakta, +1'e yaklaştıkça birimlerin doğru kümeleneceği artmaktadır. Silhouette 0 değeri için birimlerin doğru kümeleneceği kümeleneceği ile ilgili bir yargıya varılamamaktadır [41].

Silhouette değeri veri birimlerinin tek başına doğru kümeleneceği kümeleneceği ile ilgili bilgi verirken kümeleme sürecinin başarısı ile ilgili bir bilgi vermez. Genel olarak kümeleme sürecinin başarısı ile ilgili bir kaniye, ortalama Silhouette değeri

$$S = \frac{1}{n} \sum_{i=1}^n s(x_i) \quad (2.41)$$

ile hesaplanarak verilebilir.

Genel olarak, 0.5 den büyük ortalama Silhouette değerine sahip kümeleme süreçlerinin başarılı olduğu kabul edilir. Ortalama Silhouette değeri arttıkça da kümeleme başarısı artmaktadır.

## 2. Dunn Endeksi

Dunn endeksi, yoğun ve iyi ayrılmış kümeleri tespit etmek amacıyla, Dunn [42] tarafından geliştirilmiştir.  $n$  adet  $C = \{C_1, C_2, \dots, C_n\}$  kümeyle bölünen veri seti için Dunn endeksi

$$D = \min_{i=1, \dots, n} \left\{ \min_{j=i+1, \dots, n} \left\{ \frac{d(C_i, C_j)}{\max_{k=1, \dots, n} \{\Delta(C_k)\}} \right\} \right\} \quad (2.42)$$

ile bulunur.

Denklem (2.42)'de  $C_i$  ve  $C_j$  kümeleri arasındaki uzaklık fonksiyonu

$$d(C_i, C_j) = \min_{x \in C_i, y \in C_j} d(x, y) \quad (2.43)$$

biçiminde hesaplanır.

C kümesi için  $\Delta(C)$ , küme içi birimler arası uzaklığı ifade eder ve

$$\Delta(C) = \max_{x,y \in C} d(x,y) \quad (2.44)$$

Şeklinde hesaplanır.

Dunn endeksi,  $[0, +\infty)$  açık aralığında değer alabilir. Kümeleme analizi sonucu elde edilen kümeler yoğun ve iyi ayrılmışlarsa kümeler arası uzaklık maksimum ve küme içi uzaklık minimum olur. Dolayısı ile Dunn endeksinin alacağı pozitif değerle orantılı olarak küme kalitesi artmaktadır. Kümeleme analizlerinde Dunn endeksinin mümkün olduğunca büyük değer alması istenir [42].

### 3. Calinski-Harabasz Endeksi

Calinski ve Harabasz [43], bir küme doğrulama yöntemi önermişlerdir. k küme sayısı, n toplam gözlem sayısı,  $n_k$  k. kümedeki gözlem sayısını,  $x_i$  i. gözlem birimini ve  $\bar{x}_k$  k. kümedeki gözlem birimlerinin ortalamasını göstermek üzere Calinski-Harabasz (CH) [44] endeksi

$$CH = \frac{[\sum_{i=1}^n (x_i - \bar{x})^2 - \sum_k^G \sum_{i=1}^{n_k} (x_i - \bar{x}_k)^2] / (G-1)}{[\sum_k^G \sum_{i=1}^{n_k} (x_i - \bar{x}_k)^2] / (n-G)} \quad (2.45)$$

ile bulunur.

CH endeksi için belirlenmiş bir alt ve üst sınır bulunmamaktadır. Kümeleme başarısının artışına paralel olarak CH endeks değeri de artmaktadır. Bu yüzden en yüksek CH değerini veren küme sayısı o kümeleme analiz için en optimum küme sayısıdır [43][44].

### 4. Davies-Bouldin Endeksi

Davies ve Bouldin [45] tarafından da bir küme doğrulama endeksi geliştirilmiştir. Endeks, küme içi gözlemlerin küme merkezlerine olan uzaklıklarını, kümeler arası uzaklıklara oranını ölçmeyi amaçlamaktadır. n adet  $C = \{C_1, C_2, \dots, C_n\}$  kümeye bölünen veri seti için Davies-Bouldin (DB) [46] endeksi

$$DB = \frac{1}{n} \sum_{i=1}^n \max_{i \neq j} \left\{ \frac{\Delta(C_i) + \Delta(C_j)}{d(X_i, X_j)} \right\} \quad (2.46)$$

ile hesaplanır.

$C_i$  ve  $C_j$  küme merkezleri arasındaki uzaklık

$$d(C_i, C_j) = \min_{x \in C_i, x \in C_j} d(x, y) \quad (2.47)$$

ile hesaplanır.

$C$  kümesi için  $\Delta(C)$ , küme içi birimler arası uzaklığı ifade eder ve

$$\Delta(C) = \max_{x, y \in C} d(x, y) \quad (2.48)$$

biçiminde ifade edilir.

DB endeksi için belirlenmiş bir alt ve üst sınır bulunmayıp, DB değeri küçüldükçe kümelerin yoğunluğu artmakta ve küme merkezlerinin birbirlerine göre uzaklıkları artmaktadır. En küçük DB değerini veren küme sayısı o kümeleme analizi için en optimum küme sayısıdır.

### 5. Hubert-Levin Endeksi

Hubert-Levin diğer bir isimle C endeksi Hubert ve Levin [47] tarafından bir küme doğrulama endeksi olarak önerilmiştir. Kümelerin gözlem çiftleri arasındaki minimum uzaklıkları toplamı  $S_{\min}$ , maksimum uzaklıkları toplamı  $S_{\max}$  ve toplam uzaklıklar toplamı  $S$  ile gösterilmek üzere C [48] endeksi

$$C = \frac{S - S_{\min}}{S_{\max} - S_{\min}}, S_{\min} \neq S_{\max} \quad (2.49)$$

biçiminde hesaplanır.

C endeksi 0 ile 1 arasında değer olup C değeri küçüldükçe kümeleme başarısı artmaktadır. En küçük C endeks değerini veren küme sayısı o kümeleme analizi için en optimum küme sayısıdır [47][48].

### 6. Krzanowski-Lai Endeksi

Krzanowski ve Lai [49] tarafından ortaya atılan bu endeks  $n$  adet  $C = \{C_1, C_2, \dots, C_n\}$  kümeye bölünen veri seti için KL [50] endeksi

$$KL = \left| \frac{DIFF(K)}{DIFF(K+1)} \right| \quad (2.50)$$

şeklinde bulunur.

p veri setindeki deęişken sayısı olmak üzere  $DIFF(K)$

$$DIFF(K) = (K - 1)^{\frac{2}{p}} \text{trace} W_{K-1} - K^{\frac{2}{p}} \cdot \text{trace} W_K \quad (2.51)$$

ile bulunur.

$\text{trace} W_K$  K küme içi kareler toplamını ve  $\frac{2}{p}$ ,  $j \in [1, \dots, J]$  deęişkenleri üzerinde düzgün üniforma dağılmış gözlemler üzerinden türetilmiş normalleştirme faktörünü ifade eder [50].

KL endeksi için belirlenmiş bir alt ve üst sınır bulunmayıp endeks deęeri büyüdükçe kümeleme başarısı artmaktadır [49][50].

### 7. Hartigan Endeksi

n adet  $C = \{C_1, C_2, \dots, C_n\}$  kümeye bölünen veri serti için Hartigan (H) endeksi

$$H = \frac{W(k) - W(k+1)}{W(k+1)} (n - k - 1) \quad (2.52)$$

ifadesi ile hesaplanırken  $W(k)$  de

$$W(k) = \sum_{k=1}^n \sum_i^{n_k} \|x_{i(k)} - \bar{x}_{(k)}\|^2 \quad (2.53)$$

biçiminde hesaplanır.

H endeks deęeri büyüdükçe kümeleme başarısı artmaktadır [50].

### 8. Kök-Ortalama-Kare Standart Sapma Endeksi (The Root-Mean-Square Standart Index)

Kök-Ortalama-Kare standart sapma (RMSSTD) endeksi küme içi homojenliği ölçmeyi hedefleyen bir küme doğrulama endeksidir. nc küme sayısını, d deęişken sayısını, J deęişkeni için  $N_j$  toplam gözlem sayısını ve  $n_{ij}$  veri birimini ve  $\bar{x}_j$  veri ortalamasını ifade etmek üzere, RMSSTD [38][39]

$$RMSSTD = \left[ \frac{\sum_{j=1, \dots, d} \sum_{i=1, \dots, nc}^{N_{ij}} (x_k - \bar{x}_j)^2}{\sum_{j=1, \dots, d} (N_{ij} - 1)} \right]^{\frac{1}{2}} \quad (2.54)$$

ile bulunur.

RMSSTD endeks değeri küçüldükçe küme içi homojenlik artmaktadır. [38][39].

### 9. R-Kareler Endeksi (R-Squared Index)

R-Kareler (RS) endeksi, kümeler arası farklılığı ölçmeyi hedefleyen bir küme doğrulama endeksidir.  $d$  değişken sayısını,  $J$  değişkeni için  $N_j$  toplam gözlem sayısını ve  $n_{ij}$  veri birimini ve  $\bar{x}_j$  veri ortalamasını ifade etmek üzere, RS endeksi

$$RS = \frac{\left\{ \sum_{j=1, \dots, d} \left[ \sum_{k=1}^{N_j} (x_k - \bar{x}_j)^2 \right] \right\} - \left\{ \sum_{i=1, \dots, nc} \left[ \sum_{j=1, \dots, d} n_{ij} (x_k - \bar{x}_j)^2 \right] \right\}}{\sum_{j=1, \dots, d} \left[ \sum_{k=1}^{N_j} (x_k - \bar{x}_j)^2 \right]} \quad (2.55)$$

olarak bulunur [38][39].

RS endeksi  $[0,1]$  kapalı aralığında değer alabilirken endeks değeri küçüldükçe kümeler arası farklılık azalmaktadır. RS değerinin 0 olması, veri setinde anlamlı kümelerin olmadığı anlamına gelmektedir. RS değerinin 1 olması kümeler arasında anlamlı farklılıklar olduğu anlamına gelmektedir [38][39].

### 10. Homojenlik ve Ayrışma Endeksi (The Homogeneity & Separation Index)

Bir kümenin homojenliğini (HI), kümeleri oluşturan veri birimlerinin birbirlerine göre benzerliğinin ortalaması olarak düşünelim.  $M$  toplam gözlem çiftini ve  $F(i)$  her  $i$  gözlemi için bir veri vektörünü ifade etsin. Bu durumda küme homojenliği [51].

$$HI_{ort} = \frac{1}{M} \sum_{i < j} S(F(i), F(j)) \quad (2.56)$$

ile tanımlanır.

Benzer şekilde, kümeler arası ayrışma farklı (S) kümelerdeki veri birimlerinin bir birine göre benzerliklerinin ortalaması şeklinde düşünülür. Böylece kümeler arası ayrışma [51].

$$SI_{ort} = \frac{2}{n(n-1) - 2M} \sum_{i < j} S(F(i), F(j)) \quad (2.57)$$

ile tanımlanır.

$HI_{ort}$  ve  $SI_{ort}$  için belirlenmiş bir alt ve üst sınır bulunmayıp,  $HI_{ort}$  ve  $SI_{ort}$  değerleri arttıkça küme içi homojenlik ve kümeler arası ayrışma artmaktadır [51].

### 2.5.2 Görsel Küme Doğrulama (Visual Cluster Validity)

Sayısal küme doğrulama endeksleri kümeleme sonuçlarını değerlendirirken, kümeleri temsil edebilecek noktaları referans noktaları olarak seçerek, birimlerin referans noktaları ile aralarındaki uzaklıkları ve referans noktalarından hareketle varyans gibi parametreleri hesaplarlar. İdeal olmayan küme biçimlerini temsil edebilecek referans noktalarının seçilmesindeki zorluklardan dolayı bu tür endekslerle ideal olmayan küme biçimlerinin değerlendirilmesi her zaman doğru olmayabilir. Literatürde yapılan çalışmalar da incelendiğinde, sayısal küme doğrulama endekslerinin ideal olmayan küme biçimleri için her zaman en ideal çözümü sağlamadığı ve kendi içerisinde tutarsız sonuçlar verebildiği görülmektedir [1][13][40][52][53]. Bu sebeplerden dolayı, sayısal küme doğrulama endekslerine alternatif olarak, insan algı sisteminin de devreye sokulabileceği görsel küme doğrulama yöntemleri geliştirilmiştir [5].

Bu çalışmada matris grafikleri (heat maps) görsel küme doğrulama (visual cluster validity) aracı olarak kullanılmıştır. Matris grafiğinin temel amacı, veri matrisinde bulunan verileri veri büyüklüklerine göre farklı renklerle temsil ederek görsel bir anlatım sağlamaktır [1].

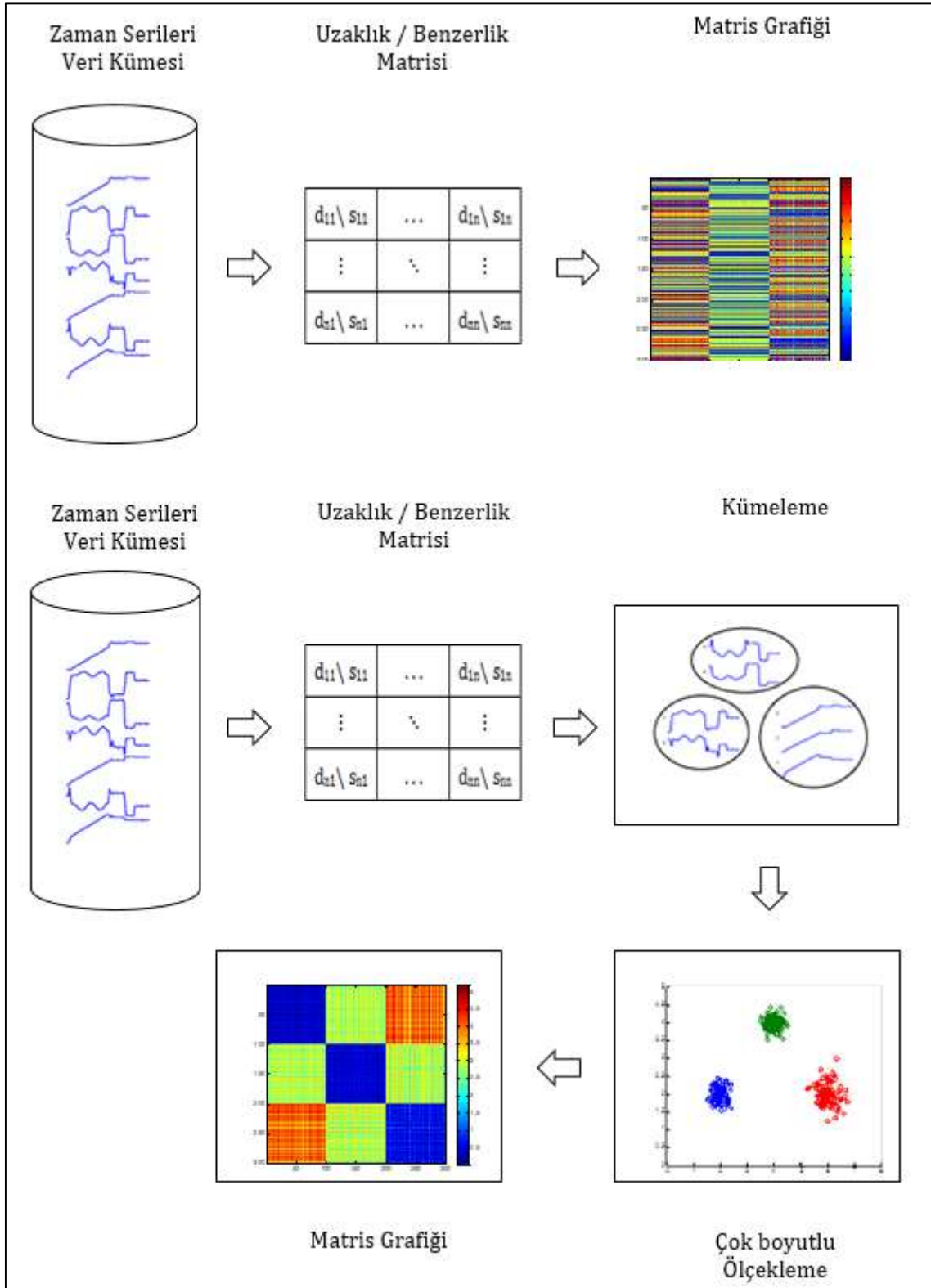
Matris grafikler, çok boyutlu veri yapılarını iki boyutlu bir düzlemde görsel olarak ifade ederek, işlevsel olarak çok boyutlu ölçekleme analizlerini hatırlatmaktadırlar. Matris grafikleri, büyük veri yapıları içerisinde doğru küme sayıları hakkında genel bir fikir edinmeyi ve buna bağlı olarak kümeleme süreçlerinde gerekli iyileştirmelerin yapılmasına olanak tanımaktadırlar [1][29].

Kümeleme sürecinde hesaplanan uzaklık veya benzerlik ölçümleri matris veri değeri gibi kullanılarak, veri birimleri arasındaki uzaklık veya benzerlik ilişkileri renklendirme vasıtasıyla gözlemlenebilir. İleri düzey bir analiz olarak, kümeleme analizi sonucu belirlenen küme sınıfı üyeliklerinin birbirlerine göre uzaklık veya benzerlik ölçümleri düzenlenerek, kümeleme analiz sonucu görsel bir şekilde ortaya konulabilir. Bu vesileyle insan algı sisteminin de devreye girmesiyle küme

sonuları iki veya  boyutlu olarak grsel algılama sistemiyle deęerlendirilir [1][29][40][53].

Őekil 2.4' de rnek bir veri kme iin matris grafikleri, kmeleme ncesi veri setine ve kmeleme sonrası elde edilen sonulara gre izilmiŐtir. Őekilden de grleceęi zere kmeleme iŐlemi yapılmamıŐ rastgele olarak sıralanmıŐ, zaman serileri iin anlamlı renk sınıflarının oluŐmadıęı gzlenmektedir. Aksine, kmeleme iŐlemi yapılmıŐ, benzer kmelerde yer alan zaman serilerinin, matris grafikleri zerinde hatta ok boyutlu lekleme grafięinde, benzer renklenmektedir





Şekil 2.4 Kümeleme Öncesi Ve Sonrası Matris Grafikleri



Bir  $y$  değişkeninin belirli bir  $t$  anında almış olduğu değer  $y_t$  ile gösterilmek üzere, zaman serileri  $\{y_t, t \in T\}$  kümesi şeklinde gösterilen bir stokastik süreçtir.  $y_t$  zaman serisinin analizinin asıl amacı, tek bir zaman serisine ait gözlemlerin dinamik veya zamana bağlı değişen yapısını ortaya çıkartmaktır. Bu bilgiler doğrultusunda tek değişkenli zaman serileri modelleri kendi geçmiş değerleri ve hatalarına göre kurulan modeller olup, temel olarak.

$$y_t = f(y_{t-1}, y_{t-2}, \dots, e_{t-1}, e_{t-2}, \dots) \quad (3.1)$$

ile ifade edilir[54][55].

Burada  $y_t$ ' nin gecikme değerleri  $y_{t-1}, y_{t-2}$ , artıklarının gecikme değeri  $e_{t-1}, e_{t-2}$  şeklinde ifade edilir. (3.1)' deki fonksiyonel ilişki en optimum seviyede hangi zaman seri modelinin uygun olduğuna karar verilmesi gerektiğini ifade eder. Buradaki model, zaman serisi gözlemlerinin arasındaki veya gözlemlerin hataları ile yada gözlemlerin hem kendi arasındaki hem de hatalar ile olan ilişkilerinin tanımlanmasından oluşabilir [55].

Bir  $y_t$  zaman serisinin hata serisi  $e_t$  gözlemlenmeyen bir değişken olmak üzere, tahmini ve gerçekleşen zaman serileri arasındaki fark

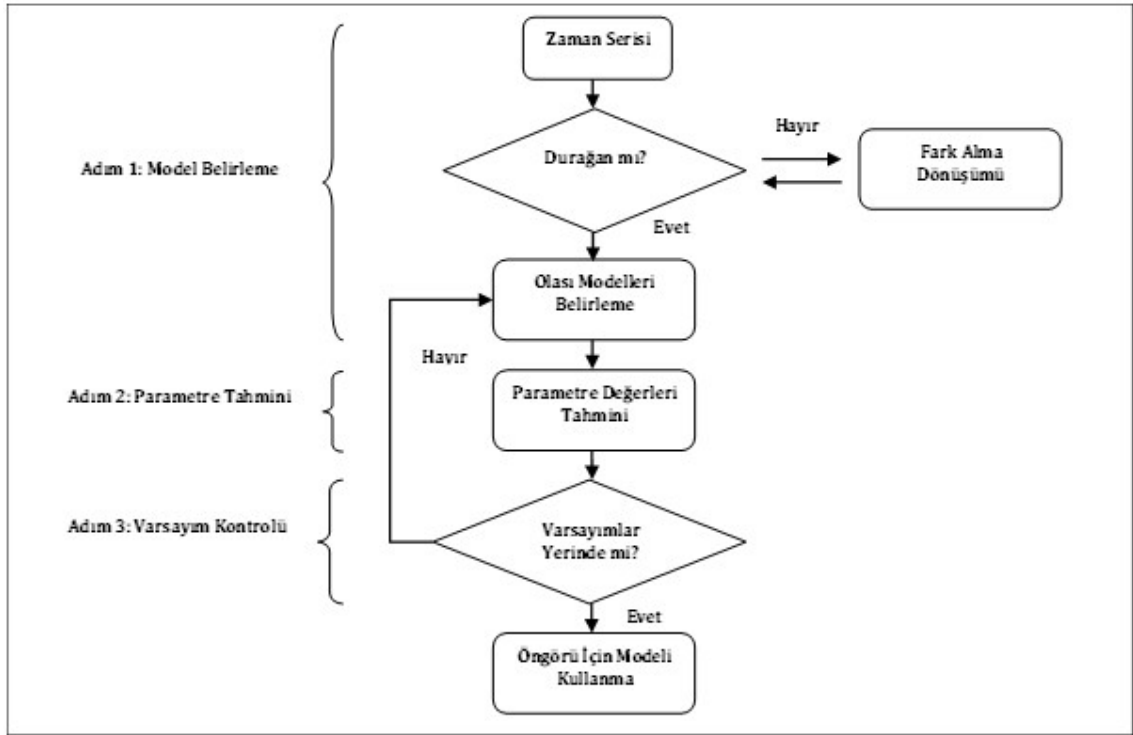
$$e_t = y_t - \hat{y}_t \quad (3.2)$$

Şeklinde ifade edilir.

Box-Jenkins yöntemi tek değişkenli zaman serileri için, en uygun geleceği öngörme fonksiyonunun belirlenmesinde bir model çerçevesi olarak Box-Jenkins [56] tarafından 1976 yılında ortaya atılmıştır. Kısa dönem öngörülerde oldukça başarılı olan bu yöntemin uygulandığı zaman serisinin, eşit zaman aralıklarıyla elde edilen gözlem değerlerinden oluşan, kesikli, durağan ve rassallık özelliği sağlaması gerekmektedir [56].

Zaman serileri modellemelerinde ön şart olarak durağanlık koşulu aranmaktadır. Durağanlık özelliği sergilemeyen zaman serileri üzerinden yapılan bir çalışmada zaman serilerini doğru tahminleyen modellerin bulunması dolayısı ile de güvenilir öngörü sonuçlarının elde edilmesi mümkün olamamaktadır [56].

Temel olarak Box-Jenkins yaklaşımı üç ana adımdan oluşmaktadır: Model Belirleme, Parametre Tahmini ve Varsayım Kontrolü (Diagnostic Checking). Bu yöntem, üç ana adım altında özetlenmektedir [56].



Şekil 3.1 Box-Jenkins Aşamaları Akış Diyagramı

Pür Rassal Süreç: Pür rassallık (White Noise) özelliği sağlayan zaman serileri sıfır ortalamalı, sabit varyanslı ve özilintisizdir (otokorelasyonsuz).  $k$  gecikme sayısı olmak üzere, hata terimi için pür rassal süreç [56]

$$E(\varepsilon_t) = 0; \quad E(\varepsilon_t^2) = \sigma^2; \quad E(\varepsilon_t, \varepsilon_{t-k}) = 0, \quad k \neq 0 \quad (3.3)$$

özelliklerini sağlar.

Pür rassal süreci, bağımsız, özdeş ve normal dağılımlı ise Gaussçu pür rassallık (Gaussian White Noise) olarak adlandırılır. Hata terimi için Gaussçu pür rassallık

$$e_t \sim \text{IID}(0, \sigma^2) \quad t = 1, 2, \dots, n \quad (3.4)$$

şeklinde ifade edilir [56].

Durağanlık: Ortalamasının ve varyansının zamana göre değişmediği ve iki dönem arasındaki kovaryansın bakılan döneme değil de dönemler arasındaki uzaklığa bağlı olduğu süreçlere durağan süreçler denir [56].

Herhangi bir  $y_t$  zaman serisinin durağan olabilmesi için aşağıdaki şartları sağlaması gerekmektedir [56]:

$$E(y_t) = \mu \quad (3.5)$$

$$E[(y_t - \mu)] = \text{Var}(y_t) = \sigma^2 \quad (3.6)$$

$$E[(y_t - \mu)(y_{t-k} - \mu)] = \text{Cov}(y_t, y_{t-k}) = \gamma_k \quad (k \neq 0) \quad (3.7)$$

(3.5) ve (3.6) denklemleri stokastik zaman serisi sürecinin ortalama ve varyansının sabit olduğunu göstermektedir. Denklem (3.7) serinin herhangi iki değer arasındaki kovaryansının belirli bir t zaman noktasına değil de, sadece iki değer arasındaki k indisine bağlıdır. Zaman serileri modellemesindeki amaç anlamlı öngörülerde bulunmak ise, bu ancak stokastik zaman serisinin zaman süresince sabit (durağan) olması ile mümkündür. Aksi halde, durağan olmayan zaman serileri ile çalışıldığında, oluşturulacak olan zaman serisi modelinin sonuçları gerçekçi olmayabilmektedir [56].

Box-Jenkins göre zaman serisi modelleri, otoregresif (autoregresif), hareketli ortalama (moving average) ve otoregresif hareketli ortalama (autoregresif moving average) modeli olmak üzere 3 şekilde olabilir. Aşağıdaki alt başlıklarda bu modelleme süreçleri açıklanmıştır.

### 3.1 Otoregresif Modelleri

Temel olarak, p. dereceden otoregresif süreç,  $AR(p)$ ,

$$y_t = \sum_{k=1}^p \phi_k y_{t-k} + e_t \quad (3.8)$$

şeklinde ifade edilir.

Bu denklemde  $e_t$  denklem (3.3)'deki özelliklerini sağlayan Gaussçu pür rassal süreci ifade etmektedir [56].

$B^k y_t = y_{t-k}$  geri öteleme operatörü (3.8)'de kullanılarak

$$(1 - \sum_{k=1}^p \phi_k B^k) y_t = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = e_t \quad (3.9)$$

kapalı ifadesi elde edilir.

1. dereceden otoregresif süreç  $AR(1)$  olmak üzere,

$$y_t = \phi_1 y_{t-1} + \varepsilon_t \quad (3.10)$$

$$(1 - \phi_1 B) y_t = e_t \quad (3.11)$$

$AR(1)$  ifadesi (3.10)'daki gibi açık ve (3.11)'deki gibi kapalı olarak ifade edilir.

### 3.2 Hareketli Ortalamalar Modelleri

Temel olarak,  $q$ . dereceden otoregresif süreç  $MA(q)$

$$y_t = \sum_{k=1}^q \theta_k e_{t-k} + e_t \quad (3.12)$$

biçiminde verilir.

Bu denklemde  $MA(q)$  doğrusal modelin açık bir gösterimidir.  $B^k e_t = e_{t-k}$  geri öteleme operatörü kullanılarak denklem (3.12)'nin kapalı biçimi

$$(1 + \sum_{k=1}^q \theta_k B^k) e_t = (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_p B^p) e_t = y_t \quad (3.13)$$

olarak bulunur.

1. dereceden otoregresif süreç  $MA(1)$  olmak üzere,

$$y_t = \theta_1 e_{t-1} + \varepsilon_t \quad (3.14)$$

$$(1 - \theta_1 B) e_t = y_t \quad (3.15)$$

şeklinde verilen ifadeler, sırasıyla,  $MA(1)$ 'in açık ve kapalı formlarıdır.

### 3.3 Otoregresif Hareketli Ortalamalar Karma Modelleri

Temel olarak,  $(p, q)$ . dereceden bir otoregresif hareketli ortalamalar süreci ARMA(p, q),

$$y_t = \mu + \sum_{k=1}^p \phi_k y_{t-k} + \sum_{k=1}^q \theta_k e_{t-k} + e_t \quad (3.16)$$

ile verilir.

Burada  $\mu$  sabit terimi ifade etmektedir.

(3.16) ifadesi, ARMA(p, q) modelinin açık bir gösterimidir.

B geri öteleme operatörü kullanılarak (3.16)'nın kapalı formu

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) y_t = \mu + (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) e_t \quad (3.17)$$

olarak verilir.

1. dereceden otoregresif hareketli ortalamalar süreci ARMA(1,1)'in açık ve kapalı formu sırasıyla

$$y_t = \phi_1 y_{t-1} + \theta_1 e_{t-1} + \varepsilon_t \quad (3.18)$$

$$(1 - \theta_1 B) y_t = (1 - \phi_1 B) e_t \quad (3.19)$$

olarak verilebilir.

### 3.4 Durağan Olmayan Otoregresif Hareketli Ortalamalar Karma Modelleri (Autoregressive Integrated Moving Average Models)

Şu ana kadar anlatılan otoregresif (AR) ve hareketli ortalamalar (MA) modelleri zaman serilerinin durağan olması şartı altında tasarlanmıştır. Zaman serilerini durağan bir zaman serisi haline getirmek için ihtiyaca göre 1 veya 2 defa fark alma işlemi uygulanır. Fark alma işlemi  $d$  sembolü ile gösterilir. Durağan olmayan ancak fark alma işlemi ile durağan hale getirilen zaman serileri üzerinde geliştirilen AR veya MA veya ARMA modellerine karma (entegre) modeller denir. Eğer AR modelinin derecesi  $p$ , MA modelinin derecesi  $q$  ve serinin de  $d$  kere fark alma işlemi ile durağan olması durumunda oluşturulan modele otoregresif entegre hareketli ortalama modeli denir ve ARIMA(p, d, q) şeklinde gösterilir [56].

Bu durumda, ARIMA(p, d, q) modeli

$$y_t = \mu + \sum_{k=1}^p \phi_k y_{t-k} + \sum_{k=1}^q \theta_k e_{t-k} + e_t \quad (3.20)$$

olarak bulunur. Burada  $\mu$  bir sabit olup (3.20) ifadesi ARMA(p, d, q) modelinin açık bir ifadesidir.

B geri operatörü dikkate alındığında

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d y_t = \mu + (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q) e_t \quad (3.21)$$

kapalı formunda ya da daha açık bir şekilde

$$y_t - y_{t-1} = \phi_1 (y_{t-1} - y_{t-2}) + \phi_2 (y_{t-2} - y_{t-3}) + \dots + \phi_p (y_{t-p} - y_{t-p-1}) \quad (3.22)$$

$$+ e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

olarak bulunur.

1. dereceden otoregresif hareketli ortalamalar süreci ARMA(1,1), sırasıyla,

$$y_t = \phi_1 y_{t-1} + \theta_1 e_{t-1} + \varepsilon_t \quad (3.23)$$

$$(1 - \phi_1 B) y_t = (1 - \theta_1 B) e_t \quad (3.24)$$

açık ve kapalı formlarda ifade edilir.

### 3.5 Durağanlık Analiz

Bir zaman serisinin durağan olup olmadığını veya durağansa hangi mertebeden durağan olduğunu belirlemek için bir takım testler geliştirilmiştir. Durağanlığın sınanmasında genel olarak birim kök testleri kullanılmaktadır. Birim kök testi genel olarak [57]

$$y_t = \rho y_{t-1} + \varepsilon_t \quad (3.25)$$

olarak ifade edilir.

Burada  $\varepsilon_t$  hata terimi pür rassaldır. ve  $y_{t-1}$ ' in katsayısı  $\rho$ , 1' e eşitse bu seri birim kök içeriyor denir. Katsayının bire eşitliğini sınanan test t testidir. Bu durumda durağanlık testi için hipotez testi aşağıdaki gibi geliştirilmiştir[57]:

$H_0: \rho = 1$  seri durağan değildir. Serinin birim kökü vardır.

$H_a: \rho < 1$  seri durağandır. Serinin birim kökü yoktur.

Hipotezlere ilişkin test istatistiği

$$t = \frac{\hat{\rho}-1}{s_p} \quad (3.26)$$

ile hesaplanır.

Herhangi bir  $y_t$  serisinin “t” istatistiğinin hesaplanabilmesi için zaman serisinin durağan olması gerekmektedir. Durağan olmayan zaman seriler için hesaplanan istatistiği t dağılımına uymaz ve sapmalı olur. Bu durumda  $H_0$  hipotezinin, “t” istatistiğinin t dağılımı ile test edilmesi gerçekliliğini yitirir. Yani (3.25)’de yer alan  $y_{t-1}$  varlığından dolayı etkinlik kaybı olur ve  $\hat{\rho}$  aşağı yönlü sapmalı olur. Bu sebeplerden dolayı standart hata büyür ve hipotez testiyle yanlış karar verilir. Bu durumun üstesinden gelmek için çeşitli birim kök testleri geliştirilmiştir [57].

En fazla karşılaşılan birim kök testleri: Genişletilmiş Dickey Fuller (ADF), Phillips-Perron (PP) ve Kwiatkowski-Phillips-Schmidt Shin (KPSS)

### 3.5.1 Genişletilmiş Dickey Fuller Testi (Augmented Dickey Fuller -ADF)

ADF testi Said ve Dickey [58] tarafından ifade edilmiştir. Bu testte  $\Delta$  fark alma operatörü kullanılarak denklem (3.25)’deki  $y_t$  zaman serisine ilişkin durağan olmama durumu ortadan kaldırılmıştır. Yeni haliyle test edilecek model aşağıdaki gibi adım adım ifade edilmiştir:

$$y_t = \rho y_{t-1} + \epsilon_t \quad (3.27)$$

$$y_t - y_{t-1} = \rho y_{t-1} - y_{t-1} + \epsilon_t \quad (3.28)$$

$$\Delta y_t = (\rho - 1)y_{t-1} + \epsilon_t \quad (3.29)$$

$$\Delta y_t = \rho^* y_{t-1} + \epsilon_t \quad (3.30)$$

Said ve Dickey, testlerinde oluşabilecek otokorelasyon sorununu ortadan kaldırabilmek için ADF modeline  $\Delta y_t$  serisinin gecikme değerleri de eklenmiştir. ADF testinde, modeller sabit terimli, sabit terimsiz veya sabit terim ve eğilim içerecek şekilde üç farklı şekilde, sırasıyla olabilmektedir [58]:

$$\Delta y_t = \rho^* y_{t-1} + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \varepsilon_t, \quad (3.31)$$

$$\Delta y_t = \mu + \rho^* y_{t-1} + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \varepsilon_t, \quad (3.32)$$

$$\Delta y_t = \mu + \beta t + \rho^* y_{t-1} + \sum_{i=1}^p \beta_i \Delta y_{t-i} + \varepsilon_t. \quad (3.33)$$

Test edilecek modellere ilişkin olarak geliştirilen hipotez aşağıdaki gibidir [58]:

$H_0: \rho^* = 0$  seri durağan değildir. Serinin birim kökü vardır.

$H_a: \rho^* < 0$  seri durağandır. Serinin birim kökü yoktur.

ADF testi için hipotez test istatistiği

$$t = \frac{\hat{\rho}}{s_p} \quad (3.34)$$

ile hesaplanır.

### 3.5.2 Phillips-Perron (PP) Testi

PP, ADF testine alternatif olarak Phillips-Perron (PP) testini geliştirmişlerdir [59]. ADF testinde hatalarda meydana gelen otokorelasyon ve değişen varyans sorununu ADF modeline eklenen gecikme değerleri ile çözüm geliştirilmişti. PP testinde ise hata terimlerindeki otokorelasyon önemsenmeyerek, t test istatistiği olacak şekilde yeniden düzenlenmiştir [59].

PP testinde kullanılan model ifadesi

$$\Delta y_t = \pi y_{t-1} + \varepsilon_t \quad (3.35)$$

$$\Delta y_t = \mu + \pi y_{t-1} + \varepsilon_t \quad (3.36)$$

$$\Delta y_t = \mu + \beta t + \pi y_{t-1} + \varepsilon_t \quad (3.37)$$

şeklindedir.

Yeniden düzenlenen PP testi

$$Z_t = \left( \frac{\hat{\sigma}^2}{\hat{\lambda}^2} \right)^{1/2} \cdot t_{\pi=0} - \frac{1}{2} \left( \frac{\hat{\lambda}^2 - \hat{\sigma}^2}{\hat{\lambda}^2} \right) \left( \frac{T \cdot SE(\hat{\pi})}{\hat{\sigma}^2} \right) \quad (3.38)$$

$$Z_{\pi} = T_{\hat{\pi}} - \frac{1}{2} \cdot \frac{T^2 \cdot SE(\hat{\pi})}{\hat{\sigma}^2} \cdot (\hat{\lambda}^2 - \hat{\sigma}^2) \quad (3.39)$$



gibidir.

$$\sigma^2 = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(\varepsilon_t^2) \quad (3.40)$$

$$\lambda^2 = \lim_{T \rightarrow \infty} \sum_{t=1}^T E(T^{-1} S_T^2) \quad (3.41)$$

$$S_T = \sum_{t=1}^T \varepsilon_t \quad (3.42)$$

$H_0: \pi = 0$  hipotezi altında, PP testi,  $Z_t$  ve  $Z_\pi$  istatistikleri, ADF t-istatistiği ve normalize önyargı istatistikleriyle aynı asimptotik dağılımlara sahiptir. Avantaj olarak, PP testi ADF testine göre, hata terimlerinde meydana gelebilecek değişen varyans sorununa karşı daha az duyarlıdır. PP testinin diğer bir avantaj ise kullanıcının regresyon testi için bir gecikme uzunluğu belirleme zorunluluğunda olmamasıdır.

### 3.5.3 Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Testi

ADF testinin zayıflığına yönelik literatürde yapılan pek çok araştırma bulunmaktadır. Örneğin Schwert [60], ADF testinin kuvvetli bir test olmadığını ve ADF testinin seçilen gecikme uzunluğuna oldukça duyarlı olduğunu göstermiştir. Bu bağlamda, ADF testi için seçilmesi gereken en uygun gecikme uzunluğunun belirlenmesi de karşılaşılan diğer bir problemdir.

Kwiatkowski ve arkadaşları [61] tarafından ADF testine alternatif olarak KPSS testi geliştirilmiştir. KPSS testindeki yaklaşım, ADF metodolojisinin tersine, durağan olamama sıfır hipotezine karşın, sıfır hipotezi durağan olacak şekilde geliştirilmiştir [61].

KPSS testine göre, bir zaman serisi, sabit terim, rassal terim ve bir deterministik eğilim bileşenlerinden oluşmaktadır. Testin amacı, zaman serisini deterministik eğiliminden arındırarak serinin durağan hale getirilmesidir. Buna göre  $Y_t$  zaman serisi, deterministik eğilim  $t$ ,  $r_t$  rassal süreç ve  $\varepsilon_t$  durağan hata terimi olacak şekilde denklem

$$y_t = \beta t + r_t + \varepsilon_t \quad (3.43)$$

ifade edilir.

Burada

$$r_t = r_{t-1} + u_t. \quad (3.44)$$

Burada ise  $u_t$ , bağımsız, özdeş ve normal dağılımlı olup  $u_t \sim \text{IID}(0, \sigma^2)$  özelliğini göstermektedir[61]. Bu durumda, KPSS testi için durağanlık testi hipotezi aşağıdaki gibi kurulur:

$H_0: \sigma_u^2 = 0$  seri durağandır. Serinin birim kökü yoktur.

$H_a: \sigma_u^2 \neq 0$  seri durağan değildir. Serinin birim kökü vardır.

Hata terimi varyansının  $\sigma_u = 0$  olması, hata teriminin  $u_t$  sabit olmasını, sonuç olarak  $r_t$  rassal sürecinin durağan olması anlamına gelecektir.

### 3.6 Model Belirleme Kriterleri

Zaman serileri modelleme çalışmalarında birden fazla anlamlı model ortaya çıkabilir. Bu anlamlı modeller arasından serinin yapısına en iyi uyan modelin tespiti, kümeleme ve öngörü çalışmalarını doğrudan etkilemektedir. Bu durumdan dolayı, serinin yapısına en iyi uyan modellerin belirlenmesi için literatürde geliştirilmiş çeşitli kriterler bulunmaktadır. Literatürde en sık kullanılan model belirleme kriterleri Akaike ve arkadaşları [62] Bayesci bilgi kriteridir [62].

#### 3.6.1 Akaike Bilgi Kriteri (AIC – Akaike Information Criteria)

Zaman serileri model belirlenmesinde en çok kullanılan kriterlerden biri Akaike bilgi kriteridir (AIC). AIC,

$$\text{AIC} = -2\log L(\hat{\theta}) + 2k \quad (3.45)$$

ile hesaplanır.

Burada  $\theta$  model parameter vektörlerini,  $L(\hat{\theta})$  aday modellerinin olabirlik fonksiyonunu göstermektedir.  $\hat{\theta}$ ,  $L$  olabirlik fonksiyonunu maksimum yapacak şekilde tahmin edilmiştir. Aday modeller için,  $k$  tahmin edilen model parametre sayısını göstermektedir [62].

Küçük örneklemler için ( $n/k < 40$ ) AIC denklemi

$$\text{AIC}_c = -2\log L(\hat{\theta}) + 2k + (2k + 1)/(n - k - 1) \quad (3.46)$$

ile bulunur.

Burada, n toplam gözlem sayısını göstermektedir.

Model belirleme esnasında minimum AIC değeri, aday modelleri arasında en iyi modeli verir. Aday modellerin en iyi model olma seviyeleri delta ( $\Delta$ )AIC ile ölçülür.  $\Delta_m$  aday modellerinin bağlı olabilirlik değerlerine göre standardize edilmiş AIC ağırlıklar değerlerini göstermek üzere, aday modellerinin model olma seviyeleri  $w_m$  aşağıdaki gibi hesaplanır.

$$w_m = \frac{\exp(-0.5\Delta_m)}{\sum_{j=1}^M \exp(-0.5\Delta_j)} \quad (3.47)$$

Daha detaylı bilgi için Akaike ve Schwarz'ın [62] çalışmasına başvurulabilir.

### 3.6.2 Schwarz Bayesci Bilgi Kriteri (BIC – Bayesian Information Criteria)

Schwarz tarafından AIC'e alternatif olarak geliştirilen diğer bir model belirleme kriteri Schwarz Bayesian kriteridir (BIC) [61][63]. BIC'in arkasında yatan temel mantık, model uzayındaki tüm olası modellerin posterior model olasılıklarını hesaplamak ve en yüksek posterior olasılık değerine sahip modeli seçmektir [62].

Modelin olasılık tahmini için, AIC hesaplamasından farklı olarak BIC bir ceza katsayısı kullanır. BIC

$$BIC = -2\log L(\hat{\theta}) - k\log n \quad (3.48)$$

ile bulunur.

Burada  $\hat{\theta}$  model parametrelerinin maksimum olabilirlik tahminini ve k modelde tahmin edilecek bağımsız parametre sayısını ifade eder. (3.48)'in birinci kısmı, daha fazla parametreye sahip daha büyük modelleri tercih eden maksimum olabilirlik sürecini, ikinci kısım, daha az parametrelili daha küçük modelleri tercih etme eğiliminde olan ceza terimini ifade eder. BIC'in amacı bir modelin sadeliği ile verilere olan tahmini arasında bir denge bulmaktır. Maksimum BIC değeri, aday modeller arasında en iyi modeli verir [62].

### 3.7 Model Artıkları (Residuals) Durağanlık Testleri

Model artıklarının durağanlığının testi için literatürde genel kabul gören birçok yöntem bulunmaktadır. Model artıklarına ilişkin örnek otokorelasyon katsayılarının incelenmesi bu yöntemlerden bir tanesidir [64].

k otoregresif süreçte gecikme sayısı olmak üzere otokorelasyon katsayısı; kovaryansın varyansa oranından,

$$\hat{\gamma}_k = E(e_t, e_{t-k}) \quad (3.49)$$

$$\hat{\gamma}_0 = E(e^2) = \sigma^2 \quad (3.50)$$

sırasıyla

$$\rho_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (3.51)$$

Şeklinde hesaplama -1 ile +1 arasında değer almaktadır. Otokorelasyon katsayısı,  $k = 0$  iken 1 değerini alır.  $k$  gecikme değerinin artmasıyla birlikte otokorelasyon katsayısı azalma eğilimi göstermektedir. Bu da ilk değerden itibaren model hatalarının durağanlığı hakkında bilgi sahibi olmamızı sağlar [64].

$\rho_k$  tamamen rassal, ise  $N$  örneklem sayısı için 0 ortalamalı,  $1/N$  varyanslı normal dağılıma sahiptir ve

$$\hat{\rho}_k \sim N(0, \frac{1}{\sqrt{N}}) \quad (3.52)$$

şeklinde bulunur.

Model hatalarının otokorelasyon özelliği sergileyip sergilemediği otokorelasyon grafiklerine bakılarak incelenebileceği gibi hipotez testleriyle de sınanabilir. Bu çalışmada, Box ve arkadaşları [64]-[66] tarafından geliştirilen ve kendi adlarıyla anılan Box-Pierce ve Ljung-Box testlerine yer verilecektir.

### 3.7.1 Box-Pierce Testi

Box-Pierce testi, Q olarak adlandırılmak üzere; test hipotezi ve test istatistiği

$H_0: \rho_1 = \rho_2 = \dots = \rho_n = 0$  Otokorelasyon yoktur, kalıntılar durağandır

$H_1: \rho_1 \neq \rho_2 \neq \dots \neq \rho_n \neq 0$  Otokorelasyon vardır, kalıntılar durağan değildir.

$$Q = N \sum_{k=1}^N \hat{\rho}_k^2 \quad (3.53)$$

ile ifade edilir.

Q test istatistiği  $(N - p - q)$  serbestlik dereceli bir ki-kare  $(\chi_{N-p-q}^2)$  dağılımına uymaktadır [64][65].

### 3.7.2 Ljung-Box Testi

Ljung-Box testi, LB olarak adlandırılmak üzere; test hipotezi ve test istatistiği

$H_0: \rho_1 = \rho_2 = \dots = \rho_n = 0$  Otokorelasyon yoktur, kalıntılar durağandır

$H_1: \rho_1 \neq \rho_2 \neq \dots \neq \rho_n \neq 0$  Otokorelasyon vardır, kalıntılar durağan değildir.

$$LB = N(N + 2) \sum_{k=1}^N \frac{\hat{\rho}_k^2}{n-k} \quad (3.54)$$

ile ortaya konur.

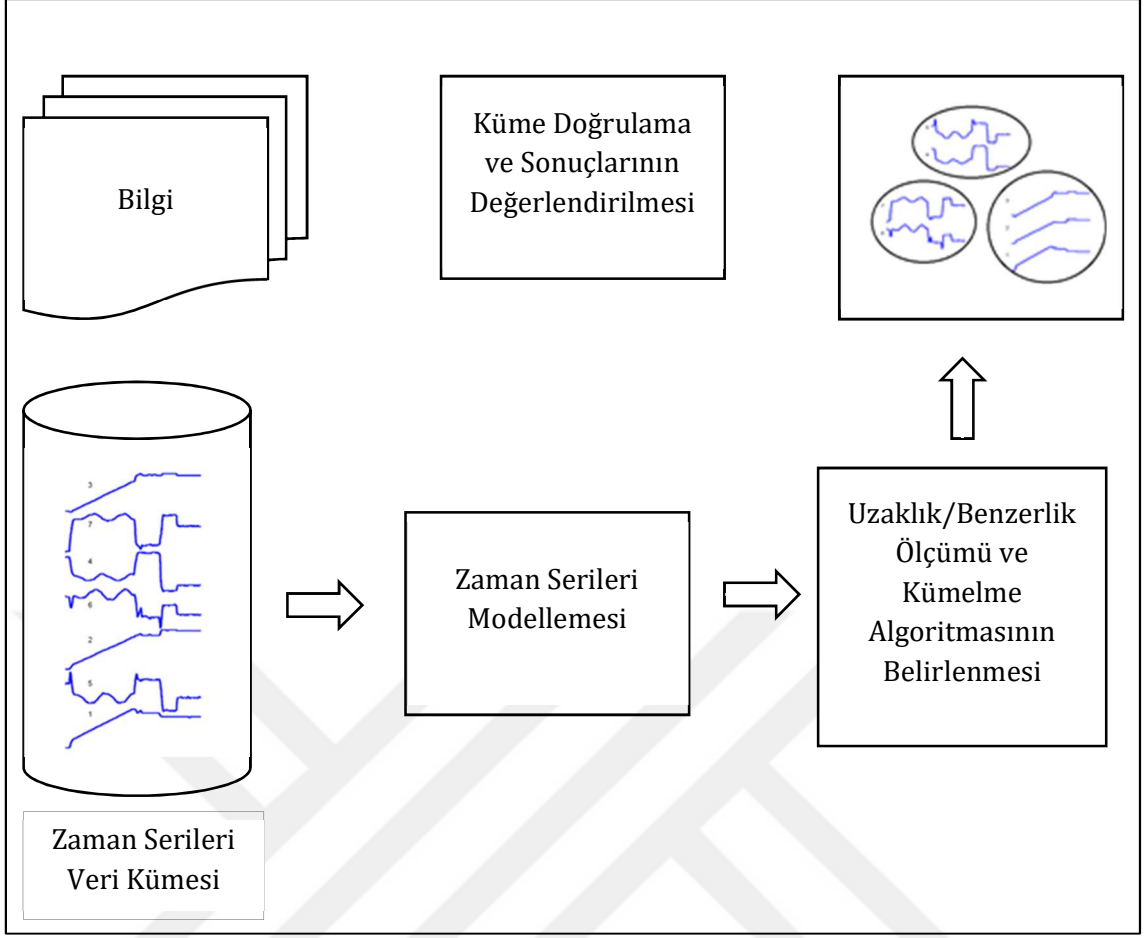
LB test istatistiği  $(N - p - q)$  serbestlik dereceli bir ki-kare  $(\chi_{N-p-q}^2)$  dağılımına uyar. uymaktadır. LB testi, Q testine kıyasla küçük örneklemlerinde daha etkili bir testtir [64][66].

## MODEL BAZLI ZAMAN SERİLERİ KÜMELEME YAKLAŞIMLARI

---

Model bazlı yaklaşımlar, her zaman serisinin belirli bir model veya olasılık dağılımlarının karışımı tarafından üretildiğini varsayarlar. Bu yaklaşımın ana fikri, her zaman serisi için ayrı ayrı model tahminlemesi yapmak ve daha sonra bu model parametreleri arasında benzerlik/uzaklık ölçümlerini dikkate alarak benzer model grupları bulmaktır. Benzer gruplar içerisinde yer alan bireysel zaman seri modellerinin yada tahminleme sonucu oluşan hata terimleri dağılımlarının benzer olduğu düşünülür [4]. Model bazlı zaman serileri kümelemenin adımları Şekil 4.1' de verilmiştir.

En sık kullanılan zaman serisi modelleri: polinom karma Modelleri ARIMA (Autoregressive Integrated Moving Average), Markov Zincirleri (Markov Chains) ve Gizli Markov Modelleri [67]. Bu çalışmada, otoregresif modeller, zaman serileri üzerinden öngörü yapmaya imkan tanınması, yorumlama kolaylığı ve hesaplama elverişliliği nedeniyle tercih edilmiştir. Buradaki önemli nokta: otoregresif modelleri kullanılarak yapılacak kümeleme çalışması ile değer bazlı bir ayrışım yerine, eğilim bazlı bir ayrışım hedeflenmektedir.



Şekil 4.1 Model Bazlı Zaman Serileri Kümeleme Adımları

Bu çalışmada, otoregresif model bazlı zaman serisi kümeleme yaklaşımı aşağıdaki adımlardan oluşmaktadır:

1. Otoregresif zaman serisi modellerinin tespiti
2. Uygun uzaklık/benzerlik ölçümü ve kümeleme algoritmasının seçimi
3. Küme sayısının belirlenmesi
4. Kümeleme ve küme sonuçlarının yorumlanması
5. Öngörü yapma ve performanslarının değerlendirilmesi

AR modeli, herhangi  $Y_t \in \mathcal{L}$  zaman serisi için öngörü serisi  $F_t = E(Y_t | y_{t-1}, Y_{t-2}, \dots)$  olmak üzere

$$F_t = \phi_{t-1}y_{t-1} + \phi_{t-2}y_{t-2} + \dots \quad (4.1)$$

ifade edilir.

$Y_t = F_t + \varepsilon_t, \forall t$  için  $F_t$  ve  $\varepsilon_t$  ortogonaldir.

$x_t \in \mathcal{L}$  ve  $y_t \in \mathcal{L}$  AR sürecinde tanımlanan iki zaman serisi ve  $p$  gecikme sayısı olmak üzere,

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t \quad (4.2)$$

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (4.3)$$

ile verilir.

AR katsayıları

$$\phi_x = (\phi_{1,x}, \phi_{2,x}, \dots, \phi_{i,x}, \dots); \quad \phi_y = (\phi_{1,y}, \phi_{2,y}, \dots, \phi_{i,y}, \dots) \quad (4.4)$$

olarak bulunur.

Uzaklık/benzerlik ölçümlerinden bir tanesi olan Piccola (Öklid) uzaklığı  $X_t$  ve  $Y_t$  AR süreçli zaman serileri için denklem (4.5)' deki gibi tanımlanır [24].

$$d(x_t, y_t) = [(\phi_x - \phi_y)'(\phi_x - \phi_y)]^{\frac{1}{2}} = \sqrt{\sum_{i=1}^{\infty} (\phi_{i,x} - \phi_{i,y})^2} \quad (4.5)$$

$d(x_t, y_t)$ ,  $x_t, y_t \in \mathcal{L}$  metrik uzayına ait, tüm zaman serisi süreçleri arasındaki yapısal farklılığı iyi tanımlayan, bir ölçüdür ve değeri karşılaştırılacak olan stokastik yapıların tüm bileşenleri tarafından belirlenir.

Buradaki Öklid mesafesinin temel özelliği: aynı başlangıç değerli zaman serileri kümeleri için, iki AR zaman süreci arasındaki mesafe, eğer karşılıklı modeller aynı öngörülerini üretiyorsa, sıfırdır.



Önceki bölümlerde, zaman serileri için önerilen otoregresif model bazlı zaman serileri kümeleme yaklaşımına ilişkin teorik yapı üzerinde durulmuş ve kümeleme metodolojisi incelenmiştir. Bu bölümde, önerilen otoregresif model bazlı zaman serileri kümeleme yaklaşımı bir gerçek veri kümesinde uygulanmış ve kümeleme sonuçları tartışılmıştır. Ek olarak, otoregresif model bazlı zaman serileri kümeleme sonuçları klasik ham veri bazlı kümeleme sonuçları ile de karşılaştırılmıştır.

### 5.1 Veri Kümesinin Açıklaması

Bu çalışmada Reidin.com Türkiye Konut Satış Fiyat Endeksi bir gerçek veri kümesi olarak kullanılmıştır. Endeksin amacı, belirli bir coğrafi bölgedeki konutların dönemsel satış fiyatlarını ve bunların zamana göre değişimini tespit etmektir [68].

Reidin.com Türkiye Konut Satış Fiyat Endeksi, Türkiye'deki konut satışları konusunda güvenilir ve tutarlı bir gösterge olması amacıyla Reidin.com tarafından geliştirilmiştir. Bu anlamda, Reidin.com satış fiyat endeksi Türkiye' nin en büyük 5 ilinin (Anakara, Bursa, İstanbul, İzmir ve Kocaeli) 196 mahallesi için düzenli ve aylık olarak hesaplanmaktadır [68].

Reidin.com satış fiyat endeksi her ay yaklaşık olarak 200.000 tane konut üzerinden, arz ve talep satış fiyat verileri kullanılarak hesaplanmaktadır. Veri kümesinde yer alan satış fiyatları gayrimenkul ofislerinden, internet sitelerinden, gazetelerden, dergilerden ve varlık yönetim şirketlerinden temin edilmektedir. Veri kümesi ayrıca geçmişte ya da hali hazırda içerisinde ikamet edilen müstakil evleri de kapsamaktadır. Veriler her ay düzenli olarak mantık ve doğruluk kontrolünden geçirilerek titizlikle hazırlanmaktadır [68].

Endeks oluşturulmasında, aylık periyotlarda, mahalledeki medyan konut satılık değerleri kullanılmaktadır. Burada medyan yönteminin kullanılmasının sebebi, aykırı değerlerin yanıltıcı sonuçlar verebilmesinin önüne geçmektir. Aylık

periyotlarda hesaplanan medyan satılık fiyat değerleri hesaplandıktan sonra, sabit esaslı basit fiyat endeksi yöntemiyle mahalle bazlı Reidin.com Türkiye Konut Satış Fiyat Endeksi hesaplanmaktadır.

$p_t$  , t inci aydaki metrekare medyan fiyat değerini ve  $p_0$ , baz alınan temel dönemdeki metrekare medyan konut fiyat değerini göstermek üzere sabit fiyatlı basit endeks

$$I_{t/0} = \frac{p_t}{p_0} \times 100 \quad (5.1)$$

ile hesaplanır. Oransal kıyaslamayı kolaylaştırması açısından endeks sonucu 100 ile çarpılır.

Veri kümesi, ulaşılabilirlik sebebiyle, 2010 Ocak – 2017 Şubat dönemlerini kapsayacak şekilde 96 aylık zaman gözleminden oluşmaktadır. Veri kümesi, model parametrelerinin tahmin performanslarının incelenebilirliği için farklı zaman periyotlarını içeren iki veri kümesine ayrılmıştır: %75 eğitim (2010 Ocak – 2016 Aralık), %25 test (2017 Ocak – 2017 Aralık).

Tablo Ek A.1, 196 mahalleye ilişkin olarak mahallelerin içinde bulunduğu il ve ilçe isimlerini; endeks hesaplamasında kullanılan temel dönem bilgilerini ve konut satış fiyat endekslerine ilişkin minimum, maksimum ve standart sapma değerlerini vermektedir.

Bu tezde, model bazlı zaman serileri kümeleme yaklaşımı aşağıdaki adımlardan oluşmaktadır:

1. Her bir zaman serisi için ayrı ayrı otoregresif modellerin tespiti
2. Uygun uzaklık/benzerlik ölçümüyle kümeleme algoritmasının seçimi
3. Küme sayısının belirlenmesi
4. Kümeleme ve küme sonuçlarının yorumlanması
5. Öngörü performanslarının değerlendirilmesi

## 5.2 Otoregresif Zaman Serisi Modellerinin Geliştirilmesi

Bu alt bölümde, mahalle bazlı 196 konut satış fiyat endeksi için bireysel otoregresif (AR) modeller tahminlenmiştir. Modelleme aşamasında Box-Jenkins yaklaşımı benimsenmiştir. Box-Jenkins yaklaşımına göre, zaman serisi modelleri için, her bir zaman serisinin ön şart olarak durağanlık koşulunu sağlaması gerekmektedir. Durağanlık olmayan zaman serileri üzerinden yapılan bir çalışmada zaman serilerini doğru tahminleyen modellerin bulunması dolayısı ile de güvenilebilir öngörü sonuçlarının elde edilmesi mümkün olmamaktadır [56].

Box-Jenkins yaklaşımından hareketle, öncelikle, mahalle bazlı 196 konut satış fiyat endeksinin durağanlığı incelenmiştir. Durağanlığın tespiti için her bir zaman serisi ADF, PP ve KPSS birim kök kriterleri ile test edilmiştir. 196 endeksin bireysel olarak bu üç birim kök testiyle eşanlı sınanabilmesi için R ortamında otomatik durağanlık sınama kodları geliştirilmiştir. Bu R kodları Ek B.1' de paylaşılmıştır.

Tablo B.2.1' de sunulan ADF, PP ve KPSS birim kök test sonuçlarına göre satış fiyat endeksleri sıfırıncı mertebeden, yüzde 5 anlamlılık düzeyinde, durağan çıkmamıştır. Bu durumda endeks verilerinin durağanlaştırılması gerekmektedir. Fark alma  $\Delta$  ( $\Delta y_t = y_t - y_{t-1}$ ) operatörü durağan olmayan zaman serilerini durağan hale getirebilmektedir. Gecikmeli bir AR modeli için birinci mertebeden fark alma işlemi

$$\Delta Y_t = \phi_1 \Delta y_{t-1} + \varepsilon_t \quad (5.2)$$

ile gerçekleştirilir.

Tablo B.3.1, birinci mertebeden fark alma işlemiyle elde edilen zaman serileri için ADF, PP ve KPSS birim kök test sonuçlarını vermektedir. Tablo B.3.1 sonuçlarına göre, fark alma işlemi 196 konut satış fiyat endeksini 0.05 anlamlılık düzeyinde durağan hale getirmektedir. Burada, endeks verilerinin en az herhangi iki birim kök test sonucuna göre durağan çıkması; endeks verilerini durağan olarak kabul etmemiz için yeterlidir.

Fark alma işlemi ile durağan hale gelen endeks zaman serileri, AR modelleme sürecine hazır hale gelmektedir. Her bir zaman serisi için ayrı ayrı AR model

parametre sayılarının ve katsayılarının optimum düzeyde tahmin edilmesi zaman açısında verimli olmayabilmektedir. Bu sebepten ötürü, otomatik olarak her bir zaman serisi modelinin belirlenmesi verimlilik ve etkinlik açısından önem arz etmektedir. Bu bağlamda, bu çalışmada, R açık kaynak kodlu bilgisayar programlama dili vasıtasıyla, zaman serileri için bir AR otomatik model belirleme yaklaşımı geliştirilmiştir. Bu yaklaşım, en düşük AIC değerini bulmayı hedefleyerek, optimum AR model parametre sayılarını tahminlemektedir. Ek C' de, n adet durağan zaman serisi için geliştirilen otomatik AR model belirleme R kodları paylaşılmıştır. Bu R kodları sabit terimi içermeyecek şekilde maksimum 10 AR katsayısı bulacak şekilde tasarlanmıştır.

Otomatik model belirleme yaklaşımı ile 0.05 anlamlılık düzeyinde, en fazla 7 AR katsayısı bulunabilmektedir. Böylece, en fazla 196x7' lik bir veri matrisine ulaşılmaktadır. Sonuç olarak, her bir zaman serisinin 7 AR katsayısı ile giriş veri setinin boyutu azaltılmaktadır.

$\Delta Y_t$ , birinci mertebeden farkı alınan bir zaman serisini göstermek üzere i. zaman serisi için belirlenen 7 gecikmeden oluşan bir AR modeli,

$$\Delta Y_{i,t} = \hat{\beta}_1 \Delta Y_{i,t-1} + \hat{\beta}_2 \Delta Y_{i,t-2} + \hat{\beta}_3 \Delta Y_{i,t-3} + \hat{\beta}_4 \Delta Y_{i,t-4} + \hat{\beta}_5 \Delta Y_{i,t-5} + \hat{\beta}_6 \Delta Y_{i,t-6} + \epsilon_t \quad (5.3)$$

ile gösterilmektedir.

$\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$  olmak üzere birinci mertebeden farkı alınan bir zaman serisi normal formuna

$$Y_t - Y_{t-1} = \hat{\beta}_1 (Y_{t-1} - Y_{t-2}) + \hat{\beta}_2 (Y_{t-2} - Y_{t-3}) + \hat{\beta}_3 (Y_{t-3} - Y_{t-4}) + \hat{\beta}_4 (Y_{t-4} - Y_{t-5}) + \hat{\beta}_5 (Y_{t-5} - Y_{t-6}) + \hat{\beta}_6 (Y_{t-6} - Y_{t-7}) + \epsilon_t \quad (5.4)$$

ulaşılır. Bu denklem tekrar düzenlenirse

$$Y_t = (1 + \hat{\beta}_1)Y_{t-1} + (\hat{\beta}_2 - \hat{\beta}_1)Y_{t-2} + (\hat{\beta}_3 - \hat{\beta}_2)Y_{t-3} + (\hat{\beta}_4 - \hat{\beta}_3)Y_{t-4} + (\hat{\beta}_5 - \hat{\beta}_4)Y_{t-5} + (\hat{\beta}_6 - \hat{\beta}_5)Y_{t-6} - \hat{\beta}_6 Y_{t-7} + \epsilon_t \quad (5.5)$$

elde edilir.

Burada (5.5) denkleminde  $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6$  and  $\hat{\beta}_7$  sıradan en küçük kareler yöntemi ile tahminlenen model parametre değerlerini ve  $\epsilon_t$  varsayım gereği pür rassallık özelliği gösteren model hata terimini göstermektedir.

n tane durağan zaman sersi için, otomatik AR modelleri belirleme sürecinde, tespit edilen modellere ilişkin eşanlı olarak AR model hata terimlerinin pür rassallık özelliği sergileyip sergilemediği test edilmiştir. Test işlemi için Ljung-Box and Box-Pierce testleri kullanılmıştır. Test sonuçlarına göre 196 gayrimenkul endeksi için belirlenen AR model hata terimleri, 0.05 anlamlılık düzeyinde, pür rassallık özelliği göstermektedir.

### **5.3 Uzaklık Benzerlik Ölçüsünün ve Kümeleme Algoritmasının Belirlenmesi**

Bu alt bölümde, eğitim seti 2010 Ocak – 2016 Aralık dönemlerinden oluşan 196 mahalle bazlı konut satış fiyat endeksleri için tahminlenen AR model parametreleri üzerinden kümeleme çalışması gerçekleştirilmiştir. Bu çalışmada Öklid, Mahalanobis, Cityblock, Minkowski, Chebychev ve Cepstral bazlı uzaklık ölçüsü ve açısal ve korelasyon benzerlik ölçümleri; kümeleme algoritması olarak tek bağlantılı, tam bağlantılı, grup ortalama bağlantılı, ağırlıklı ortalama bağlantılı, merkezi bağlantılı, medyan bağlantılı, Ward hiyerarşik, K-ortalamlar, özdenetimli haritalar (SOM) ve bulanık C-ortalamlar kümeleme yöntemleri kullanılmıştır. 8 uzaklık benzerlik ölçümü ile 10 kümeleme algoritması denenerek toplamda 80 farklı kümeleme çalışması yapılmıştır.

Aynı veri seti üzerinde, farklı kümeleme algoritmalarıyla farklı sonuçlar bulunabilmektedir. Hatta farklı uzaklık benzerlik ölçümleri ile aynı kümeleme algoritmaları kullanılarak farklı sonuçlar da elde edilebilmektedir. Bu sebeplerden dolayı, kümeleme çalışmalarında uygun uzaklık benzerlik ölçümü ve kümeleme yönteminin seçilmesi kritik bir önem taşımaktadır [13][40][69].

Kümeleme algoritmalarının maksimum performansını sağlayan en uygun ve tutarlı uzaklık/ benzerlik ölçümünün seçilebilmesi adına, yukarıda bahsi geçen 10 kümeleme algoritması ve 8 uzaklık/benzerlik ölçümü kullanılarak veri kümesi sırasıyla 2, 3, 4, 5, 6, 7, 8, 9 ve 10 kümeğe ayrılmıştır. Kümeleme sürecinde,

sonuçların tutarlılığının kıyaslanması adına Silhouette, Davies-Bouldin, Calinski-Harabasz, Dunn, R-Kareler, C-Endeksi, Krzanowski-Lai, Hartigan, Homojenlik ve Ayrışma küme doğrulama endeksleri birlikte kullanılmıştır. Hesaplama işlemleri MATLAB Cluster Validity Analysis Platform'unda hesaplanmıştır [70]. Endekslerin kümeleme başarısını maksimum düzeyde sağlayan optimum seviyeler dikkate alınmıştır. Buna göre elde edilen sonuçlar Tablo 5.1' de paylaşılmıştır.

Tablo 5.1, farklı küme sayıları için (2, 3, 4, 5, 6, 7, 8, 9 ve 10 küme sayısı) bulunan en uygun uzaklık/benzerlik ölçümlerinin görülme sıklıklarını dikkate alarak, nihayetinde kümeleme algoritmalarının en iyi ve tutarlı sonucu verdiği düşünülen uzaklık/benzerlik ölçüsünü göstermektedir. Buna göre tek, ortalama, ağırlıklı ortalama, merkezi, medyan bağlantılı ve Ward hiyerarşik yöntemlerine göre en uygun uzaklık/benzerlik ölçümleri sırasıyla öklid, chebychev, cityblock, chebychev, chebychev, mahalonobis, korelasyondur. Diğer taraftan, K-ortalamlar, bulanık C-ortalamlar ve özdüzenleyici haritalar kümeleme algoritmaları için en uygun uzaklık/benzerlik ölçümleri sırasıyla Chebychev, Öklid ve Chebychev'dir. Sonuçlardan anlaşılacağı üzere, farklı küme doğrulama endekslerine göre farklı uzaklık/benzerlik ölçümleri seçilmektedir. Benzer şekilde, aynı küme doğrulama endeksi için bile kümeleme algoritmaları arasında tutarlı bir sonuç bulunmamaktadır.

**Tablo 5.1** Farklı Kümeleme Algoritmaları İçin En Uygun Uzaklık Benzerlik Ölçüleri

Endeksler / Kümeleme Algoritmaları	Tek Bağlantılı	Tam Bağlantılı	Ortalama Bağlantılı	Ağırlıklı Ortalama Bağlantılı	Merkezi Bağlantılı	Medyan Bağlantılı	Ward Yöntemi	K-Ortalamalar	Bulanık C-Ortalamalar	Özdüzenleyici Haritalar
Silhouette	Öklid	Cityblock	Öklid	Chebychev	Öklid	Öklid	Açısal Benzerlik	Açısal Benzerlik	Cityblock	Korelasyon
Davies-Bouldin	Minkowski	Chebychev	Chebychev	Chebychev	Chebychev	Chebychev	Chebychev	Chebychev	Öklid	Öklid
Calinski-Harabasz	Öklid	Chebychev	Cityblock	Chebychev	Chebychev	Korelasyon	Korelasyon	Chebychev	Öklid	Minkowski
Dunn	Chebychev	Cityblock	Cityblock	Chebychev	Chebychev	Cityblock	Korelasyon	Açısal Benzerlik	Açısal Benzerlik	Öklid
R-Kareler	Öklid	Chebychev	Cityblock	Chebychev	Chebychev	Açısal Benzerlik	Korelasyon	Chebychev	Açısal Benzerlik	Minkowski
C-Endeks	Chebychev	Chebychev	Cityblock	Chebychev	Chebychev	Öklid	Cityblock	Öklid	Öklid	Chebychev
Krzanowski-Lai	Korelasyon	Chebychev	Açısal Benzerlik	Öklid	Chebychev	Mahalanobis	Mahalanobis	Chebychev	Öklid	Chebychev
Hartigan	Açısal Benzerlik	Chebychev	Chebychev	Chebychev	Minkowski	Öklid	Korelasyon	Mahalanobis	Açısal Benzerlik	Korelasyon
Homojenlik	Mahalanobis	Mahalanobis	Mahalanobis	Mahalanobis	Mahalanobis	Mahalanobis	Mahalanobis	Mahalanobis	Mahalanobis	Öklid
Ayrışma	Minkowski	Açısal Benzerlik	Cityblock	Açısal Benzerlik	Minkowski	Mahalanobis	Korelasyon	Açısal Benzerlik	Minkowski	Açısal Benzerlik
En Sık Rastlanan	Öklid	Chebychev	Cityblock	Chebychev	Chebychev	Mahalanobis	Korelasyon	Chebychev	Öklid	Minkowski

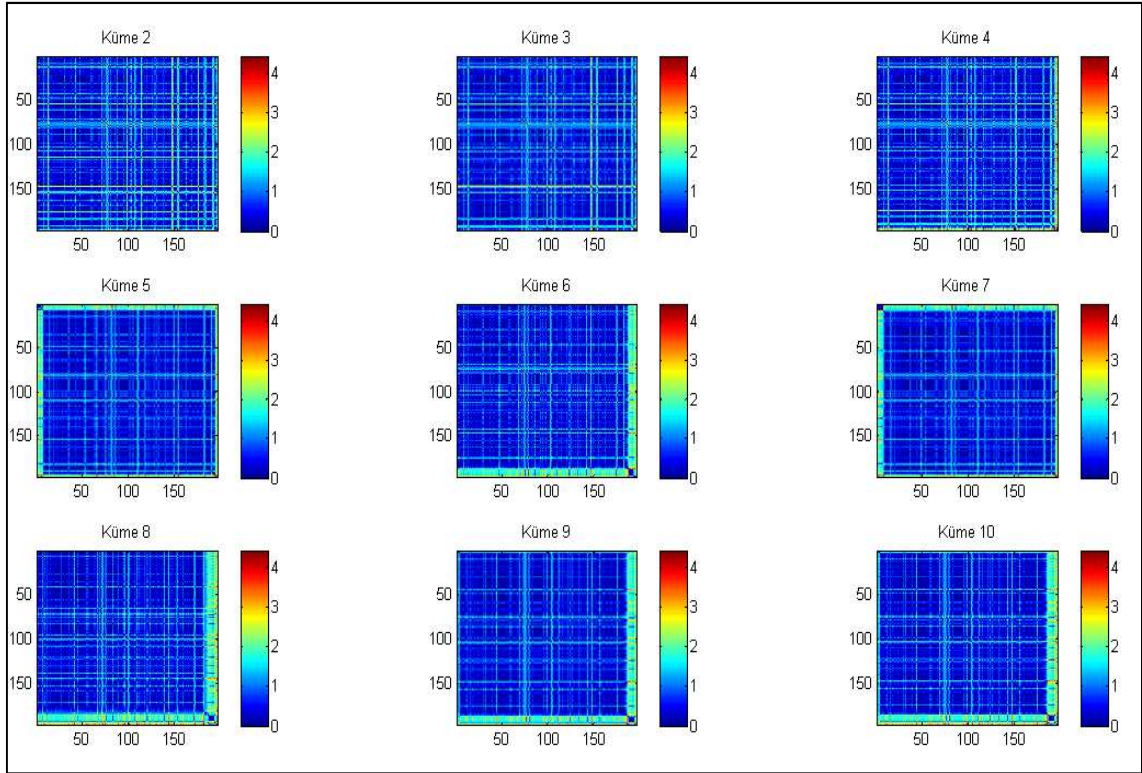
Literatürde yapılan çalışmalara göre, sayısal küme doğrulama endeksleri her durumda, özellikle endeks özelinde gerekli varsayımların karşılanmaması durumlarında en ideal çözümü her zaman sağlayamamakta ve kendi içerisinde tutarsız sonuçlar verebilmektedir [13][40][52][69].

Küme doğrulama endekslerinin ideal olmayan küme biçimleri için bazı dezavantajları bulunmaktadır. Bu teknikler kümeleme sonuçlarını değerlendirirken, kümeleri temsil edebilecek noktaları referans noktalar olarak, veri birimlerinin referans noktalarına olan uzaklıklarını ve referans noktalarından hareketle varyans gibi parametrelerini hesaplarlar. İdeal olmayan küme biçimlerini temsil edebilecek referans noktalarının seçilmesindeki zorluklardan dolayı bu tür endekslerle ideal olmayan küme biçimlerinin değerlendirilmesi her zaman doğru olmayabilir. Sayısal küme doğrulama endekslerine alternatif olarak, bu tezde görsel küme doğrulama (sayısal olmayan - Visual Cluster Validity, VCV) yöntemleri de kullanılmıştır. MATLAB programının sağladığı görselleştirme araçları yardımı ile görsel küme doğrulama yöntemleri için MATLAB ortamında kodlar üretilmiştir [13][40][52][68][69].

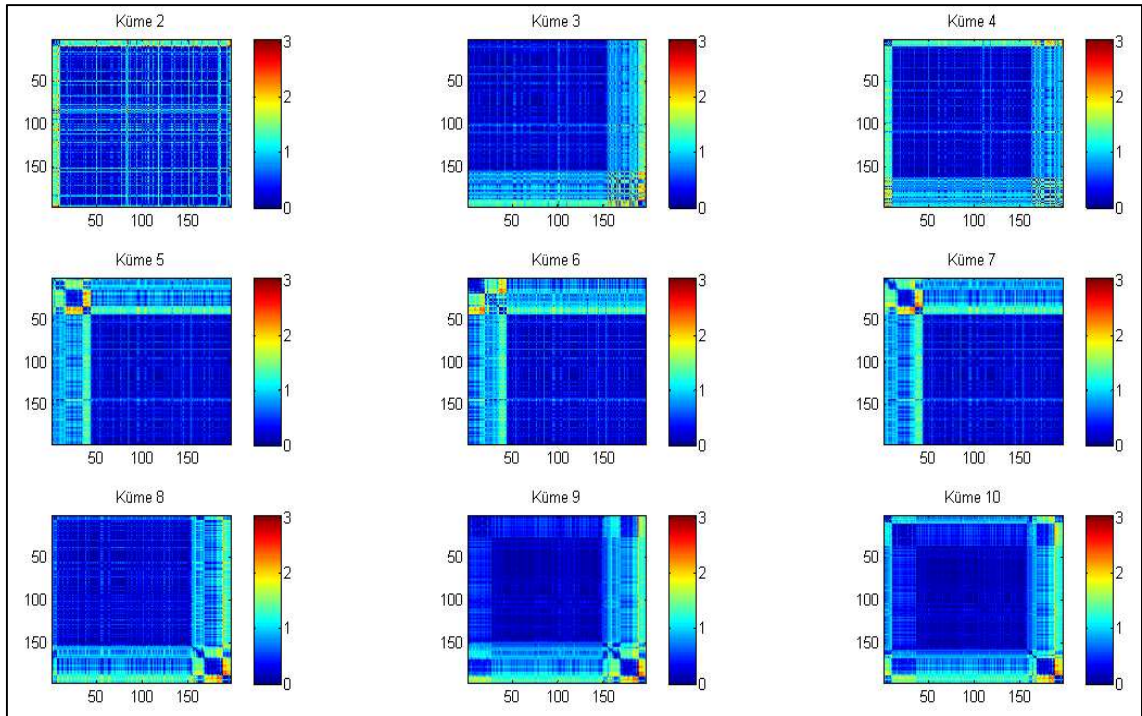
Bu çalışmada, VCV yöntemi küme doğrulama endekslerine alternatif bir yöntem olarak kullanılmıştır. Şekil 5.1' deki matris grafiği, Öklid uzaklığına dayalı tekil bağlantılı hiyerarşik kümeleme yöntemiyle tespit edilen 2, 3, 4, 5, 6, 7, 8, 9 ve 10 küme için kümelememe sonuçlarını VCV yöntemine göre göstermektedir. Benzer şekilde, Şekil 5.2 Chebychev uzaklığına dayalı tam bağlantılı, Şekil 5.3 Cityblock uzaklığına dayalı ortalama bağlantılı, Şekil 5.4 Chebychev uzaklığına dayalı ağırlıklı ortalama, Şekil 5.5 Chebychev uzaklığına dayalı merkezi bağlantılı, Şekil 5.6 Mahalanobis uzaklığına dayalı medyan bağlantılı, Şekil 5.7 korelasyon benzerliğine dayalı Ward, Şekil 5.8 Chebychev uzaklığına dayalı K-ortalamlar, Şekil 5.9 Öklid uzaklığına dayalı bulanık C-ortalamlar ve Şekil 5.10 Chebychev uzaklığına dayalı özdüzenleyici haritalar kümeleme sonuçlarını VCV yöntemine göre göstermektedir. Burada koyu mavi renkler konut satış fiyat endekslerinin birbirine benzerliğini gösterirken, kırmızı renkler birbirlerine benzemezliğini göstermektedir. Böylece, renk araçları kullanılarak kümeleme performansı diğer bir deyişle konut satış fiyat davranışlarının kendi içinde homojen, kendi aralarında



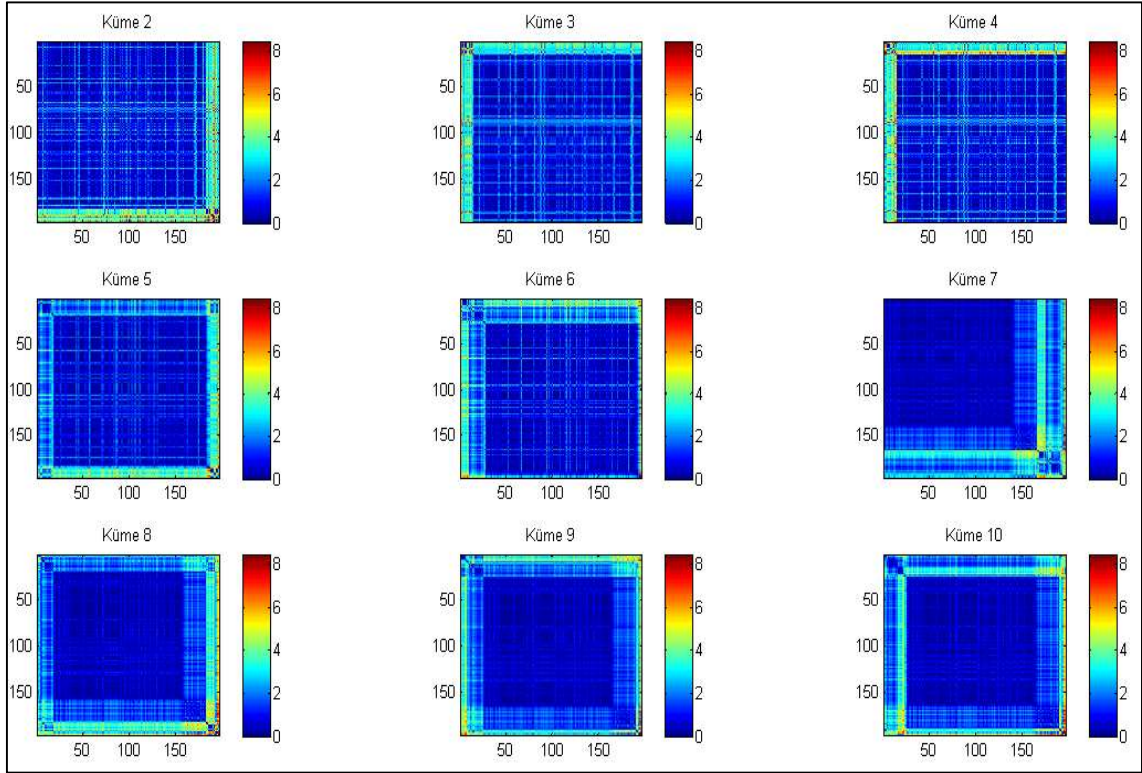
heterojen bir yapıda kümelenip kümelenmediği görsel bir şekilde ortaya konmaktadır.



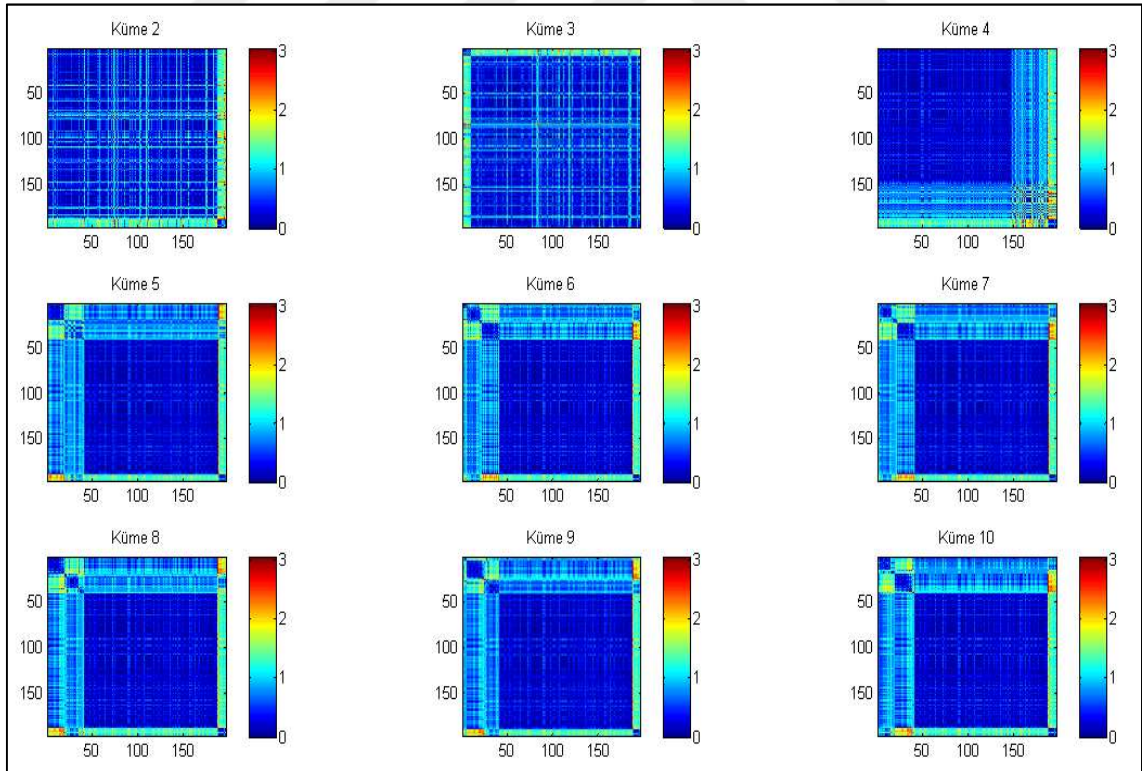
Şekil 5.1 Öklid Uzaklığına Dayalı Tekil Bağlantılı VCV Sonuçları



Şekil 5.2 Chebychev Uzaklığına Dayalı Tam Bağlantılı VCV Sonuçları

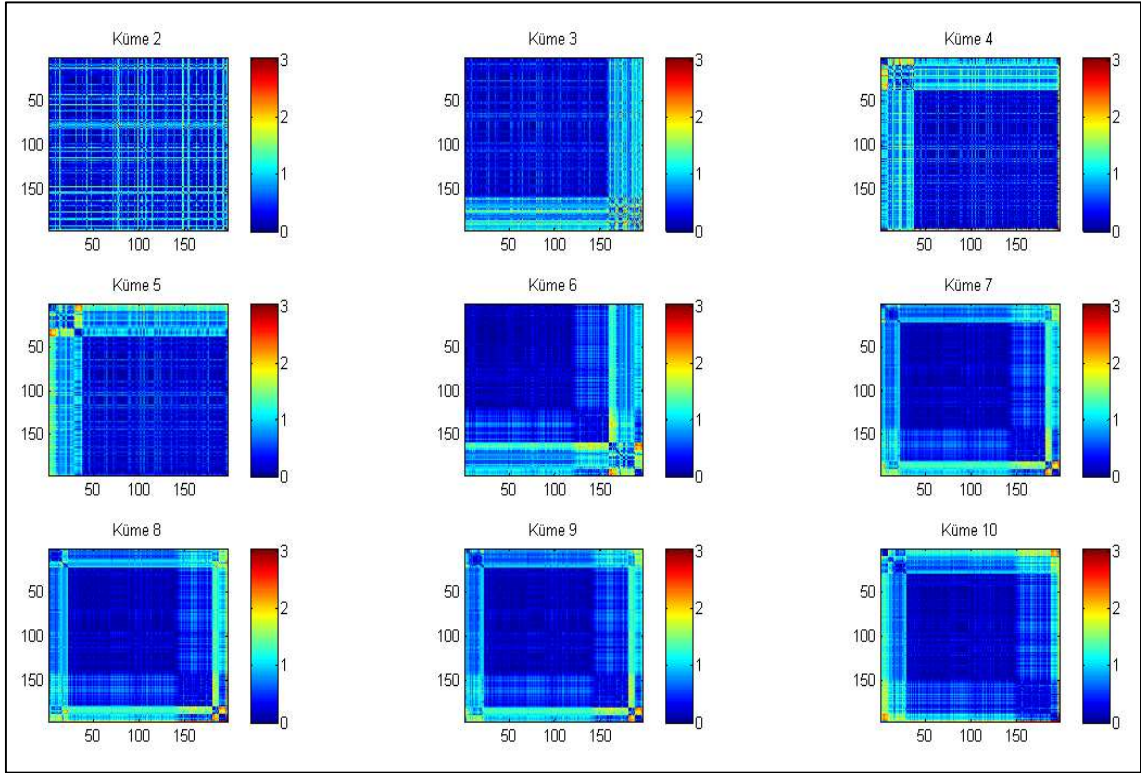


Şekil 5.3 Cityblock Uzaklığına Dayalı Ortalama Bağlantılı VCV Sonuçları

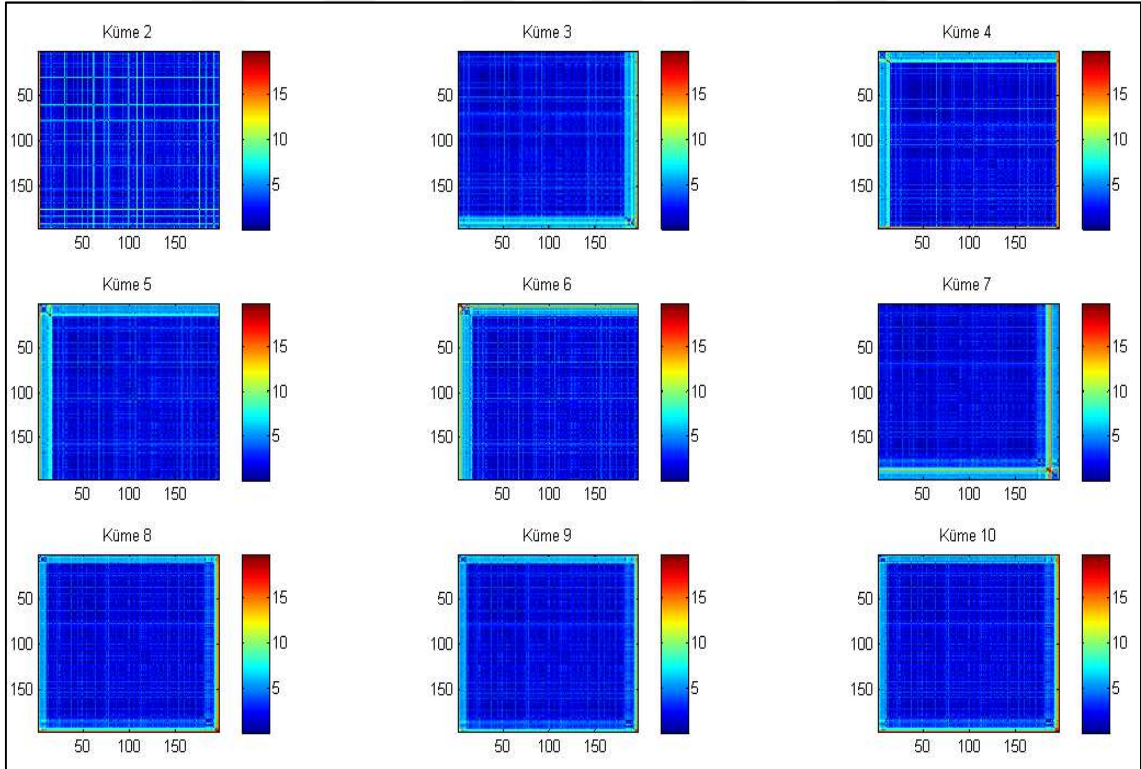


Şekil 5.4 Chebychev Uzaklığına Dayalı Ağırlıklı Ortalama Bağlantılı VCV Sonuçları

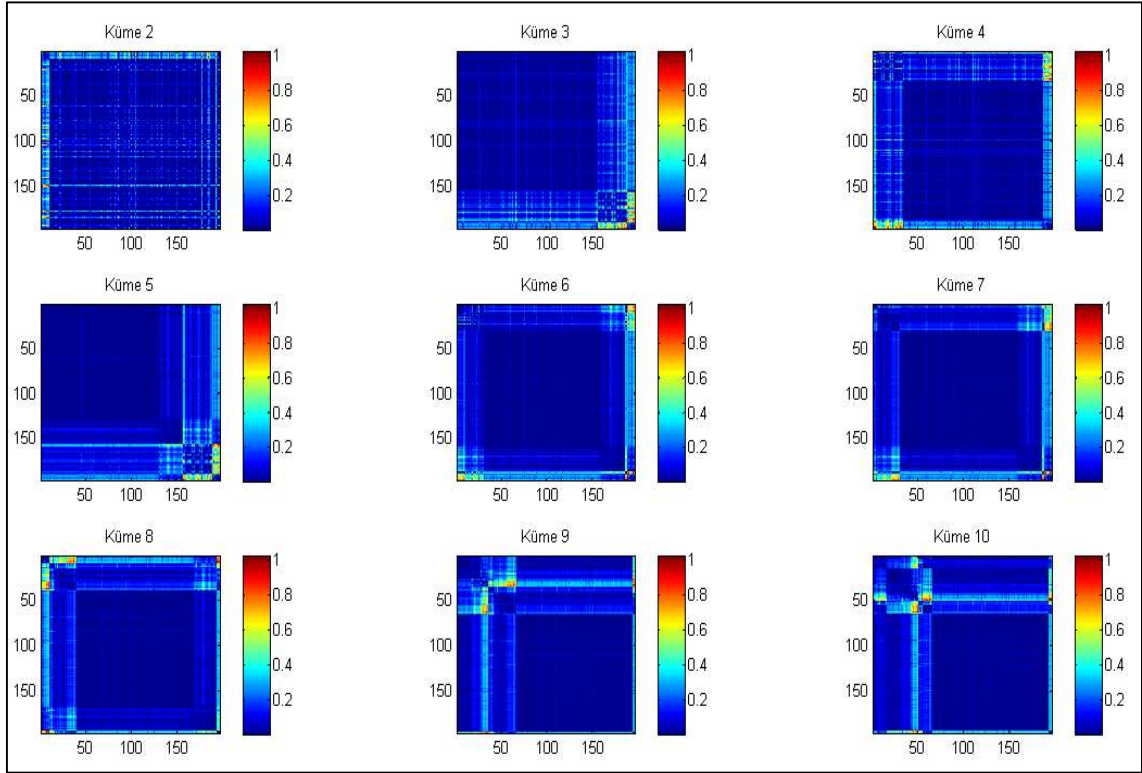




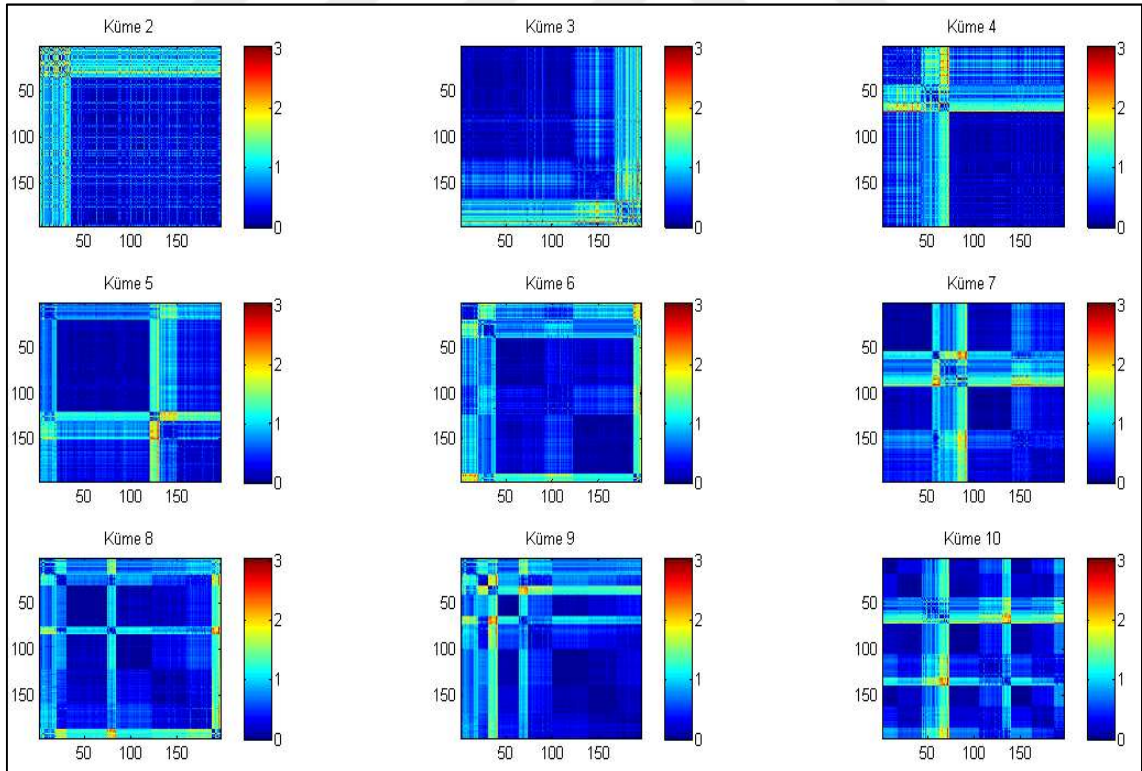
Şekil 5.5 Chebychev Uzaklığına Dayalı Merkezi Bağlantılı VCV Sonuçları



Şekil 5.6 Mahalanobis Uzaklığına Dayalı Medyan Bağlantılı VCV Sonuçları

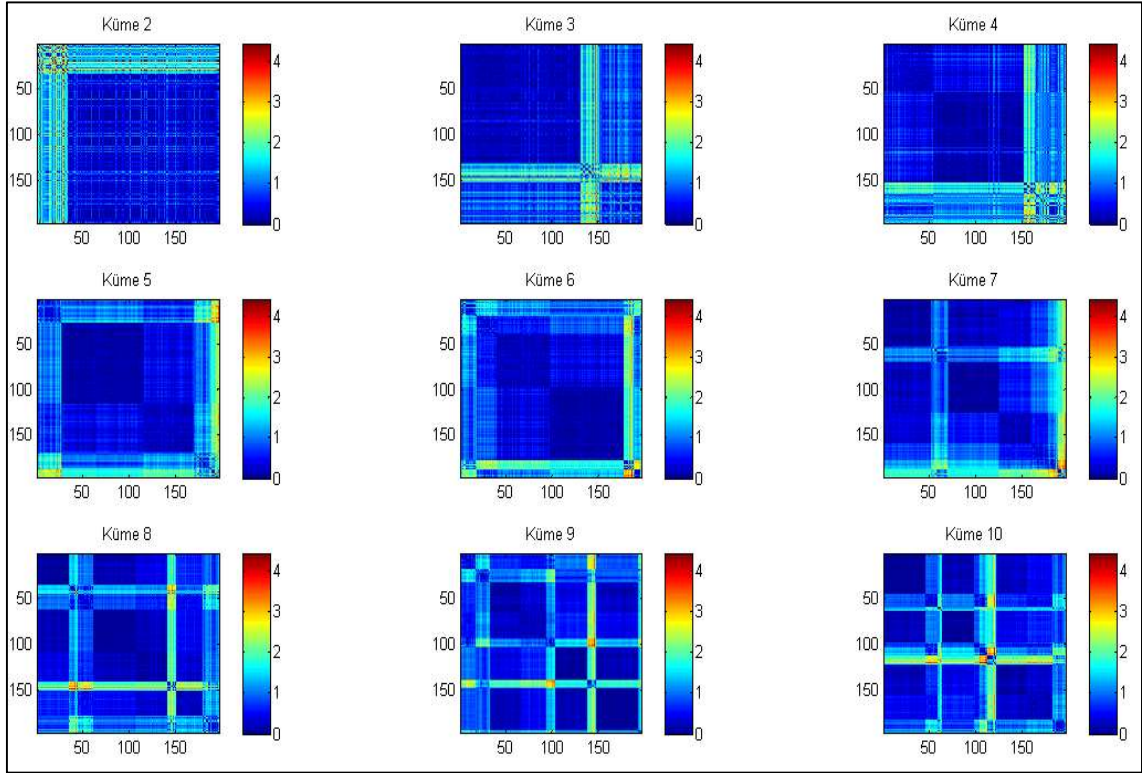


Şekil 5.7 Korelasyon Benzerliğine Dayalı Ward VCV Sonuçları

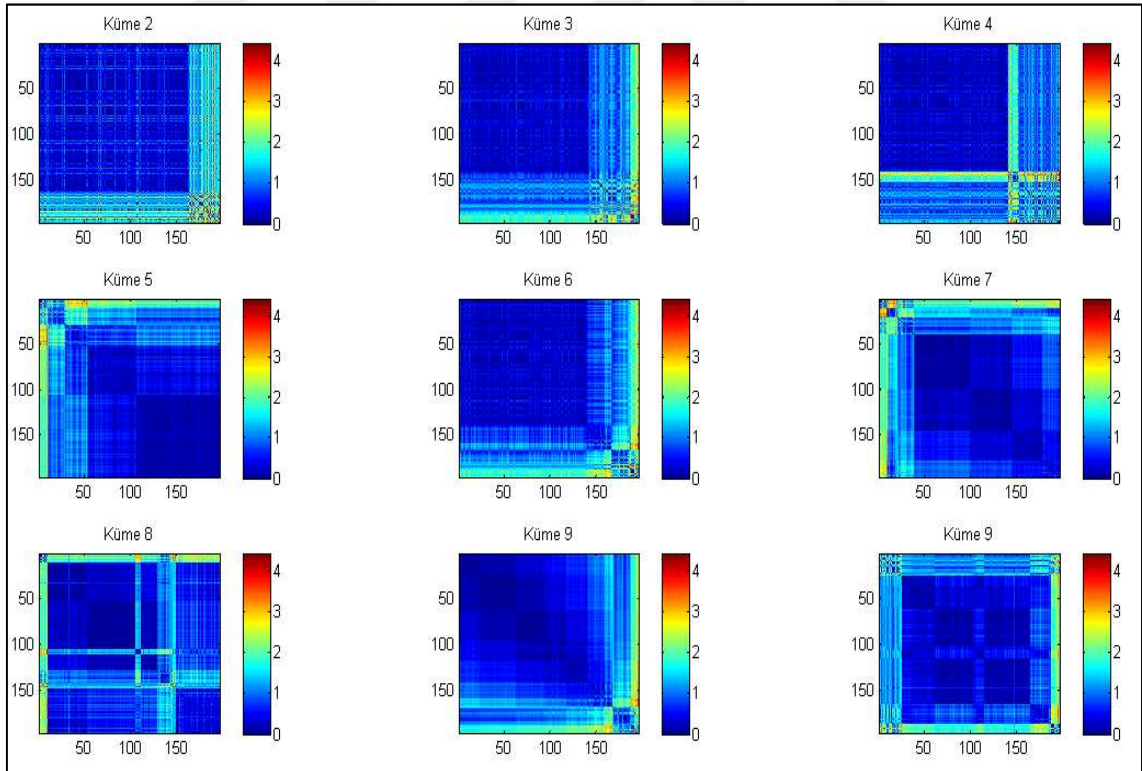


Şekil 5.8 Chebychev Uzaklığına Dayalı K-Ortalamlar VCV Sonuçları





Şekil 5.9 Öklid Uzaklığına Dayalı Bulanık C-Ortalamalar VCV Sonuçları



Şekil 5.10 Minkowski Uzaklığına Dayalı Özdüzenleyici Haritalar VCV Sonuçları

Küme doğrulama endekslerinin, farklı kümeleme algoritmaları için, optimum seviyelerinde belirlediği uzaklık/benzerlik ölçüleri sonuçlarının görsel doğrulama yaklaşımına göre değerlendirilmesi neticesinde; Öklid uzaklık ölçüsü ile kullanılan bulanık C-ortalamlar kümeleme yönteminin tahminlenen AR model katsayılarından oluşan veri kümesi için kümeler içi homojenliği en yüksek, kümeler arası heterojenliği en yüksek küme sonuçlarını verdiği görülmektedir.

Sonuç olarak, Silhouette, Davies-Bouldin, Calinski-Harabasz, Dunn, R-Kareler, C-Endeks, Krzanowski-Lai, Hartigan, Homojenlik ve Ayrışma endeks sonuçları ve görsel küme doğrulama yaklaşımlarına göre veri setimize en uygun, kümeleme algoritmasının bulanık C-ortalamlar olduğu sonucuna varılmıştır. Uzaklık ölçüsü olarak Öklid uzaklığı seçilmiştir.

#### **5.4 Küme Sayısının Belirlenmesi**

Kümeleme analizlerinde uygun küme sayısının tespiti kümeleme işleminin ayrılmaz bir parçasıdır. Geleneksel kümeleme yöntemleri, veri setinde yer alan doğal küme sayıları hakkında herhangi bir ön bilgiye sahip olmadan kümeleme işlemlerini gerçekleştirirler. Burada; sezgisel kararlara dayanılarak karar verilen küme sayıları ölçüsünde kümeleme algoritmaları istenilen sayıda küme bulacaktır. Diğer taraftan, bir veri setinde küme yapıları olmasa bile kümeleme algoritmaları bu veri setinde istenilen sayıda küme tespit edecektir. Bu alt bölümde, görsel ve analitik yöntemler birlikte kullanılarak veri seti içerisindeki doğal küme sayılarının tespiti hedeflenmektedir.

196 adet gayrimenkul satış fiyat endeksi için tahminlenen AR model parametreleri, düzlemde iki noktanın birbirine göre uzaklığını ifade eden Öklid uzaklığının hesaplaması için kullanılmıştır. Bulanık C-ortalamlar kümeleme yöntemi, bu 10x196 Öklid uzaklık matrisi için çalıştırılmıştır. Öklid uzaklık matrisi en uygun sayıda küme sayısının tespiti için sırasıyla 2, 3, 4, 5, 6, 7, 8, 9 ve 10 kümeye ayrılmıştır. Literatürde en çok tercih edilen Silhouette, Davies-Bouldin, Calinski-Harabasz, Dunn, R-Kareler, C-Endeks, Krzanowski-Lai, Hartigan, Homojenlik ve Ayrışma küme doğrulama endeksleri küme sayısının tespiti için kullanılmıştır. Tablo 5.2, MATLAB İstatistik Araç Kutusu (MATLAB Statistics Toolbox) ve küme

doğrulama platformundan elde edilen küme doğrulama endeks sonuçlarına göre en uygun küme sayılarını göstermektedir [3].

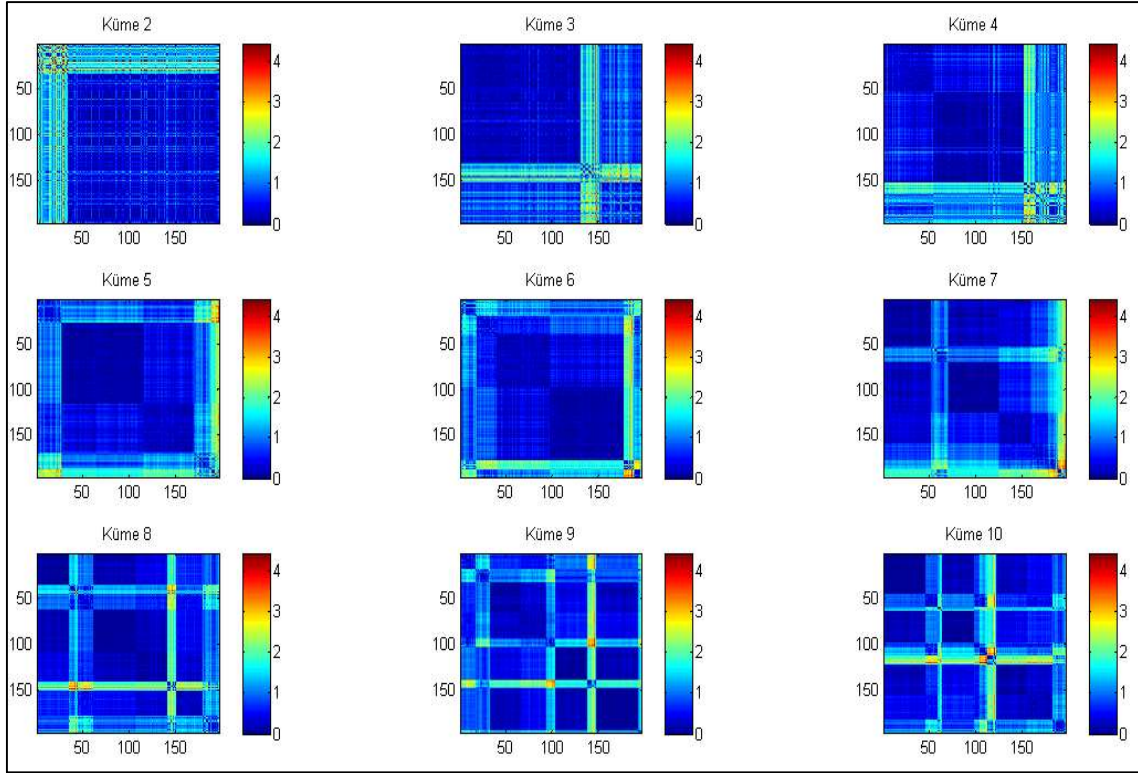
**Tablo 5.2** En Uygun Küme Sayısı

Küme Doğrulama Endeksleri	En Uygun Küme Sayısı
Silhouette	2
Davies-Bouldin	3
Calinski-Harabasz	3
Dunn	3
R-Kareler	10
C-Endeks	5
Krazanowki-Lai	5
Hartigan	3
Homojenlik	2
Ayrışma	2

Tablo 5.2' den de görüldüğü üzere farklı deney sonuçları için hesaplanan küme doğrulama endeksleri, en uygun küme sayısı bulma konusunda ortak sonuçlar vermemektedir. Veri kümesi içerisinde Silhouette, Homogeneity ve Separation iki küme bulmakta iken, Davies-Bouldin, Calinski-Harabasz, Dunn ve Hartigan endeksleri üç küme bulmaktadır. Diğer taraftan C-Endeks ve Krazanowski-Lai endeksleri 5 küme ve R-Kareler 10 küme bulmaktadır. Çoğunluk itibariyle endeksler veri setinde 2 ya da 3 küme bulunabileceğini söylemektedir. Anlaşılacağı üzere, kümeleme işleminde doğru küme sayısı hala soru işaretidir. Bezdek ve Hataway'e göre görsel küme doğrulama yöntemleri aracılığı ile doğru küme sayısı, insan görme sisteminin de devreye sokulmasıyla, sayısal yöntemlere alternatif olarak, başarılı bir şekilde tespit edilebilmektedir [2][4][5][6].

Bu tezde, VCV yöntemi küme doğrulama endekslerine alternatif bir yöntem olarak kullanılmıştır. Şekil 5.11' deki matris grafiği, bulanık C-ortalamlar kümeleme yöntemiyle tespit edilen 2, 3, 4, 5, 6, 7, 8, 9 ve 10 küme için kümeleme sonuçlarını VCV yöntemine göre göstermektedir. Burada koyu mavi renkler konut satış fiyat endekslerinin birbirine benzerliğini gösterirken, kırmızı renkler birbirlerine benzemezliğini göstermektedir. Böylece, renk araçları kullanılarak kümeleme

performansı, diğer bir deyişle de satış fiyat endekslerinin kendi içinde homojen, kendi aralarında heterojen bir yapıda kümelenip kümelenmediği, görsel bir şekilde ortaya konmaktadır.



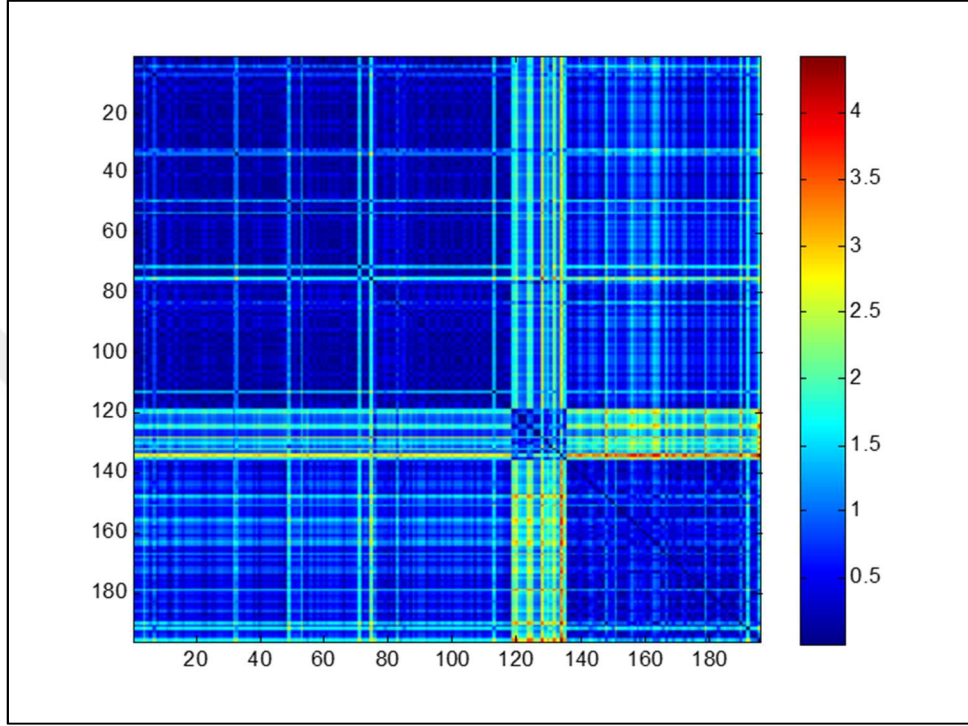
**Şekil 5.11** Bulanık C-ortalamalar görsel doğrulama sonuçları

Şekil 5.11'den de anlaşılacağı üzere kendi içinde homojenliği, kendi aralarında heterojenliği en yüksek renklemeye veren küme sayısı 3' tür. VCV ile tespit edilen küme sayısı da Davies-Bouldin, Calinski-Harabasz, Dunn ve Hartigan küme doğrulama endeks sonuçlarıyla da uyuşmaktadır.

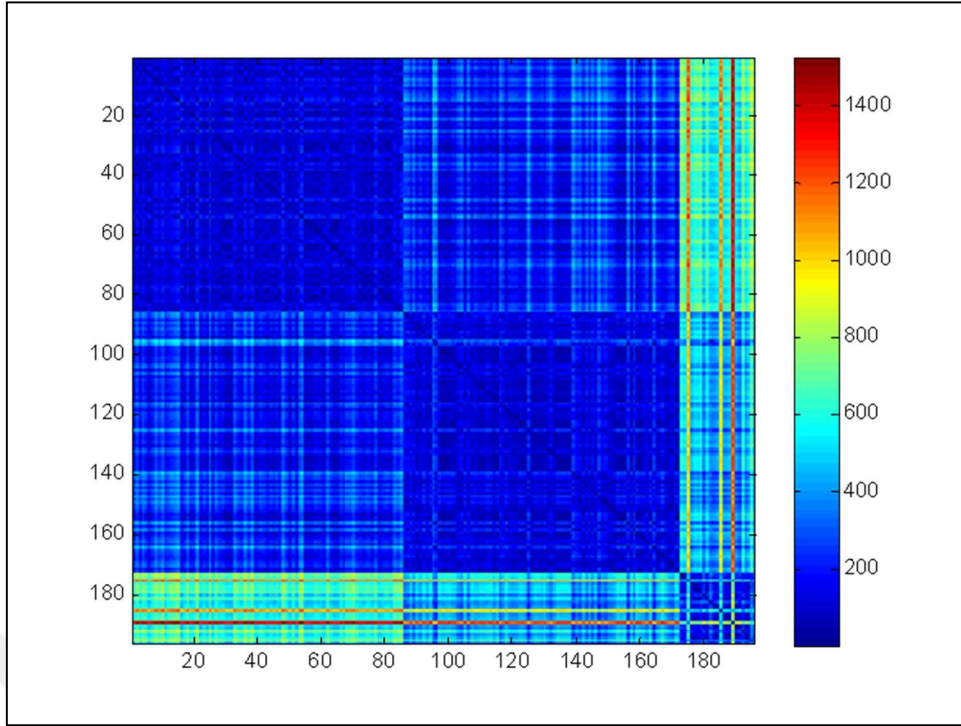
Bu çalışmada önerilen otoregresif model bazlı zaman serileri kümeleme yaklaşımı sonuçları, klasik ham veri bazlı zaman serileri kümeleme yaklaşımı ile de karşılaştırılmıştır. Şekil 5.12, model parametreleri üzerinden bulanık C-ortalamalar kümeleme algoritması ile elde edilen 3 kümeyi gösterirken, ham veri kümesi kullanılarak elde edilen 3 küme sonucu Şekil 5.13'de gösterilmektedir. Şekillerden de görüleceği üzere, görsel küme doğrulama yöntemine göre otoregresif model bazlı zaman serisi yaklaşımı ile elde edilen küme sonuçları kendi içerisinde daha homojen ve kendi aralarında daha heterojendir. Buradan hareketle, ham veri bazlı zaman serileri kümeleme yöntemine kıyasla, model bazlı



zaman serileri kümeleme yöntemiyle elde edilen küme yapılarının daha belirgin ve anlamlı olduğunu söyleyebiliriz. Buradaki önemli nokta: otoregresif model bazlı kümeleme çalışması ile, ham veri bazlı kümeleme çalışmalarındaki gibi miktar bazlı kümelemeden ziyade, eğilim bazlı bir gruptama yapılmaktadır. Dolayısıyla iki yaklaşım arasındaki en temel ve en önemli farklılık performanstan ziyade budur.

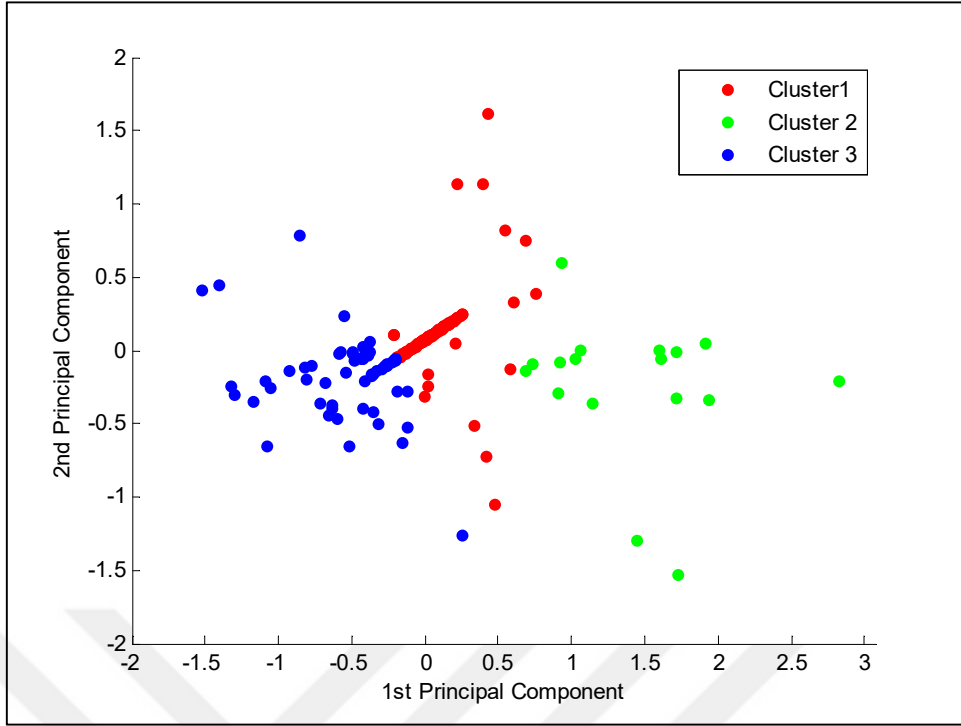


**Şekil 5.12** 3 Küme İçin Model Bazlı Bulanık C-Ortalamalar VCV Sonuçları

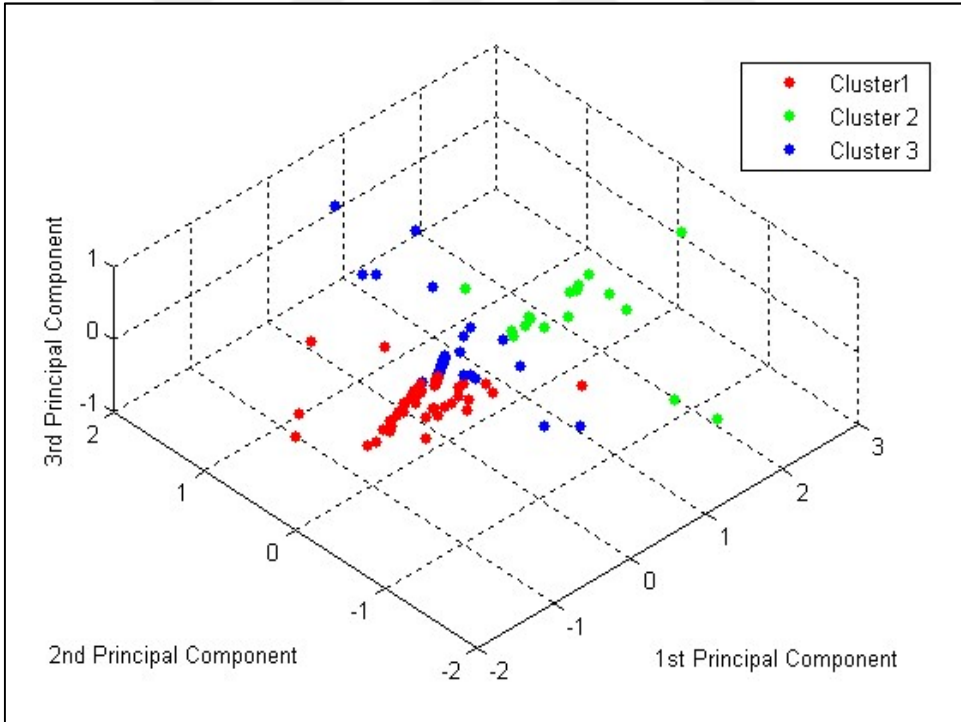


**Şekil 5.13** 3 Küme İçin Ham Bazlı Bulanık C-Ortalamalar VCV Sonuçları

Temel bileşenler analizi boyut indirgemesi yaparak çok değişkenli veri kümesini iki yada üç boyutta görselleştirmemize olanak tanımaktadır. Şekil 5.14, temel bileşenler analizine göre, varyans açıklayıcılığı en yüksek ilk iki bileşen kullanılarak, model parametre dağılımlarını iki boyutlu uzayda göstermektedir. Kümeler farklı renklere gösterilmektedir. İlgili şekilden de anlaşılacağı üzere aynı küme üyeleri iki boyutlu koordinat düzleminde birbirine yakın dağılmaktadır. Şekil 5.15, açıklayıcılığı en yüksek olan ilk üç temel bileşene göre, model parametre dağılımlarını üç boyutlu olarak göstermektedir. İki boyutlu düzlemde olduğu gibi, üç boyutlu uzayda da aynı küme üyeleri bir birlerine yakın dağılmaktadır. Bu iki görsel sonuç da, temel bileşenler analizinin, kümeleme sürecinde bulunan küme gruplarını ne ölçüde doğru seçtiğimizi desteklemektedir.



Şekil 5.14 Üç Küme İçin İki Boyutlu Temel Bileşenler Analizi



Şekil 5.15 Üç Küme İçin Üç Boyutlu Temel Bileşenler Analizi

Kümeleme sonucunda elde edilen bu üç küme için: birinci kümede 118, ikinci kümede 17 ve üçüncü kümede 61 adet mahalle bulunmaktadır. Tablo 5.3, 5.4, 5.5, 5.6 ve 5.7, il bazlı, benzer konut satış fiyat değişim dinamiklerine sahip olan mahallelerin bağlı bulunduğu kümeleri göstermektedir.

**Tablo 5.3 Üç Küme İçin Ankara İli Kümeleme Sonuçları**

İlçeler	Küme 1-Mahalleler	Küme 2-Mahalleler	Küme 3-Mahalleler
<b>Çankaya</b>	Ahlatlıbel, Anıttepe, Aşağı Dikmen, Aşağı Öveçler, Ata, Aydınlar, Ayrancı, Aziziye, Barbaros, Beytepe, Birlik, Büyükesat, Cebeci, Cevizlidere, Çamlıtepe, Çayyolu, Çiğdem, Çukurambar, Emek, Gaziosmanpaşa, Harbiye, Hilal, İlkadım, İlkbahar, İlker, Kavaklıdere, Keklik Pınarı, Kırkkonaklar, Mutlukent, Öveçler, Seyranbağları, Sokullu Mehmet Paşa, Yaşamkent, Yıldızevler	Akpınar, Kızılay, Oran, Prof. Dr. Ahmet Taner Kışlalı, Ümit	Balgat, Güvenevler, Huzur, İşçi Blokları, Karapınar, Küçükesat, Maltepe, Sancak, Yukarı Dikmen
<b>Keçiören</b>	Aktepe, Aşağı Eğlence, Atapark, Ayvalı, Bademlik, Bağlarbaşı, Basınevleri, Etlik, İncirli, Kalaba, Kanuni, Köşk, Kuşcağız, Şefkat, Ufuktepe, Uyanış, Yayla	Esertepe, Tepebaşı	23 Nisan, Adnan Menderes, Çaldıran, Çalseki, Çiçekli, Emrah, Güçlükaya, Gümüşdere, Kamil Ocak, Kavacık Subayevleri, Osmangazi, Pınarbaşı, Sancaktepe, Şehit Kubilay, Şenlik, Yakacık, Yükseltepe

**Tablo 5.4 Üç Küme İçin Bursa İli Kümeleme Sonuçları**

İlçeler	Küme 1-Mahalleler	Küme 2-Mahalleler	Küme 3-Mahalleler
<b>Nilüfer</b>	Altınşehir, Ataevler, Barış, Beşevler, Çamlıca, Ertuğrul, Fethiye, İhsaniye, Konak, Özlüce, Üçevler	Görükle	Balat, Cumhuriyet, Demirci, Kültür, Yüzüncüyıl
<b>Osmangazi</b>	Altıparmak, Çekirge, Demirtaş Cumhuriyet, Dikkaldırım, İbrahimpaşa	Kükürtlü, Yunuseli	Ahmetpaşa, Çırpan, Demirtaşpaşa, Doburca, Hocahasan, Hüdavendigar, Ovaakça Merkez

**Tablo 5.5 Üç Küme İçin İstanbul İli Kümeleme Sonuçları**

İlçeler	Küme 1-Mahalleler	Küme 2-Mahalleler	Küme 3-Mahalleler
<b>Beşiktaş</b>	Abbasağa, Akat, Arnavutköy, Bebek, Etiler, Gayrettepe, Kuruçeşme, Levazım, Levent, Muradiye, Ortaköy, Türkali	Dikilitaş, Vişnezade	Balmumcu, Cihannüma, Konaklar, Mecidiye, Nispetiye, Sinanpaşa, Ulus, Yıldız
<b>Kadıköy</b>	Acıbadem, Bostancı, Caferağa, Dumlupınar, Eğitim, Erenköy, Fenerbahçe, Fikirtepe, Göztepe, Hasanpaşa, Koşuyolu, Kozyatağı, Merdivenköy, Rasimpaşa, Sahrayı Cedit, Suadiye	Caddebostan, Feneryolu	19 Mayıs, Osmanağa, Zühtüpaşa

**Tablo 5.6 Üç Küme İçin İzmir İli Kümeleme Sonuçları**

İlçeler	Küme 1-Mahalleler	Küme 2-Mahalleler	Küme 3-Mahalleler
<b>Karşıyaka</b>	Aksoy, Alaybey, Bahçelievler, Bahriye Üçok, Bostanlı, Dedebaşı, Demirköprü, Mavişehir, Nergiz, Örnekköy, Şemikler, Yalı	Donanmacı	Atakent, Bahariye, Fikri Altay, Goncalar
<b>Konak</b>	Akın Simav, Altıntaş, Çankaya, Güzelyalı	Mithatpaşa	Alsancak, Atilla, Kılıç Reis, Namık Kemal, Zafertepe

**Tablo 5.7 Üç Küme İçin Kocaeli İli Kümeleme Sonuçları**

İlçeler	Küme 1-Mahalleler	Küme 2-Mahalleler	Küme 3-Mahalleler
<b>Gölcük</b>	Değirmendere Merkez	Merkez	-
<b>İzmit</b>	28 Haziran, Cedit, Yahyakaptan, Yenişehir	-	Akarca, Alikahya Merkez, Kadıköy

Tablo 5.8 il ve ilçe bazlı küme adetleri bilgisini vermektedir. Söz konusu, tablo ek olarak, Ocak 2010 - Aralık 2016 dönemlerinden oluşan eğitim kümesi için, kümelere göre değişkenliği ifade eden standart sapma değerlerini de vermektedir. Tablodan da görüleceği üzere üçüncü küme 29.54 en yüksek ortalama standart sapmayla zaman düzleminde konut fiyat değerlerinin en çok oynadığı kümeyi ifade etmektedir. Diğer taraftan, birinci küme 21.88 en düşük ortalama standart sapma değeri ile zaman düzleminde en istikrarlı kümeyi; diğer bir deyişle konut fiyat değerlerinin en az oynadığı kümeyi ifade etmektedir.

**Tablo 5.8** Kümelere Göre Standart Sapma Değerleri

İller	İlçeler	Küme 1	Küme 2	Küme 3	Toplam Adet
Ankara	Çankaya	34	5	9	48
	Keçiören	17	2	17	36
Bursa	Nilüfer	11	1	5	17
	Osmangazi	7	2	7	16
İstanbul	Beşiktaş	12	2	8	22
	Kadıköy	16	2	3	21
İzmir	Karşıyaka	12	1	4	17
	Konak	4	1	5	10
Kocaeli	Gölcük	1	1	0	2
	İzmit	4	0	3	7
Birim Sayısı		118	17	61	196
Ortalama Standart Sapma		29.54	25.76	21.88	26.83

## 5.5 Otoregresif ve Bulanık C-Ortalamlar Model Öngörü Performanslarının Karşılaştırılması

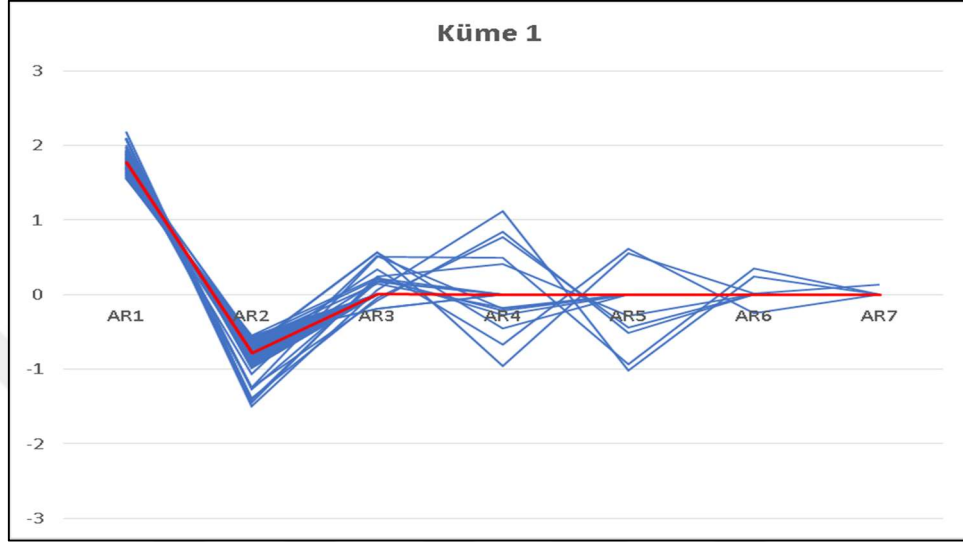
Bulanık C-ortalamlar kümeleme algoritması, kümeleme işlemi sonucunda kümeleri temsil eden küme merkezlerini, parametre tahmin değeri olarak, vermektedir. Her bir satış endeksi bir üyelik değerine göre bu parametreler dikkate alınarak kümelere atanmaktadır. Bu parametre tahmin değerlerini otoregresif model parametresi şeklinde düşünerek; parametreler temsil ettikleri kümeler için öngörü yapılabilmektedir.

Tablo 5.9, bulanık C-ortalamlar kümeleme yönteminden elde edilen parametre tahmin değerlerini göstermektedir.

**Tablo 5.9** Bulanık C-Ortalamlar Parametre Tahmin Değerleri

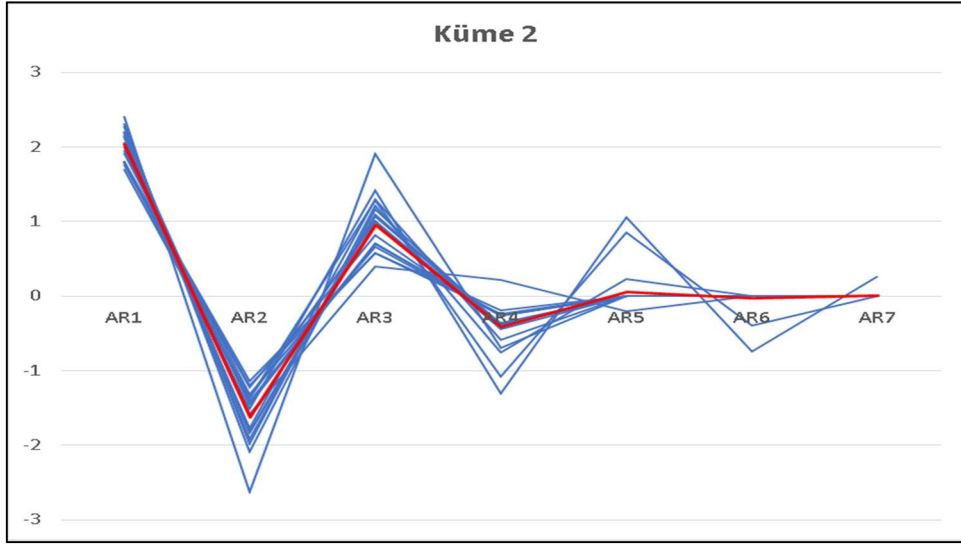
Kümelere	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$
<b>Küme 1</b>	1.7749	-0.7791	0.0112	-0.0012	-0.0054	-0.0004	0.0004
<b>Küme 2</b>	2.0398	-1.6221	0.9586	-0.4068	0.0543	-0.0303	0.0087
<b>Küme 3</b>	1.4581	-0.2800	-0.1290	-0.0310	-0.0087	-0.0095	0.0006

Şekil 5.16, 118 gözlemden oluşan 1. grup küme elemanlarının 7 tane AR parametre değerini çizgi grafiklerle göstermektedir. Şekildeki kırmızı çizgi Tablo 5.9’da yer alan 1. grup küme için bulanık C-ortalamlar parametre tahmin değerlerini göstermektedir. Şekilden görüldüğü üzere bulanık C-ortalamlar parametre tahmin değerleri kümeyi merkez olacak şekilde temsil etmektedir.



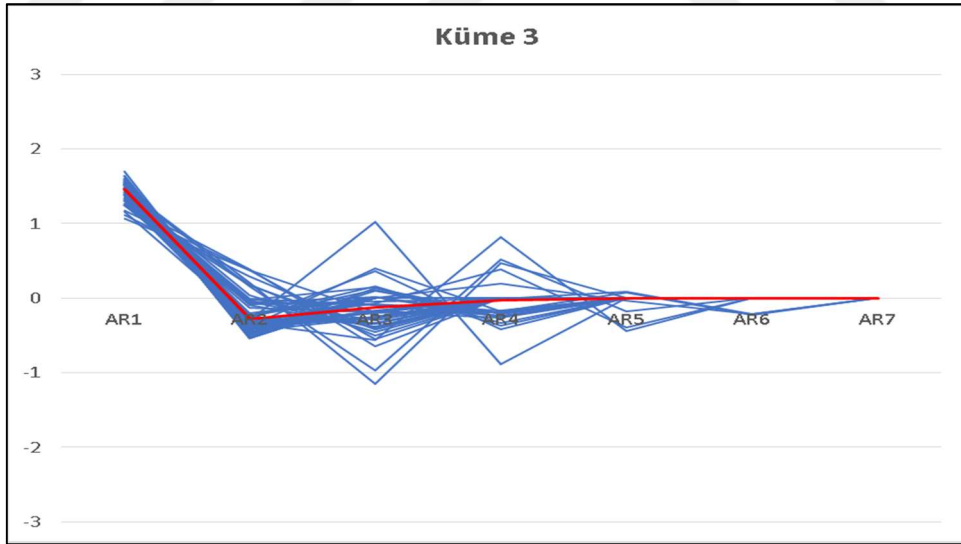
Şekil 5.16 Küme 1 İçin Küme Profilleri

Şekil 5.17, 17 gözlemden oluşan 2. grup küme elemanlarının 7 tane AR parametre değerini çizgi grafiklerle göstermektedir. Şekildeki kırmızı çizgi Tablo 5.9’ da yer alan 2. grup küme için bulanık C-ortalamlar parametre tahmin değerlerini göstermektedir. Şekilden görüldüğü üzere bulanık C-ortalamlar parametre tahmin değerleri kümeyi merkez olacak şekilde temsil etmektedir.



**Şekil 5.17** Küme 2 İçin Küme Profilleri

Şekil 5.18, 61 gözlemden oluşan 3. grup küme elemanlarının 7 tane AR parametre değerini çizgi grafiklerle göstermektedir. Şekildeki kırmızı çizgi Tablo 5.9’da yer alan 3. grup küme için bulanık C-ortalamalar parametre tahmin değerlerini göstermektedir. Şekilden görüldüğü üzere bulanık C-ortalamalar parametre tahmin değerleri kümeyi merkez olacak şekilde temsil etmektedir.



**Şekil 5.18** Küme 3 İçin Küme Profilleri

Modellerin parametre tahmini ve öngörüdeki performanslarını karşılaştırmak amacıyla her mahalle gayrimenkul satış fiyat endeks veri seti, %75’i eğitim %25’i test seti olacak şekilde iki döneme ayrılmıştır. Buna göre eğitim dönemi 2010 Ocak



- 2016 Aralık, test dönemi 2017 Ocak - 2017 Aralık dönemlerinden oluşmaktadır. Eğitim seti bireysel otoregresif model parametre ve otoregresif bazlı bulanık C-ortalamalar küme merkezlerini tahmin etmek, test ise elde edilen modellerin öngörüdeki başarısını değerlendirmek amacıyla kullanılmıştır.

Otoregresif bazlı bulanık C-ortalamalar kümeleme yöntemi sonucu elde edilen küme merkezleri, içinde buldukları kümeleri temsil eden otoregresif bazlı bir model gibi düşünülebilir. Bu durumda her üç küme için öngörü fonksiyonu aşağıdaki denklemlerle ifade edilir.

Küme 1'de yer alan i. zaman serisi için:

$$Y_{i,t} = 1.7749Y_{i,t-1} - 0.7791Y_{i,t-2} + 0.0112Y_{i,t-3} - 0.0012Y_{i,t-4} - 0.0054Y_{i,t-5} - 0.0004Y_{i,t-6} + 0.0004Y_{i,t-7} + \epsilon_t \quad (5.6)$$

Küme 2'de yer alan i. zaman serisi için:

$$Y_{i,t} = 2.0398Y_{i,t-1} - 1.6221Y_{i,t-2} + 0.9586Y_{i,t-3} - 0.4068Y_{i,t-4} + 0.0543Y_{i,t-5} - 0.0303Y_{i,t-6} + 0.0087Y_{i,t-7} + \epsilon_t \quad (5.7)$$

Küme 3 de yer alan i. zaman serisi için:

$$Y_{i,t} = 1.4581Y_{i,t-1} - 0.2800Y_{i,t-2} - 0.1290Y_{i,t-3} - 0.0310Y_{i,t-4} - 0.0087Y_{i,t-5} - 0.0095Y_{i,t-6} + 0.0006Y_{i,t-7} + \epsilon_t \quad (5.8)$$

Otoregresif bazlı bulanık C-ortalamalar kümeleme yöntemine göre, 2. küme içerisinde yer alan Kocaeli ilinin Gölcük İlçesinin Merkez mahallesi Ocak-2017 ve Aralık-2017 dönemi için öngörü

$$Y_{i,t+h} = 2.0398Y_{i,t+h-1} - 1.6221Y_{i,t+h-2} + 0.9586Y_{i,t+h-3} - 0.4068Y_{i,t+h-4} + 0.0543Y_{i,t+h-5} - 0.0303Y_{i,t+h-6} + 0.0087Y_{i,t+h-7} + \epsilon_t \quad (5.9)$$

ile hesaplanır.

$h = 1$ , Ocak 2017 için, tahmini  $\hat{Y}_{i,t+1}$ , otoregresif bazlı bulanık C-ortalamalar küme merkezine göre, öngörü değeri

$$\hat{Y}_{i,t+1} = 2.0398 * 122.81 - 1.6221 * 120.05 + 0.9586 * 116.15 \quad (5.10)$$

$$-0.4068 * 114.97 + 0.0543 * 117.65$$

$$-0.0303 * 117.9 + 0.0087 * 118.98$$

ile ve dolayısıyla

$$\hat{Y}_{i,t+1} = 124.20$$

şeklinde bulunur.

Aynı şekilde,  $h = 2$ , Ocak 2017 için, tahmini  $\hat{Y}_{i,t+2}$ , otoregresif modele göre, öngörü değeri aşağıdaki gibi hesaplanır.

$$\hat{Y}_{i,t+1} = 2.0398 * 124.20 - 1.6221 * 122.81 + 0.9586 * 120.05 \quad (5.11)$$

$$-0.4068 * 116.15 + 0.0543 * 114.97$$

$$-0.0303 * 117.65 + 0.0087 * 117.9$$

$$\hat{Y}_{i,t+1} = 125.59$$

Devam edecek olursak,  $t+3 =$  Mart 2017,  $t+4 =$  Nisan 2017,  $t+5 =$  Mayıs 2017,  $t + 6 =$  Haziran 2017,  $t + 7 =$  Temmuz 2017,  $t + 8 =$  Ağustos 2017,  $t + 9 =$  Eylül 2017,  $t + 10 =$  Ekim 2017,  $t + 11 =$  Kasım 2017 ve  $t + 12 =$  Aralık 2017 dönemleri için benzer şekilde tahmini  $\hat{Y}_{i,t+3}, \hat{Y}_{i,t+4}, \hat{Y}_{i,t+5}, \hat{Y}_{i,t+6}, \hat{Y}_{i,t+7}, \hat{Y}_{i,t+8}, \hat{Y}_{i,t+9}, \hat{Y}_{i,t+10}, \hat{Y}_{i,t+11}$  ve  $\hat{Y}_{i,t+12}$  öngörü değerleri hesaplanır.

Bireysel otoregresif modellere göre öngörü fonksiyonu, her bir  $i$ . konut satış fiyat endeksi için (5.11) denklemdeki gibi ifade edilir. Buna göre Kocaeli ilinin Gölcük ilçesinin Merkez Mahallesi için, sıradan enküçük kareler yöntemine göre tahminlenen model parametreleri  $\hat{\beta}_1 = 0.8204, \hat{\beta}_2 = -0.6875, \hat{\beta}_3 = 0.7257, \hat{\beta}_4 = -0.5781, \hat{\beta}_5 = 0.4765$  ve  $\hat{\beta}_6 = -0.2641$  olmak üzere bireysel otoregresif öngörü fonksiyonu

$$Y_{i,t+h} = (1 + \hat{\beta}_1)Y_{i,t+h-1} + (\hat{\beta}_2 - \hat{\beta}_1)Y_{i,t+h-2} + (\hat{\beta}_3 - \hat{\beta}_2)Y_{i,t+h-3} \quad (5.12)$$

$$+(\hat{\beta}_4 - \hat{\beta}_3)Y_{i,t+h-4} + (\hat{\beta}_5 - \hat{\beta}_4)Y_{i,t+h-5} + (\hat{\beta}_6 - \hat{\beta}_5)Y_{i,t+h-6} \\ - \hat{\beta}_6 Y_{i,t+h-7} + \epsilon_t$$

ile ifade edilir.

h = 0 için, denklem

$$Y_{i,t} = 1.8204Y_{i,t-1} - 1.5079Y_{i,t-2} + 1.4131Y_{i,t-3} - 1.3038Y_{i,t-4} \quad (5.13) \\ + 1.0546Y_{i,t-5} - 0.7406Y_{i,t-6} + 0.2641Y_{i,t-7} + \epsilon_t$$

şeklinde ifade edilir.

h = 1, Ocak 2017 için, tahmini  $\hat{Y}_{i,t+1}$ , otoregresif modele göre, öngörü değeri

$$\hat{Y}_{i,t+1} = 1.820 * 122.81 - 1.5079 * 120.05 + 1.4131 * 116.15 \quad (5.14) \\ - 1.3038 * 114.97 + 1.0546 * 117.65 - 0.7406 * 117.9 + 0.2641 * 118.98$$

ile ve netice itibariyle

$$\hat{Y}_{i,t+1} = 126.29$$

olarak bulunur.

Aynı şekilde, h = 2, Ocak 2017 için, tahmini  $\hat{Y}_{i,t+2}$ , otoregresif modele göre, öngörü değeri

$$\hat{Y}_{i,t+2} = 1.820 * 126.29 - 1.5079 * 122.81 + 1.4131 * 120.05 \quad (5.15) \\ - 1.3038 * 116.15 + 1.0546 * 114.97 - 0.7406 * 117.65 + 0.2641 * 117.9$$

olarak ve sonuç olarak

$$\hat{Y}_{i,t+2} = 128.64$$

bulunur.

Devam edecek olursak, t+3 = Mart 2017, t+4 = Nisan 2017, t+5 = Mayıs 2017, t + 6 = Haziran 2017, t + 7 = Temmuz 2017, t + 8 = Ağustos 2017, t + 9 = Eylül 2017, t + 10 = Ekim 2017, t + 11 = Kasım 2017 ve t + 12 = Aralık 2017 dönemleri

için benzer şekilde tahmini  $\hat{Y}_{i,t+3}, \hat{Y}_{i,t+4}, \hat{Y}_{i,t+5}, \hat{Y}_{i,t+6}, \hat{Y}_{i,t+7}, \hat{Y}_{i,t+8}, \hat{Y}_{i,t+9}, \hat{Y}_{i,t+10}, \hat{Y}_{i,t+11}$  ve  $\hat{Y}_{i,t+12}$  öngörü değerleri hesaplanır.

Şekil 5.19, 2017 Ocak -2017 Aralık dönemleri için Kocaeli, Gölcük, Merkez Mahallesi: Gerçek, OM ve OMBBK tahmini öngörü değerlerini birlikte göstermektedir. Şekilden de görüleceği üzere OMBBK tahmini öngörü değerleri OM'ye göre daha gerçekçidir.

Öngörü performanslarını değerlendirmek için bu tezde ortalama mutlak hata (MAE), ortalama mutlak yüzde hata (MAPE) ve ortalama karesel hata (RMSE) metrikleri kullanılır.

$e_{t+h}$  öngörü hata terimi, gerçekleşen  $y_{t+h}$  ve tahmini  $\hat{y}_{t+h}$  öngörü değerleri arasındaki fark olmak üzere

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h/t}, t = 1, 2, \dots, n \text{ ve } h = 1, 2, \dots, m \quad (5.16)$$

biçiminde hesaplanır.

Burada  $\{y_1, \dots, y_t\}$  eğitim periyodunu ve  $\{y_{t+1}, \dots, y_{t+h}\}$  test dönemini ifade etmektedir.

$e_{t+h}$  öngörü hata terimine göre MAE, MAPE ve RMSE, sırasıyla,

$$MAE_h = \sqrt{\frac{1}{m} \sum_{h=1}^m e_{t+h}}, \quad (5.17)$$

$$MAPE_h = \sqrt{\frac{1}{m} \sum_{h=1}^m \frac{(y_{t+h} - \hat{y}_{t+h})}{y_{t+h}}}, \quad (5.18)$$

$$RMSE_h = \sqrt{\frac{1}{m} \sum_{h=1}^m e_{t+h}^2} \quad (5.19)$$

formülleri ile hesaplanır.

Tablo 5.10, Merkez mahallesi için otoregresif model (OM) ve otoregresif model bazlı bulanık kümeleme (OMBBK) yöntemine göre elde edilen öngörü tahmin değerlerini Ocak-Aralık 2017 dönemleri için hata terimleri ile beraber vermektedir. Tablo 5.11' de öngörü tahminlerine ilişkin MAE, MAPE ve RMSE

sonuçlarını vermektedir. Tablolara göre OMBBK, OM ye göre daha düşük hatalı tahmini öngörülerde bulunmaktadır.

Merkez mahallesi için, Tablo 5.10' daki OM ve OMBBK öngörü hataları kullanılarak MAE, MAPE ve RMSE aşağıdaki gibi hesaplanır.

Otoregresif model için:

$$MAE_h =$$

$$= \sqrt{\frac{1}{12}(1.32 + 2.87 + 0.55 + 0.44 + 2.15 + 3.94 + 3.84 + 1.49 + 4.26 + 1.78 + 4.58 + 6.44)}$$

$$= 2.80$$

$$MAPE_h =$$

$$= \frac{1}{12} \left[ \left( \frac{126.29 - 124.97}{126.29} \right) + \left( \frac{128.64 - 125.77}{128.64} \right) + \left( \frac{126.52 - 125.97}{126.52} \right) + \left( \frac{127.54 - 127.10}{127.54} \right) + \right.$$

$$\left. \left( \frac{129.65 - 127.50}{129.65} \right) + \left( \frac{130.98 - 127.04}{130.98} \right) + \left( \frac{130.74 - 126.90}{130.74} \right) + \left( \frac{128.12 - 126.63}{128.12} \right) + \right.$$

$$\left. \left( \frac{130.68 - 126.42}{130.68} \right) + \left( \frac{128.27 - 126.49}{128.27} \right) + \left( \frac{130.83 - 126.25}{130.83} \right) + \left( \frac{132.50 - 126.06}{132.50} \right) \right]$$

$$= 2.22$$

$$RMSE_h =$$

$$= \sqrt{\frac{1}{12}(1.32^2 + 2.87^2 + 0.55^2 + 0.44^2 + 2.15^2 + 3.94^2 + 3.84^2 + 1.49^2 + 4.26^2 + 1.78^2 + 4.58^2 + 6.44^2)}$$

$$= 3.31$$

Bulanık C-oralamalar için:

$$MAE_h =$$

$$= \sqrt{\frac{1}{12}(2.09 + 3.05 - 0.23 - 0.21 + 1.05 + 1.65 + 0.78 - 2.40 - 0.34 - 3.21 - 1.06 + 0.23)}$$

$$= 1.36$$

$$MAPE_h =$$

$$\sqrt{\frac{1}{12} \left[ \left( \frac{126.29 - 124.20}{126.29} \right) + \left( \frac{128.64 - 125.59}{128.64} \right) + \left( \frac{126.52 - 126.75}{126.52} \right) + \left( \frac{127.54 - 127.75}{127.54} \right) + \left( \frac{129.65 - 128.60}{129.65} \right) + \left( \frac{130.98 - 129.33}{130.98} \right) + \left( \frac{130.74 - 129.96}{130.74} \right) + \left( \frac{128.12 - 130.52}{128.12} \right) + \left( \frac{130.68 - 131.02}{130.68} \right) + \left( \frac{128.27 - 131.48}{128.27} \right) + \left( \frac{130.83 - 131.89}{130.83} \right) + \left( \frac{132.50 - 132.27}{132.50} \right) \right]}$$

$$= 1.06$$

$$RMSE_h =$$

$$= \sqrt{\frac{1}{12} (2.09^2 + 3.05^2 + 0.23^2 + 0.21^2 + 1.05^2 + 1.65^2 + 0.78^2 + 2.40^2 + 0.34^2 + 3.21^2 + 1.06^2 + 0.23^2)}$$

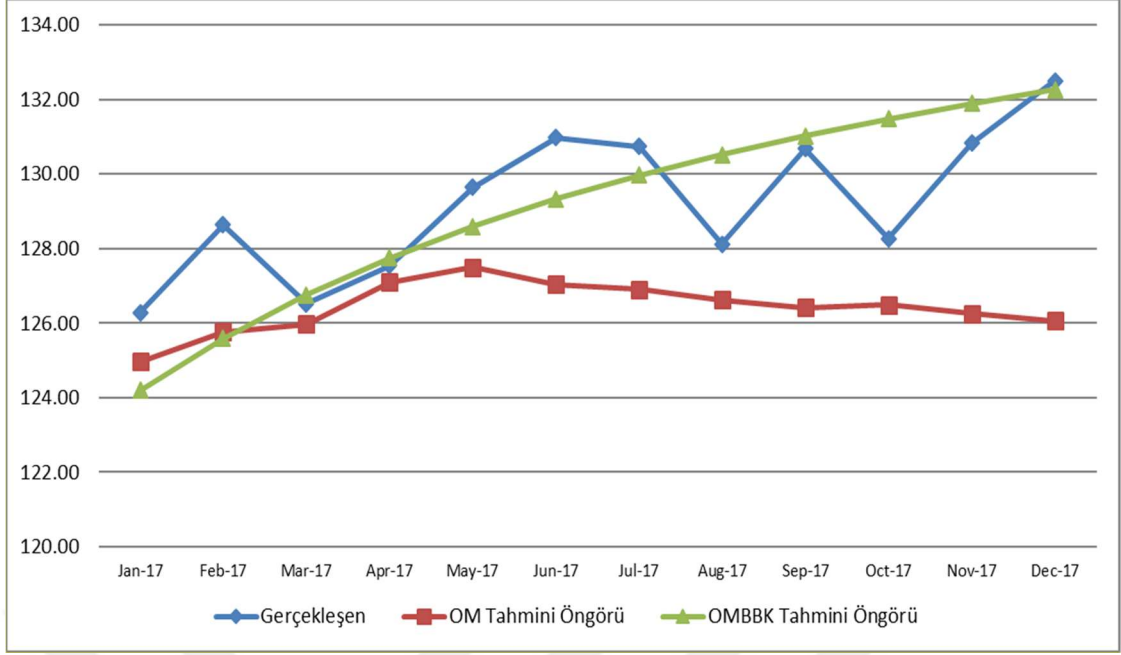
$$= 1.72$$

**Tablo 5.10** Kocaeli, Gölcük, Merkez Mahallesi: OM ve OMBBK Tahmini Öngörü Değerleri

Tarih (t)	Endeks	OM Tahmini Öngörü	OM Öngörü Hata	OMBBK Tahmini Öngörü	OMBBK Öngörü Hata
Oca-2016	122.15	-	-	-	-
Şub-2016	125.12	-	-	-	-
Mar-2016	123.42	-	-	-	-
Nis-2016	123.63	-	-	-	-
May-2016	122.37	-	-	-	-
Haz-2016	118.98	-	-	-	-
Tem-2016	117.9	-	-	-	-
Ağu-2016	117.65	-	-	-	-
Eyl-2016	114.97	-	-	-	-
Eki-2016	116.15	-	-	-	-
Kas-2016	120.05	-	-	-	-
Ara-2016	122.81	-	-	-	-
Oca-2017	126.29	124.97	1.32	124.20	2.09
Şub-2017	128.64	125.77	2.87	125.59	3.05
Mar-2017	126.52	125.97	0.55	126.75	-0.23
Nis-2017	127.54	127.10	0.44	127.75	-0.21
May-2017	129.65	127.50	2.15	128.60	1.05
Haz-2017	130.98	127.04	3.94	129.33	1.65
Tem-2017	130.74	126.90	3.84	129.96	0.78
Ağu-2017	128.12	126.63	1.49	130.52	-2.40
Eyl-2017	130.68	126.42	4.26	131.02	-0.34
Eki-2017	128.27	126.49	1.78	131.48	-3.21
Kas-2017	130.83	126.25	4.58	131.89	-1.06
Ara-2017	132.5	126.06	6.44	132.27	0.23

**Tablo 5.11** Kocaeli, Gölcük, Merkez Mahallesi: OM ve OMBBK Tahmini Öngörü Performans Değerlendirme

	OM Tahmini Öngörü Değerlendirme	OMBBK Tahmini Öngörü Değerlendirme
<b>MAE</b>	2.80	1.36
<b>MAPE</b>	2.22	1.06
<b>RMSE</b>	3.31	1.72



**Şekil 5.19** Kocaeli, Gölcük, Merkez Mahallesi: Gerçek, OM ve OMBBK Tahmini Öngörü Değerleri

Her bir 196 mahalle bazlı konut satış fiyat endeksi, Ocak 2017-Aralık 2017 dönemleri için, bireysel tahminlenen otoregresif modeller ve endeksin ait olduğu kümenin bulanık küme merkezleri üzerinden öngörüler yapılmıştır. Öngörü başarı performansları ortalama mutlak hata (MAE), ortalama mutlak yüzde hata (MAPE) ve ortalama karesel hata (RMSE) metrikleri kullanılarak değerlendirilmiştir. Yapılan analizlerde, hem tahminlemede hem de öngöründe otoregresif bazlı bulanık C-ortalamlar kümeleme yönteminin daha doğru performans gösterdiği görülmektedir. Diğer bir ifadeyle, otoregresif bazlı bulanık C-ortalamlar kümeleme yöntemiyle belirlenen 3 kümeyle, 0.98, oranında öngörü model sayısı azaltılarak, bireysel otoregresif modellere göre daha başarılı öngörü sonuçları elde edilmektedir. Rakamlarla, 196 mahalle bazlı gayrimenkul satış fiyatları endeksinin 136 tanesinin OMBBK öngörü tahmin değerleri gerçek değerlerini gösteren test kümesine daha yakın çıkmıştır. Bu da adet olarak 70% oranında öngörü performans iyileşmesi demektir. Ek E, 196 endeksin Ocak 2017-Aralık 2017 dönemleri için OM ve OMBBK' ye göre yapılan tahmini öngörü değerlerini niteliksel olarak göstermektedir. Ek F, OM ve OMBBK öngörü tahminleri için MAE, MAPE ve RMSE başarı performanslarını vermektedir.



Bu çalışmada, Türkiye' nin 5 büyük ilinin toplamda 196 mahallesi için benzer konut satış fiyat davranışlarını tespit etmek amacıyla otoregresif model bazlı bir zaman serisi kümeleme yaklaşımı geliştirildi. Bu model bazlı zaman serisi kümeleme yaklaşımı ile, her bir zaman serisinin bireysel tahmini otoregresif modellerini, benzerliklere göre sınıflandırılması da gerçekleştirildi.

Benzer gayirmenkul satış fiyat endeks otoregresif modelleri, model parametrelerinden hesaplanan Öklid uzaklığına bulanık C-ortalamlar kümeleme yöntemi uygulanarak tespit edildi. Sayısal ve görsel (sayısal olmayan) küme doğrulama yaklaşımlarına göre, fiyat eğilimleri açısından, üç küme tespit edildi. Küme 1, 118 adet mahalle ve 29.54 standart sapma ile en yüksek hareketliliğe sahip grup olduğu gözlemlendi. Küme 2' nin, 17 mahalle ve 25.76 standart sapma ile nispeten fiyat açısından daha kararlı bir grup olduğu sonucuna varıldı. Geri kalan 61 mahallenin 21.88 standart sapma ile en kararlı fiyat hareketlerinden oluştuğu ortaya çıktı. Ayrıca bu sonuçlar ile coğrafi mesafeden bağımsız olarak aynı şehirde, hatta aynı ilçede bile farklı konut satış fiyat davranışları gösterebilen mahallelerin olduğu öngörüldü.

Önerilen otoregresif model bazlı zaman serileri kümeleme yaklaşımı ile, zaman serileri için bir kümeleme işleminin yapıldığı gibi; aynı anda da bireysel zaman serileri için öngörülerde yapıldı. Model öngörü performanslarının incelenebilirliği için veri kümesi farklı zaman periyotlarını içeren iki veri kümesine ayrıldı: %75 eğitim (2010 Ocak – 2016 Aralık), %25 test (2017 Ocak – 2017 Aralık). Eğitim seti bireysel otoregresif ve otoregresif bazlı bulanık C-ortalamlar modellerini tahmin etmek, test ise elde edilen modellerin öngörüdeki başarısını değerlendirmek amacıyla kullanıldı.

Her bir 196 mahalle bazlı konut satış fiyat endeksi, Ocak 2017-Aralık 2017 dönemi için otoregresif ve otoregresif bazlı bulanık C-ortalamlar kümeleme yöntemleriyle öngörüldü. Veri seti için yapılan analizlerde, otoregresif bazlı

bulanık C-ortalamalar kümeleme yönteminin bireysel otoregresif modellerine göre daha başarılı sonuçlar verdiği görüldü. Diğer bir ifadeyle, otoregresif bazlı bulanık C-ortalamalar kümeleme yöntemiyle belirlenen 3 modelle, 0.98 oranında öngörü model sayısı azaltılarak, bireysel otoregresif modellere göre daha başarılı sonuçlar elde edildiği gözlemlendi. Rakamsal olarak 196 endeksin 140' ı için model bazlı bulanık C-ortalamalar yöntemi daha başarılı sonuçlar verdiği tespit edilmiştir.

Bu sonuçlar doğrultusunda, otoregresif model bazlı zaman serileri kümelemesinin, genel olarak, otoregresif modeline uyan çok sayıda zaman serilerinin kümelenebilmesi ve aynı zamanda öngörülebilmesi için kullanılabileceği sonucuna varıldı.

Gelecek dönemlerde yapılacak çalışmalar için, gayrimenkul satış fiyat davranışının tespitine yönelik çeşitli politik ve sosyo ekonomik değişkenler tespit edilebilir. Anlamli olduğu düşünölen bu değişkenlerle bireysel otoregresif modellerden ziyade çok değişkenli zaman serileri modelleri belirlenebilir. Bu çok değişkenli modellerle daha başarılı kümeler tespit edilebilir ve bu meyanda daha başarılı öngörüler gerçekleştirilebilir. Bu çalışmanın gelecekte yapılacak bu tür çalışmalara ışık tutacağı düşünölmektedir.

- [1] M. Vatansever, "Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizinde Kullanılması ve Uygulanması", Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, İstanbul, 2008.
- [2] R. Xu ve D. Wunsch II, "Survey of Clustering Algorithms", IEEE Transactions on Neural Networks, no 16, cilt 3, ss. 645-678, 2005.
- [3] R. Xu ve D. Wunsch II, Clustering, John Wiley and Sons, Inc., Hoboken, New Jersey, 2009.
- [4] T. W. Liao, "Clustering of Time Series Data – a Survey", Pattern Recognition, cilt 38, ss. 1857-1874, 2005.
- [5] S. Aghabozorgi, A. S. Shirkhorshidi ve T. Y. Wah, "Time-series clustering – A decade review", Information System, cilt 53, ss. 16-38, 2015.
- [6] J. Han, M. Kamber ve J. Pei, Data Minin Concepts and Techniques, 1. Baskı, Morgan Kaufmann Publishers, Waltham, 2011
- [7] Y. Kakizawa, R. H. Shumway ve M. Taniguchi, "Discrimination and Clustering for Multivariate Time Series", Journal of American Statistical Association, cilt 93, sayı 441, ss. 328-340, 1998.
- [8] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis ve P. Boesiger, "A New Correlation-Based Fuzzy Logic Clustering Algorithm for FMRI", Magnetic Resonance in Medicine, cilt 40, sayı 2, ss. 249-260, 1998.
- [9] J. J. Van Wijk ve E. R. Van Selow, "Cluster and Calendar Based Visualization of Time Series Data", Proceedings of IEEE Symposium on Information Visualization, cilt 10, sayı 199, ss. 25-26.
- [10] T. W. Liao, B. Bolt, J. Forester, E. Hailman, C. Hansen, R. C. Kaste ve J. O' May, "Understanding and Projecting the Battle State", 23rd Army Science Conference, Orlando, USA, 2002.
- [11] C. S. Möller-Levet, F. Klawonn, K.- H. Cho ve O. Wolkenhauer, "Fuzzy Clustering of Short Time Series and Unevenly Distributed Sampling Points", Proceedings of the 5th International Symposium on Intelligent Data Analysis, ss. 330-340, 2003.
- [12] C. A. Ratanamahatana ve V. Niennattrakul, "Clustering Multimedia Data Using Timeseries" Proceedings of the 1th International Conference on Hybrid Information Technology, Cheju Island, South Korea, 2006.
- [13] A. Hepşen ve M. Vatansever, "Using Hierarchical Algorithms for Turkish Residential Market", International Journal of Economics and Finance, cilt 4, sayı 1, ss. 138-150, 2012.
- [14] C. T. Shaw ve G. P. King, "Using Cluster Analysis to Classify Time Series", Physica D: Nonlinear Phenomena cilt 58, sayı 1-4, ss. 288-298, 1992

- [15] L. M. D. Owsley, L. E. Atlas ve G. D. Bernard, "Self-Organizing Feature Maps and Hidden Markov Models for Machine Tool Monitoring", *IEEE Transactions Signal Processing*, cilt 45, sayı 11, ss. 2787-2798, 1997.
- [16] M. Vlachos, J. Lin, E. Keogh ve D. Gunopulos, "A Wavelet Based Anytime Algorithm for k-means Clustering of Time Series", *Proceedings of the Third SIAM International Conference on Data Mining*, San Francisco, USA, 2003.
- [17] T. Rasanen ve M. Kolehmainen, "Feature-Based Clustering for Electricity Use Time Series Data", *Adaptive and Natural Computing Algorithms: 9th International Conference ICANNGA*, Kuopio, Finland, 2009.
- [18] B. D. Fulcher ve N. S. Jones, "Highly Comparative Feature-Based Time-Series Classification", *IEEE Transaction in Knowledge and Data Engineering*, cilt 26, sayı 12, ss. 3026-3037, 2014.
- [19] C. Biernacki, G. Celeux ve G. Govaert, "Assessing a Mixture Model for Clustering with the Integrated Completed Likelihood", *IEEE Transaction on Pattern Analysis and Machine Intelligence*, sayı 22, cilt 7, ss. 719-725, 2000.
- [20] M. Ramoni, P. Sebastiani ve P. Cohen, "Multivariate Clustering by Dynamics", *Proceedings of the Seventeenth National Conference on Artificial Intelligence*, Austin, Texas, USA, 2000.
- [21] D. Tran ve M. Wagner, "Fuzzy C-Means Clustering-Based Speaker Verification", *Advances in Soft Computing - AFSS 2002*, 2002 AFSS International Conference on Fuzzy Systems, Calcutta, India, USA, 2002, ss. 318-324.
- [22] Y. Xiong ve D. Y. Yeung, "Mixture of ARMA Models for Model-Based Time Series Clustering", *Proceedings of IEEE International Conference on Data Mining*, Maebaghi City, Japan, 2002.
- [23] J. Lin, M. Vlachos, E. Keogh ve D. Gunopulos, "Iterative Incremental Clustering of Time Series" *9th International Conference on Extending Database Technology*, Heraklion, Grete, Greece, 2004
- [24] M. Corduas ve D. Piccolo, "Time Series Clustering and Classification by the Autoregressive Metric", *Computational Statistics & Data Analysis*, cilt 52, sayı 4, ss. 1860-1872, 2008.
- [25] T. A. Runkler ve H. G. Seeding, "Fuzzy C-Auto Regression Models", *IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, Hong Kong, China, 2008.
- [26] N. Güler ve M. Karahasan, "Zaman Serileri Kümelemesi Yaklaşımı ile OECD Ülkelerinin İthalatının Tahmini", *Finans Politik & Ekonomik Yorumlar*, cilt 50, sayı 586, ss. 61-72, 2008.
- [27] S. Ghassempour, F. Girosi ve A. Maeder, "Clustering Multivariate Time Series Using Hidden Markov Models", *International Journal of Environment Research and Public Health*, cilt 11, sayı 3, ss. 2741-2763, 2014.
- [28] E. Keogh ve S. Kasetty, "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration", *Data Mining and Knowledge Discovery*, cilt 7, sayı 4, ss. 349-371, 2003.

- [29] W. L. Martinez ve A. R. Martinez, *Exploratory Data Analysis with MATLAB*, Chapman & CRC, Mishawaka, USA, 2005.
- [30] Z. Çakmak, "Kümeleme Analizinde Geçerlilik Problemi ve Kümeleme Sonuçlarının Değerlendirilmesi", *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, sayı 3, ss. 187-205, 1999.
- [31] P. Tan, M. Steinbach ve V. Kumar, *Introduction To Data Mining*, Birinci Baskı, Addison Wesley, USA, 2006.
- [32] T. Kohonen, *The Self Organizing Maps*, Üçüncü Baskı Springer Series in Information Sciences, 2001.
- [33] M. Zontul, O. Kaynar ve H. Bircan, (2004), "SOM Tipimde Yapay Sinir Ağlarını Kullanarak Türkiye' nin İthalat Yaptığı Ülkelerin Kümelenmesi Üzerine Bir Çalışma", *Cumhuriyet Üniversitesi, İktisadi ve İdari Bilimler Dergisi* sayı 2, ss. 47-68, 2004.
- [34] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics*, cilt 3, ss. 32-57, 1973.
- [35] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms", Springer US, New York, 1981.
- [36] S. Atal, "Bulanık Kümeleme Analizi ve OECD Ülkelerinin Gelişmişlik Bakımından Kümelenmesi", *Yüksek Lisans Tezi, Eskişehir Osman Gazi Üniversitesi*, 2015.
- [37] C. Koçak, "Bulanık Zaman Serileri Öngörüsü İçin Yeni Bir Model Sınıfı", *Doktora Tezi, Ondokuz Mayıs Üniversitesi, Samsun*, 2011.
- [38] M. Halkidi, Y. Batistakis ve M. Vazirgiannis, "Cluster Validity Methods: Part I", *SIGMOD Record*, cilt 31, sayı 2, ss. 40-45, 2002.
- [39] M. Halkidi, Y. Batistakis ve M. Vazirgiannis, "Cluster Validity Methods: Part II", *SIGMOD Record*, cilt 31, sayı 3, ss. 19-27, 2002.
- [40] M. Vatansever ve A. Hakan Büyüklü, "Görsel Veri Madenciliği Tekniklerinin Kümeleme Analizlerinde Kullanımı ve Uygulanması", *Sigma Journal of Engineering and Natural Sciences*, cilt 2, sayı 2, ss. 83-104, 2009.
- [41] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis", *Journal of Computational and Application Mathematics* vol. 20, pp. 53-65, 1987.
- [42] J. Dunn (1974), "Well Separated Clusters and Optimal Fuzzy Partitions", *Journal of Cybernetics*, cilt 4, ss. 95-104.
- [43] T. Calinski ve J. Harabasz, "A Dendrite Method for Cluster Analysis", *Communication in Statistics*, cilt 3, sayı 1, ss. 1-27, 1974.
- [44] G. Shu, B. Zeng, Y. P. Chen ve O. H. Smith O., "Performance Assessment of Kernel -Density Clustering for Gene Expression Profile Data", *Comparative and Functional Genomics*, cilt 4, sayı 3, ss. 287-299, 2003.

- [45] D. L. Davies ve D.W. Bouldin, "A Cluster Separation Measure.", IEEE Transactions on Pattern Analysis and Machine Intelligence, cilt 1, sayı 2, ss. 224-227, 1979.
- [46] N. Bolshakova ve F. Azuaje, "Cluster Validation Techniques for Genome Expression Data", Signal Processing, cilt 83, sayı 4, ss. 825-833, 2003.
- [47] L. J. Hubert ve J. R. Levin, "A General Statistical Framework for Assessing Categorical Clustering in Free Recall", Psychological Bulletin, cilt 83, sayı 6, ss. 1072-1080, 1976.
- [48] Y. Lechevallier, M. B. Ahmed, ve G. Saporta, "On the Number of Clusters in Block Clustering Algorithms", Proceedings of the Twenty-Third International Florida Artificial Intelligence Research Society Conference, ss. 392-397, 2010
- [49] W. Krzanowski ve Y. Lai, "A Criterion for Determining the Number of Groups in a Data Set using Sum-of-Squares Clustering" Biometrics cilt 44, sayı 1, ss. 23-34, 1988.
- [50] S. Dudoit, J. Fridlyand, "A Prediction-Based Resampling Method for Estimating the Number of Clusters in a Dataset", Genome Biology, cilt 3, sayı 7, ss. 1-21, 2002.
- [51] R. Saharan, A. Maron-Katz ve R. Shamir, "CLICK and EXPANDER: A System for Clustering and Visualizing Gene Expression Data", Bioninformatics, cilt 19, sayı 14, ss. 1787-1799, 2003.
- [52] M. Vatansever ve İ. Demir, "A Segmentation Study of Non-Performing Loans Rates in Turkish Credit Market", International Business Research, cilt 10, sayı 11, ss. 29- 41, 2017.
- [53] Y. Ding ve R. F. Harrison, "Relational visual cluster validity (RVCV)", Pattern Recognition Letters, cilt 28, sayı 15, ss. 2071-2079, 2007.
- [54] C. Chatfield, The Analysis of Time Series: An Introduction. Chapman, London, 1989
- [55] C. Koçak, "Bulanık Zaman Serileri Öngörüsü için Yeni Bir Model Sınıfı", Doktora Tezi, Ondokuz Mayıs Üniversitesi, Samsun, 2011.
- [56] G. E. P. Box, G. M. Jenkins, Time Series Analysis: Forecasting and Control, Holden Day, San Francisco, 1976
- [57] D. N. Gujarati ve D. C. Porter, Basic Econometrics, McGraw-Hill Education, New York, 4th Edition, 2009.
- [58] S.E. Said and D.A. Dickey, "Testing for Unit Roots in Autoregressive-Moving Average Models of Unkown Order", Biometrika, cilt 71, ss. 599-607.
- [59] P. C. B. Phillips ve P. Perron, "Testing for a Unit Root in Time Series Regression", Biometrika, cilt 75, sayı 2, ss. 335-346, 1988.
- [60] G. Schwert, "Tests for Unit Roots: A Monte Carlo Investigation", Journal of Business & Economics Statistics, 1989, cilt 7, sayı 2, ss. 147-159.

- [61] D. Kwiatkowski, P. C. B. Phillips, P. Schmidt ve Y. Shin, "Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root", *Journal of Econometrics*, cilt 54, ss. 159-178, 1992
- [62] F. J. Fobozzi, S. M. Focardi, S. T. Rachev ve B. G. Arshanapalli, "The Basics of Financial Econometrics: Tools, Concepts AND Asset Management Applications", John Wiley & Sons, 2014.
- [63] G. Schwarz, "Estimating the Dimension of a Model", *The Annals of Statistics*, sayı 6, cilt 2, ss. 461- 464, 1978.
- [64] E. Uğurlu, Durağanlık ve Birim Kök Sınamaları, İstanbul Aydın Üniversitesi Ekonomi ve Finansal Bölümü, 2009.
- [65] G. E. P. Box ve D. A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive Integrated Moving Average Time Series Models", *Journal of the American Statistical Association*, cilt 65, ss. 1509-1526, 1970.
- [66] G. M. Ljung ve G. E. P. Box, "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, cilt 65, ss. 297-303, 1978.
- [67] A. J. Bagnall, G. Janacek ve M. Zhang, "Clustering Time Series from Mixture Polynomial Models with Discretized Data", Technical Report CMP-C03-17, School of Computer Sciences, University of East Anglia, England, 2003 [Online]. Hazır: <https://core.ac.uk/download/pdf/2770326.pdf> [Erişim 10 Nisan, 2019]
- [68] Hazır: <https://www.reidin.com/en/> [Erişim 03 Mayıs, 2019]
- [69] Bezdek J. C. ve Hathaway R. J., "VAT: A Tool For Visual Assesment Of (Cluster) Tendency", *Proceedings of the 2002 International Joint Conference on, Honolulu, HI, USA, May, 2002*, 2225-2230.
- [70] Wang, K.J.(2007), Cluster Validation ToolBox CVAP, Hazır: <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do?objectId=14620> [Erişim 10 Nisan, 2019]

Tablo A.1 Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 1

Şehirler	Mahalleler	Temel Dönem	Ortalama	Minimum	Maksimum	Std. Sap.
Ankara	19 Mayıs	June 2007=100	143.78	95.35	230.35	42.31
Ankara	23 Nisan	Jan 2012=100	112.81	85.36	141.07	16.00
Kocaeli	28 Haziran	Jan 2015=100	97.09	69.68	145.69	20.72
İstanbul	Abbasağa	June 2007=100	135.42	79.32	233.76	51.04
İstanbul	Acıbadem	June 2007=100	124.18	90.10	194.80	34.85
Ankara	Adnan Menderes	Jan 2012=100	112.75	85.36	151.67	17.75
Ankara	Ahlatlıbel	Jan 2015=100	94.77	68.68	129.83	17.57
Bursa	Ahmetpaşa	Jan 2012=100	108.90	85.97	158.23	17.31
Kocaeli	Akarca	Jan 2012=100	119.64	85.97	188.39	26.32
İstanbul	Akat	June 2007=100	188.91	82.67	370.68	90.43
İzmir	Akın Simav	June 2007=100	108.94	82.58	171.79	26.23
Ankara	Akpınar	Jan 2015=100	93.67	68.68	119.35	16.18
İzmir	Aksoy	June 2007=100	143.71	94.69	266.79	52.29
Ankara	Aktepe	June 2007=100	127.79	95.24	176.19	25.78
İzmir	Alaybey	June 2007=100	137.82	96.10	248.87	46.86
Kocaeli	Alikahya Merkez	June 2007=100	127.23	105.33	172.06	18.06
İzmir	Alsancak	June 2007=100	104.21	80.90	145.17	19.77
Bursa	Altınşehir	June 2007=100	130.06	85.29	224.25	44.13
İzmir	Altıntaş	June 2007=100	123.42	90.20	212.09	34.93
Bursa	Altıparmak	June 2007=100	100.26	86.97	153.79	16.15
Ankara	Anıttepe	June 2007=100	114.80	76.49	187.45	34.59
İstanbul	Arnavutköy	June 2007=100	151.86	88.59	186.79	33.44
Ankara	Aşağı Dikmen	June 2007=100	106.55	78.07	169.35	26.77
Ankara	Aşağı Eğlence	Jan 2012=100	103.92	85.36	146.03	12.80
Ankara	Aşağı Öveçler	Jan 2015=100	95.12	68.68	126.51	18.12
Ankara	Ata	Jan 2015=100	91.40	68.68	114.66	13.64
Bursa	Ataevler	June 2007=100	122.56	91.77	219.98	38.13
İzmir	Atakent	June 2007=100	153.01	95.06	262.41	51.93
Ankara	Atapark	June 2007=100	121.32	91.92	156.55	19.91
İzmir	Atilla	Jan 2012=100	127.28	85.13	214.51	34.44
Ankara	Aydınlar	Jan 2012=100	121.21	85.36	163.76	25.65
Ankara	Ayrancı	June 2007=100	118.73	82.89	161.57	27.06
Ankara	Ayvalı	Jan 2012=100	111.48	85.36	136.24	14.57
Ankara	Aziziye	Jan 2015=100	92.61	68.68	118.77	14.89
Ankara	Bademlik	June 2007=100	127.84	95.33	168.05	22.12



**Tablo A.2 Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 2**

Şehirler	Mahalleler	Temel Dönem	Ortalama	Minimum	Maksimum	Std. Sap.
Ankara	Bağlarbaşı	Jan 2012=100	113.81	85.36	147.10	17.56
İzmir	Bahariye	Jan 2012=100	127.96	85.13	214.71	36.74
Ankara	Bahçelievler	June 2007=100	138.20	96.10	236.65	44.73
İzmir	Bahriye Üçok	June 2007=100	142.69	95.21	239.06	46.58
Bursa	Balat	Jan 2012=100	115.84	85.97	183.49	27.13
Ankara	Balgat	June 2007=100	119.20	83.24	181.79	30.60
İstanbul	Balmumcu	June 2007=100	182.42	85.98	341.08	77.91
Ankara	Barbaros	Jan 2015=100	92.24	68.68	123.27	14.79
Bursa	Bariş	June 2007=100	120.07	87.18	204.06	35.41
Ankara	Basınevleri	June 2007=100	110.78	85.32	157.17	22.13
İstanbul	Bebek	June 2007=100	231.68	82.89	469.63	112.49
Bursa	Beşevler	June 2007=100	120.18	89.31	212.05	36.72
Ankara	Beytepe	Jan 2015=100	89.86	68.68	104.38	11.64
Ankara	Birlik	June 2007=100	115.46	87.92	159.97	25.59
İstanbul	Bostancı	June 2007=100	170.71	91.95	276.06	62.48
İzmir	Bostanlı	June 2007=100	140.79	92.60	248.46	48.49
Ankara	Büyükesat	June 2007=100	121.49	83.09	185.29	31.45
İstanbul	Caddebostan	June 2007=100	179.57	95.75	290.07	66.07
İstanbul	Caferağa	June 2007=100	179.50	95.75	307.74	71.36
Ankara	Cebeci	June 2007=100	104.98	83.66	163.71	24.09
Kocaeli	Cedit	June 2007=100	135.80	107.67	222.38	31.05
Ankara	Cevizlidere	June 2007=100	111.45	84.35	163.05	27.64
İstanbul	Cihannüma	Jan 2012=100	126.87	86.80	177.11	28.74
Bursa	Cumhuriyet	June 2007=100	129.48	91.72	235.68	41.58
Ankara	Çaldıran	Jan 2012=100	112.40	85.36	147.50	16.76
Ankara	Çalseki	Jan 2012=100	117.66	85.36	155.03	22.18
Bursa	Çamlıca	Jan 2012=100	120.50	85.97	190.43	27.35
Ankara	Çamlıtepe	June 2007=100	128.17	97.19	198.05	31.04
Ankara	Çankaya	Jan 2015=100	123.61	90.20	206.25	35.54
Ankara	Çayyolu	June 2007=100	122.79	88.65	167.18	23.58
Bursa	Çekirge	June 2007=100	102.88	82.20	172.69	25.19
Bursa	Çırpan	Jan 2012=100	111.47	85.97	169.98	19.66
Ankara	Çiçekli	Jan 2012=100	113.66	85.36	153.43	18.53
Ankara	Çiğdem	June 2007=100	119.36	85.77	172.56	27.48
Ankara	Çukurambar	Jan 2015=100	94.08	68.68	120.96	16.73

**Tablo A.3 Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 3**

Şehirler	Mahalleler	Temel Dönem	Ortalama	Minimum	Maksimum	Std. Sap.
İzmir	Dedebaşı	Jan 2012=100	130.13	85.13	205.46	35.89
Kocaeli	Merkez	June 2007=100	109.94	92.84	156.00	17.38
Bursa	Demirci	Jan 2012=100	123.60	85.97	195.79	31.00
İzmir	Demirköprü	June 2007=100	147.06	93.57	246.85	44.74
Bursa	Cumhuriyet	Jan 2015=100	95.24	69.96	142.34	17.99
Bursa	Demirtaşpaşa	Jan 2012=100	116.24	85.97	164.48	22.97
İstanbul	Dikilitaş	June 2007=100	130.46	76.76	239.49	48.49
Bursa	Dikkaldırım	June 2007=100	117.63	88.61	195.08	31.63
Bursa	Doburca	Jan 2012=100	122.14	85.97	201.48	32.40
İzmir	Donanmacı	June 2007=100	134.81	96.10	249.73	45.41
Bursa	Dumlupınar	June 2007=100	181.36	88.85	312.18	82.72
İstanbul	Eğitim	June 2007=100	164.54	95.75	315.10	74.39
Ankara	Emek	June 2007=100	131.64	90.24	190.77	33.20
Ankara	Emrah	Jan 2012=100	112.88	85.36	160.19	18.57
İstanbul	Erenköy	June 2007=100	179.07	94.80	299.47	73.35
Bursa	Ertuğrul	June 2007=100	132.29	91.75	228.32	41.45
Ankara	Esertepe	June 2007=100	127.26	95.34	171.33	23.03
İstanbul	Etiler	June 2007=100	162.28	87.46	219.78	41.76
Ankara	Etlük	June 2007=100	127.36	95.33	184.20	26.71
İstanbul	Fenerbahçe	June 2007=100	165.76	90.99	278.28	63.28
İstanbul	Feneryolu	June 2007=100	166.56	95.75	273.75	62.36
Bursa	Fethiye	June 2007=100	114.09	87.87	187.72	29.92
İstanbul	Fikirtepe	June 2007=100	212.39	96.63	378.57	111.03
İzmir	Fikri Altay	Jan 2012=100	129.66	85.13	213.50	35.55
İstanbul	Gayrettepe	June 2007=100	148.11	83.32	254.70	55.11
Ankara	Gaziosmanpaşa	June 2007=100	118.26	86.72	184.87	30.59
İzmir	Goncalar	Jan 2012=100	130.98	85.13	209.03	36.24
Bursa	Görükle	June 2007=100	119.86	93.04	201.38	26.28
İstanbul	Göztepe	June 2007=100	167.05	95.64	275.01	64.15
Ankara	Güçlükaya	Jan 2012=100	118.34	85.36	167.96	21.84
Ankara	Gümüşdere	Jan 2012=100	110.95	85.36	134.72	14.23
Ankara	Güvenevler	Jan 2012=100	122.51	85.36	162.04	25.35
İzmir	Güzelyalı	June 2007=100	128.15	89.63	218.62	36.69
Ankara	Harbiye	June 2007=100	116.55	87.90	181.73	32.07
İstanbul	Hasanpaşa	June 2007=100	166.97	93.69	290.97	70.29

**Tablo A.4 Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 4**

Şehirler	Mahalleler	Temel Dönem	Ortalama	Minimum	Maksimum	Std. Sap.
Ankara	Hilal	Jan 2015=100	91.56	68.68	109.56	13.47
Bursa	Hocahasan	Jan 2012=100	108.88	85.97	160.74	17.39
Ankara	Huzur	Jan 2012=100	127.34	85.36	175.80	30.35
Bursa	Hüdavendigar	Jan 2012=100	119.77	85.97	198.72	29.66
Bursa	İbrahimpaşa	Jan 2012=100	108.38	85.97	161.72	18.10
Bursa	İhsaniye	June 2007=100	121.48	90.02	205.57	35.07
Ankara	İlkadım	June 2007=100	98.00	75.45	152.85	24.45
Ankara	İlkbahar	Jan 2015=100	90.02	68.68	106.24	11.89
Ankara	İlker	June 2007=100	119.08	84.28	178.38	29.56
Ankara	İncirli	June 2007=100	117.97	92.65	156.61	22.10
Ankara	İşçi Blokları	Jan 2012=100	132.45	85.36	189.27	32.48
Kocaeli	Kadıköy	Jan 2012=100	117.65	85.97	173.15	23.40
Ankara	Kalaba	June 2007=100	120.04	94.22	168.97	23.20
Ankara	Kamil Ocak	Jan 2012=100	111.52	85.36	139.52	14.68
Ankara	Kanuni	June 2007=100	140.08	94.39	197.11	35.63
Ankara	Karapınar	Jan 2012=100	127.40	85.36	170.37	29.82
Ankara	Subayevleri	Jan 2012=100	115.78	85.36	152.54	18.63
Ankara	Kavaklıdere	June 2007=100	121.65	89.34	165.67	25.35
Ankara	Keklik Pınarı	June 2007=100	112.17	81.41	163.54	28.09
İzmir	Kılıç Reis	Jan 2012=100	127.17	85.13	217.31	34.01
Ankara	Kırkkonaklar	June 2007=100	115.67	86.95	163.98	27.58
Ankara	Kızılay	June 2007=100	124.25	96.55	169.43	23.54
Bursa	Kirazlı	Jan 2015=100	98.92	69.96	142.74	22.90
Bursa	Konak	Jan 2012=100	127.76	85.97	223.98	37.60
İstanbul	Konaklar	Jan 2012=100	126.39	86.80	193.11	30.31
İstanbul	Koşuyolu	June 2007=100	149.23	91.30	257.73	55.78
İstanbul	Kozyatağı	June 2007=100	176.23	95.75	310.72	68.04
Ankara	Köşk	Jan 2012=100	113.97	85.36	145.39	16.42
İstanbul	Kuruçeşme	June 2007=100	133.88	83.12	245.14	52.45
Ankara	Kuşçagız	June 2007=100	119.62	90.93	158.84	21.40
Ankara	Küçüksat	June 2007=100	121.56	92.71	179.47	24.26
Bursa	Kükürtlü	June 2007=100	111.17	87.52	197.11	30.32
Ankara	Kültür	Jan 2015=100	120.97	85.97	203.50	31.27
İstanbul	Levazım	June 2007=100	251.36	86.91	467.74	115.29
İstanbul	Levent	June 2007=100	141.66	84.47	246.05	50.42

**Tablo A.5 Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 5**

Şehirler	Mahalleler	Temel Dönem	Ortalama	Minimum	Maksimum	Std. Sap.
Bursa	Maksem	Jan 2012=100	107.04	85.97	154.73	15.65
Ankara	Maltepe	June 2007=100	129.15	97.93	184.49	27.56
İzmir	Mavişehir	June 2007=100	144.21	94.70	240.07	45.36
İstanbul	Mecidiye	Jan 2012=100	125.16	86.80	187.21	30.67
İstanbul	Merdivenköy	June 2007=100	171.23	95.75	319.43	72.22
Kocaeli	Merkez	June 2007=100	106.05	92.51	132.50	12.52
İzmir	Mithatpaşa	June 2007=100	97.26	72.88	140.48	18.23
İstanbul	Muradiye	June 2007=100	140.83	82.47	240.25	49.55
Ankara	Mutlukent	Jan 2012=100	132.12	85.36	199.40	33.66
İzmir	Namık Kemal	June 2007=100	109.44	89.55	122.56	10.47
İzmir	Nergiz	June 2007=100	164.46	96.10	271.60	55.83
İstanbul	Nispetiye	Jan 2012=100	145.45	86.80	227.20	49.98
Ankara	Oran	June 2007=100	111.75	82.69	169.06	29.17
İstanbul	Ortaköy	June 2007=100	137.13	82.54	250.11	49.53
İstanbul	Osmanağa	Jan 2012=100	145.26	86.80	259.64	51.06
Ankara	Osmangazi	Jan 2012=100	114.10	85.36	141.51	17.06
Bursa	Ovaakça Merkez	Jan 2012=100	111.45	85.97	161.37	19.22
İzmir	Örnekköy	June 2007=100	137.15	96.10	218.14	36.64
Ankara	Öveçler	June 2007=100	108.20	82.88	153.48	24.35
Bursa	Özlüce	June 2007=100	135.14	93.04	219.87	43.92
Ankara	Pınarbaşı	Jan 2012=100	114.95	85.36	148.13	18.83
Ankara	Taner Kışlalı	Jan 2012=100	108.09	85.36	136.27	14.41
İstanbul	Rasimpaşa	June 2007=100	186.00	95.75	334.39	86.82
İstanbul	Sahrayı Cedit	June 2007=100	166.35	95.75	284.22	63.88
Ankara	Sancak	June 2007=100	130.74	96.13	180.29	28.50
Ankara	Sancaktepe	Jan 2012=100	121.73	85.36	171.66	24.27
Ankara	Seyranbağları	June 2007=100	118.86	90.39	182.82	29.98
İstanbul	Sinanpaşa	Jan 2012=100	125.79	86.80	200.90	34.23
Ankara	Paşa	June 2007=100	113.69	83.15	173.11	30.50
İstanbul	Suadiye	June 2007=100	174.47	95.75	278.74	64.64
Ankara	Şefkat	Jan 2012=100	122.16	85.36	180.90	27.78
Ankara	Şehit Kubilay	Jan 2012=100	112.30	85.36	144.80	16.33
İzmir	Şemikler	June 2007=100	150.77	95.20	257.22	48.71
Ankara	Şenlik	June 2007=100	123.86	94.62	168.53	25.29
Ankara	Tepebaşı	June 2007=100	115.16	91.72	170.65	20.53

**Tablo A.6 Konut Satış Fiyat Endeksleri Gösterge İstatistikleri 6**

Şehirler	Mahalleler	Temel Dönem	Ortalama	Minimum	Maksimum	Std. Sap.
İstanbul	Türkali	June 2007=100	137.84	82.75	234.79	54.16
Ankara	Ufuktepe	June 2007=100	120.60	92.63	153.53	20.12
İstanbul	Ulus	June 2007=100	190.49	85.32	277.64	66.88
Ankara	Uyanış	June 2007=100	128.55	93.66	196.32	31.11
Bursa	Üçevler	June 2007=100	118.80	84.76	197.66	31.44
Ankara	Ümit	June 2007=100	131.57	91.14	183.28	25.37
İstanbul	Vişnezade	June 2007=100	154.24	84.16	238.51	50.14
Kocaeli	Yahyakaptan	June 2007=100	141.44	104.24	247.67	42.69
Ankara	Yakacık	Jan 2012=100	121.81	85.36	178.77	25.89
İzmir	Yalı	June 2007=100	153.94	96.10	259.35	47.90
Ankara	Yaşamkent	June 2007=100	134.30	91.06	185.40	29.48
Ankara	Yayla	June 2007=100	130.46	95.67	174.06	22.86
Kocaeli	Yenişehir	June 2007=100	129.34	108.07	208.88	25.76
İstanbul	Yıldız	June 2007=100	127.71	82.50	204.54	34.01
Ankara	Yıldızevler	June 2007=100	101.18	77.26	158.46	25.95
Ankara	Yukarı Dikmen	June 2007=100	134.85	93.14	198.24	33.88
Bursa	Yunuseli	June 2007=100	141.20	94.90	235.49	42.77
Ankara	Yükseltepe	Jan 2012=100	104.46	85.36	126.10	11.05
Bursa	Yüzüncüyıl	Jan 2012=100	125.47	85.97	207.77	34.61
İzmir	Zafertepe	Jan 2012=100	128.89	85.13	205.92	34.52
İstanbul	Zühtüpaşa	Jan 2015=100	92.22	69.44	112.74	14.39

## B.1 Durağanlık Testi R Kodları

### B.1.1 Sıfıncı Mertebeden R Birim Kök Durağanlık Testi Kodları

```

library(readxl)
library(tseries)
Sale_Indices <- read_excel("data path ...")
Sale_Indices=ts(Sale_Indices, start = c(2010,01), frequency = 12)

# ADF Test #
l <- length(Sale_Indices[1,])
namedistrict <- matrix(, nrow = l, ncol = 1)
statistics <- matrix(, nrow = l, ncol = 1)
parameter <- matrix(, nrow = l, ncol = 1)
p.value <- matrix(, nrow = l, ncol = 1)
method <- matrix(, nrow = l, ncol = 1)
data.name <- matrix(, nrow = l, ncol = 1)

for(i in 1:l)
{
  results <- adf.test(Sale_Indices[,i], k=10)
  namedistrict[i,1] <- names(Sale_Indices[1,i])
  statistics[i,1] <- results$statistic
  parameter[i,1] <- results$parameter
  p.value[i,1] <- results$p.value
  method[i,1] <- results$method
  data.name[i,1] <- results$data.name
}
stationary_results <- data.frame(namedistrict, statistics, parameter, p.value,
method, data.name)
write.csv(stationary_results, file = "adf_test_results.csv")
rm(l)
rm(i)
rm(results)
rm(namedistrict)
rm(statistics)
rm(parameter)
rm(p.value)
rm(method)
rm(data.name)
rm(stationary_results)

```

```

# PP Test #

l <- length(Sale_Indices[1,])
namedistrict <- matrix(, nrow = l, ncol = 1)
statistics <- matrix(, nrow = l, ncol = 1)
parameter <- matrix(, nrow = l, ncol = 1)
p.value <- matrix(, nrow = l, ncol = 1)
method <- matrix(, nrow = l, ncol = 1)
data.name <- matrix(, nrow = l, ncol = 1)

for(i in 1:l)
{
  results <- pp.test(Sale_Indices[i])
  namedistrict[i,1] <- names(Sale_Indices[1,i])
  statistics[i,1] <- results$statistic
  parameter[i,1] <- results$parameter
  p.value[i,1] <- results$p.value
  method[i,1] <- results$method
  data.name[i,1] <- results$data.name
}

stationary_results <- data.frame(namedistrict, statistics, parameter, p.value,
method, data.name)
write.csv(stationary_results, file = "pp_test_results.csv")

rm(l)
rm(i)
rm(results)
rm(namedistrict)
rm(statistics)
rm(parameter)
rm(p.value)
rm(method)
rm(data.name)
rm(stationary_results)

```

```

# KPSS Test #

l <- length(Sale_Indices[1,])
namedistrict <- matrix(, nrow = l, ncol = 1)
statistics <- matrix(, nrow = l, ncol = 1)
parameter <- matrix(, nrow = l, ncol = 1)
p.value <- matrix(, nrow = l, ncol = 1)
method <- matrix(, nrow = l, ncol = 1)
data.name <- matrix(, nrow = l, ncol = 1)

for(i in 1:l)
{
  results <- kpss.test(Sale_Indices[i], null = "Trend", lshort = FALSE)
  namedistrict[i,1] <- names(Sale_Indices[1,i])
  statistics[i,1] <- results$statistic
  parameter[i,1] <- results$parameter
  p.value[i,1] <- results$p.value
  method[i,1] <- results$method
  data.name[i,1] <- results$data.name
}

stationary_results <- data.frame(namedistrict, statistics, parameter, p.value,
method, data.name)
write.csv(stationary_results, file = "kpss_test_results.csv")

rm(l)
rm(i)
rm(results)
rm(namedistrict)
rm(statistics)
rm(parameter)
rm(p.value)
rm(method)
rm(data.name)
rm(stationary_results)

```



### B.1.2 Birinci Mertebeden R Birim Kök Duraganlık Testi Kodları

```
library(readxl)
library(tseries)

Sale_Indices <- read_excel("data path ...")
Sale_Indices=ts(Sale_Indices, start = c(2010,01), frequency = 12)

# ADF Test #

l <- length(Sale_Indices_D1[1,])
namedistrict <- matrix(, nrow = l, ncol = 1)
statistics <- matrix(, nrow = l, ncol = 1)
parameter <- matrix(, nrow = l, ncol = 1)
p.value <- matrix(, nrow = l, ncol = 1)
method <- matrix(, nrow = l, ncol = 1)
data.name <- matrix(, nrow = l, ncol = 1)

for(i in 1:l)
{
  results <- adf.test(Sale_Indices_D1[,i], k=10)
  namedistrict[i,1] <- names(Sale_Indices[1,i])
  statistics[i,1] <- results$statistic
  parameter[i,1] <- results$parameter
  p.value[i,1] <- results$p.value
  method[i,1] <- results$method
  data.name[i,1] <- results$data.name
}

stationary_results <- data.frame(namedistrict, statistics, parameter, p.value,
method, data.name)
write.csv(stationary_results, file = "ADF_Test_Tesults_D1.csv")

rm(l)
rm(i)
rm(results)
rm(namedistrict)
rm(statistics)
rm(parameter)
rm(p.value)
rm(method)
rm(data.name)
rm(stationary_results)
rm(Sale_Indices_D1)
```

```

# PP Test #

l <- length(Sale_Indices_D1[1,])
namedistrict <- matrix(, nrow = l, ncol = 1)
statistics <- matrix(, nrow = l, ncol = 1)
parameter <- matrix(, nrow = l, ncol = 1)
p.value <- matrix(, nrow = l, ncol = 1)
method <- matrix(, nrow = l, ncol = 1)
data.name <- matrix(, nrow = l, ncol = 1)

for(i in 1:l)
{
  results <- pp.test(Sale_Indices_D1[i])
  namedistrict[i,1] <- names(Sale_Indices[1,i])
  statistics[i,1] <- results$statistic
  parameter[i,1] <- results$parameter
  p.value[i,1] <- results$p.value
  method[i,1] <- results$method
  data.name[i,1] <- results$data.name
}

stationary_results <- data.frame(namedistrict, statistics, parameter, p.value,
method, data.name)
write.csv(stationary_results, file = "PP_Test_Tesults_D1.csv")

rm(l)
rm(i)
rm(results)
rm(namedistrict)
rm(statistics)
rm(parameter)
rm(p.value)
rm(method)
rm(data.name)
rm(stationary_results)
rm(Sale_Indices_D1)

```

```

# KPSS Test #

l <- length(Sale_Indices_D1[1,])
namedistrict <- matrix(, nrow = l, ncol = 1)
statistics <- matrix(, nrow = l, ncol = 1)
parameter <- matrix(, nrow = l, ncol = 1)
p.value <- matrix(, nrow = l, ncol = 1)
method <- matrix(, nrow = l, ncol = 1)
data.name <- matrix(, nrow = l, ncol = 1)

for(i in 1:l)
{
  results <- kpss.test(Sale_Indices_D1[i], null = "Trend", lshort = FALSE)
  namedistrict[i,1] <- names(Sale_Indices[1,i])
  statistics[i,1] <- results$statistic
  parameter[i,1] <- results$parameter
  p.value[i,1] <- results$p.value
  method[i,1] <- results$method
  data.name[i,1] <- results$data.name
}

stationary_results <- data.frame(namedistrict, statistics, parameter, p.value,
method, data.name)
write.csv(stationary_results, file = "KPSS_Test_Tesults_D1.csv")

rm(l)
rm(i)
rm(results)
rm(namedistrict)
rm(statistics)
rm(parameter)
rm(p.value)
rm(method)
rm(data.name)
rm(stationary_results)
rm(Sale_Indices_D1)

```

## B.2 Konut Satış Fiyat Endeksleri ADF, PP ve KPSS Birim Kök Test Sonuçları (Sıfıncı Mertebeden)

Tablo B.1 Sıfıncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 1

Mahalleler	ADF Test Olasılığı	PP Test Olasılığı	KPSS Test Olasılığı	Sonuçlar
19 May	0.530	0.968	0.010	Durağan Değil
23 Nisan	0.668	0.676	0.017	Durağan Değil
28 Haziran	0.982	0.946	0.010	Durağan Değil
Abbasağa	0.990	0.990	0.010	Durağan Değil
Acıbadem	0.715	0.990	0.010	Durağan Değil
Adnan Menderes	0.420	0.729	0.100	Durağan Değil
Ahlalibel	0.741	0.835	0.010	Durağan Değil
Ahmetpaşa	0.672	0.990	0.100	Durağan Değil
Akarca	0.968	0.980	0.048	Durağan Değil
Akat	0.580	0.977	0.010	Durağan Değil
Akın Simav	0.990	0.990	0.010	Durağan Değil
Akpınar	0.381	0.681	0.010	Durağan Değil
Aksoy	0.990	0.990	0.010	Durağan Değil
Aktepe	0.960	0.977	0.010	Durağan Değil
Alaybey	0.990	0.990	0.010	Durağan Değil
Alikahya Merkez	0.076	0.561	0.035	Durağan Değil
Alsancak	0.657	0.882	0.010	Durağan Değil
Altınşehir	0.990	0.990	0.010	Durağan Değil
Altıntaş	0.990	0.990	0.010	Durağan Değil
Altıparmak	0.990	0.990	0.010	Durağan Değil
Anıttepe	0.806	0.857	0.010	Durağan Değil
Arnavutköy	0.990	0.990	0.010	Durağan Değil
Aşağı Dikmen	0.989	0.990	0.015	Durağan Değil
Aşağı Eğlence	0.228	0.614	0.100	Durağan Değil
Aşağı Öveçler	0.984	0.889	0.010	Durağan Değil
Ata	0.410	0.716	0.033	Durağan Değil
Ataevler	0.307	0.990	0.010	Durağan Değil
Atakent	0.990	0.990	0.010	Durağan Değil
Atapark	0.542	0.664	0.100	Durağan Değil
Atilla	0.832	0.990	0.045	Durağan Değil
Aydınlr	0.990	0.919	0.012	Durağan Değil
Ayrancı	0.582	0.874	0.010	Durağan Değil
Ayvalı	0.616	0.746	0.036	Durağan Değil
Aziziye	0.589	0.704	0.076	Durağan Değil
Bademlik	0.773	0.472	0.014	Durağan Değil
Bağlarbaşı	0.276	0.670	0.100	Durağan Değil
Bahariye	0.990	0.990	0.013	Durağan Değil
Bahçelievler	0.990	0.990	0.010	Durağan Değil
Bahriye Üçok	0.990	0.990	0.010	Durağan Değil
Balat	0.881	0.990	0.013	Durağan Değil

**Tablo B.2** Sıfıncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 2

<b>Mahalleler</b>	<b>ADF Test Olasılığı</b>	<b>PP Test Olasılığı</b>	<b>KPSS Test Olasılığı</b>	<b>Sonuçlar</b>
Balgat	0.426	0.846	0.010	Durağan Değil
Balmumcu	0.436	0.847	0.010	Durağan Değil
Barbaros	0.348	0.484	0.095	Durağan Değil
Barış	0.990	0.990	0.010	Durağan Değil
Basınevleri	0.921	0.925	0.010	Durağan Değil
Bebek	0.195	0.854	0.042	Durağan Değil
Beşevler	0.990	0.990	0.010	Durağan Değil
Beytepe	0.649	0.980	0.015	Durağan Değil
Birlik	0.990	0.987	0.010	Durağan Değil
Bostancı	0.744	0.853	0.010	Durağan Değil
Bostanlı	0.767	0.990	0.010	Durağan Değil
Büyükesat	0.490	0.761	0.010	Durağan Değil
Caddebostan	0.793	0.894	0.010	Durağan Değil
Caferağa	0.990	0.990	0.010	Durağan Değil
Cebeci	0.984	0.990	0.010	Durağan Değil
Cedit	0.987	0.990	0.010	Durağan Değil
Cevizlidere	0.702	0.990	0.010	Durağan Değil
Cihannüma	0.441	0.896	0.098	Durağan Değil
Cumhuriyet	0.990	0.990	0.010	Durağan Değil
Çaldıran	0.272	0.688	0.098	Durağan Değil
Çalseki	0.510	0.812	0.010	Durağan Değil
Çamlıca	0.990	0.989	0.010	Durağan Değil
Çamlıtepe	0.990	0.990	0.010	Durağan Değil
Çankaya	0.990	0.990	0.010	Durağan Değil
Çayyolu	0.728	0.987	0.040	Durağan Değil
Çekirge	0.990	0.990	0.010	Durağan Değil
Çirpan	0.109	0.990	0.100	Durağan Değil
Çiçekli	0.010	0.790	0.100	Durağan Değil
Çiğdem	0.426	0.715	0.010	Durağan Değil
Çukurambar	0.923	0.853	0.010	Durağan Değil
Dedebaşı	0.716	0.990	0.010	Durağan Değil
Değirmendere Merkez	0.822	0.764	0.014	Durağan Değil
Demirci	0.949	0.990	0.010	Durağan Değil
Demirköprü	0.912	0.957	0.010	Durağan Değil
Demirtaş Cumhuriyet	0.173	0.447	0.030	Durağan Değil
Demirtaşpaşa	0.804	0.985	0.040	Durağan Değil
Dikilitaş	0.958	0.897	0.010	Durağan Değil
Dikkaldırım	0.990	0.990	0.010	Durağan Değil
Doburca	0.990	0.990	0.010	Durağan Değil
Donanmacı	0.990	0.990	0.010	Durağan Değil
Dumlupınar	0.289	0.951	0.010	Durağan Değil
Eğitim	0.990	0.990	0.010	Durağan Değil
Emek	0.884	0.778	0.010	Durağan Değil
Emrah	0.415	0.858	0.088	Durağan Değil
Erenköy	0.990	0.980	0.010	Durağan Değil

**Tablo B.3** Sıfırıncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 3

<b>Mahalleler</b>	<b>ADF Test Olasılığı</b>	<b>PP Test Olasılığı</b>	<b>KPSS Test Olasılığı</b>	<b>Sonuçlar</b>
Ertuğrul	0.990	0.990	0.010	Durağan Değil
Esertepe	0.523	0.564	0.100	Durağan Değil
Etiler	0.010	0.972	0.010	Durağan Değil
Etlük	0.976	0.988	0.010	Durağan Değil
Fenerbahçe	0.906	0.990	0.010	Durağan Değil
Feneryolu	0.888	0.990	0.010	Durağan Değil
Fethiye	0.819	0.990	0.010	Durağan Değil
Fikirtepe	0.488	0.960	0.010	Durağan Değil
Fikri Altay	0.835	0.990	0.015	Durağan Değil
Gayrettepe	0.959	0.984	0.010	Durağan Değil
Gaziosmanpaşa	0.854	0.981	0.010	Durağan Değil
Goncalar	0.976	0.990	0.010	Durağan Değil
Görükle	0.960	0.990	0.031	Durağan Değil
Göztepe	0.951	0.976	0.010	Durağan Değil
Güçlükaya	0.037	0.936	0.100	Durağan Değil
Gümüşdere	0.362	0.818	0.030	Durağan Değil
Güvenevler	0.407	0.500	0.037	Durağan Değil
Güzelyalı	0.990	0.990	0.010	Durağan Değil
Harbiye	0.666	0.990	0.010	Durağan Değil
Hasanpaşa	0.749	0.957	0.010	Durağan Değil
Hilal	0.983	0.788	0.074	Durağan Değil
Hocahasan	0.070	0.986	0.036	Durağan Değil
Huzur	0.526	0.853	0.037	Durağan Değil
Hüdavendigar	0.672	0.990	0.010	Durağan Değil
İbrahimpaşa	0.984	0.990	0.024	Durağan Değil
İhsaniye	0.990	0.990	0.010	Durağan Değil
İlkadım	0.990	0.990	0.010	Durağan Değil
İlkbahar	0.476	0.818	0.023	Durağan Değil
İlker	0.903	0.894	0.010	Durağan Değil
İncirli	0.849	0.975	0.010	Durağan Değil
İşçi Blokları	0.506	0.891	0.100	Durağan Değil
Kadıköy	0.538	0.663	0.024	Durağan Değil
Kalaba	0.921	0.987	0.016	Durağan Değil
Kamil Ocak	0.910	0.615	0.014	Durağan Değil
Kanuni	0.302	0.711	0.019	Durağan Değil
Karapınar	0.414	0.690	0.010	Durağan Değil
Kavacık Subayevleri	0.544	0.861	0.034	Durağan Değil
Kavaklıdere	0.894	0.903	0.010	Durağan Değil
Keklik Pınarı	0.420	0.911	0.010	Durağan Değil
Kılıç Reis	0.533	0.990	0.067	Durağan Değil
Kırkkonaklar	0.649	0.964	0.010	Durağan Değil
Kızılay	0.538	0.740	0.010	Durağan Değil
Kirazlı	0.990	0.990	0.010	Durağan Değil
Konak	0.716	0.990	0.021	Durağan Değil
Konaklar	0.129	0.937	0.100	Durağan Değil

**Tablo B.4** Sıfıncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 4

<b>Mahalleler</b>	<b>ADF Test Olasılığı</b>	<b>PP Test Olasılığı</b>	<b>KPSS Test Olasılığı</b>	<b>Sonuçlar</b>
Koşuyolu	0.990	0.990	0.010	Durağan Değil
Kozyatağı	0.990	0.964	0.010	Durağan Değil
Köşk	0.972	0.945	0.010	Durağan Değil
Kuruçeşme	0.990	0.990	0.010	Durağan Değil
Kuşcağız	0.022	0.840	0.027	Durağan Değil
Küçükcesat	0.917	0.775	0.015	Durağan Değil
Kükürtlü	0.990	0.990	0.010	Durağan Değil
Kültür	0.305	0.990	0.010	Durağan Değil
Levazım	0.567	0.812	0.023	Durağan Değil
Levent	0.990	0.990	0.010	Durağan Değil
Maksem	0.564	0.990	0.060	Durağan Değil
Maltepe	0.912	0.897	0.010	Durağan Değil
Mavişehir	0.928	0.990	0.010	Durağan Değil
Mecidiye	0.990	0.990	0.010	Durağan Değil
Merdivenköy	0.978	0.988	0.010	Durağan Değil
Merkez	0.990	0.792	0.010	Durağan Değil
Mithatpaşa	0.187	0.913	0.010	Durağan Değil
Muradiye	0.969	0.989	0.010	Durağan Değil
Mutlukent	0.014	0.729	0.100	Durağan Değil
Namık Kemal	0.868	0.956	0.010	Durağan Değil
Nergiz	0.935	0.976	0.010	Durağan Değil
Nispetiye	0.864	0.963	0.010	Durağan Değil
Oran	0.990	0.990	0.010	Durağan Değil
Ortaköy	0.900	0.958	0.010	Durağan Değil
Osmanağa	0.831	0.990	0.010	Durağan Değil
Osmangazi	0.508	0.633	0.100	Durağan Değil
Ovaakça Merkez	0.689	0.968	0.072	Durağan Değil
Örnekköy	0.990	0.990	0.010	Durağan Değil
Öveçler	0.990	0.990	0.010	Durağan Değil
Özlüce	0.591	0.990	0.010	Durağan Değil
Pınarbaşı	0.047	0.888	0.100	Durağan Değil
Prof. Dr. Ahmet Taner Kışlalı	0.928	0.990	0.010	Durağan Değil
Rasimpaşa	0.990	0.990	0.010	Durağan Değil
Sahrayı Cedit	0.915	0.958	0.010	Durağan Değil
Sancak	0.983	0.954	0.010	Durağan Değil
Sancaktepe	0.555	0.860	0.100	Durağan Değil
Seyranbağları	0.977	0.972	0.010	Durağan Değil
Sinanpaşa	0.990	0.990	0.010	Durağan Değil
Sokullu Mehmet Paşa	0.990	0.990	0.010	Durağan Değil
Suadiye	0.990	0.965	0.010	Durağan Değil
Şefkat	0.794	0.990	0.018	Durağan Değil
Şehit Kubilay	0.609	0.730	0.098	Durağan Değil
Şemikler	0.990	0.982	0.010	Durağan Değil
Şenlik	0.068	0.945	0.010	Durağan Değil
Tepebaşı	0.435	0.828	0.014	Durağan Değil

**Tablo B.5** Sıfırncı Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 5

<b>Mahalleler</b>	<b>ADF Test Olasılığı</b>	<b>PP Test Olasılığı</b>	<b>KPSS Test Olasılığı</b>	<b>Sonuçlar</b>
Türkali	0.952	0.984	0.010	Durağan Değil
Ufuktepe	0.127	0.685	0.058	Durağan Değil
Ulus	0.242	0.990	0.013	Durağan Değil
Uyanış	0.990	0.990	0.010	Durağan Değil
Üçevler	0.767	0.990	0.010	Durağan Değil
Ümit	0.488	0.657	0.096	Durağan Değil
Vişnezade	0.663	0.735	0.042	Durağan Değil
Yahyakaptan	0.990	0.990	0.010	Durağan Değil
Yakacık	0.034	0.958	0.100	Durağan Değil
Yalı	0.990	0.990	0.010	Durağan Değil
Yaşamkent	0.641	0.736	0.100	Durağan Değil
Yayla	0.684	0.915	0.046	Durağan Değil
Yenişehir	0.990	0.990	0.017	Durağan Değil
Yıldız	0.620	0.953	0.040	Durağan Değil
Yıldızevler	0.990	0.990	0.010	Durağan Değil
Yukarı Dikmen	0.794	0.971	0.010	Durağan Değil
Yunuseli	0.983	0.990	0.010	Durağan Değil
Yükseltepe	0.920	0.829	0.034	Durağan Değil
Yüzüncüyıl	0.916	0.990	0.010	Durağan Değil
Zafertepe	0.790	0.990	0.050	Durağan Değil
Zühtüpaşa	0.096	0.398	0.044	Durağan Değil



### B.3 Konut Satış Fiyat Endeksleri ADF, PP ve KPSS Birim Kök Test Sonuçları (Birinci Mertebeden)

Tablo B.6 Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 1

Mahalleler	ADF Test Olasılığı	PP Test Olasılığı	KPSS Test Olasılığı	Sonuçlar
19 May	0.021	0.013	0.100	Stationary
23 Nisan	0.464	0.010	0.100	Stationary
28 Haziran	0.056	0.010	0.100	Stationary
Abbasağa	0.435	0.010	0.100	Stationary
Acıbadem	0.486	0.045	0.525	Stationary
Adnan Menderes	0.498	0.010	0.100	Stationary
Ahlatlıbel	0.206	0.010	0.100	Stationary
Ahmetpaşa	0.717	0.010	0.054	Stationary
Akarca	0.379	0.010	0.100	Stationary
Akat	0.550	0.010	0.100	Stationary
Akın Simav	0.640	0.010	0.100	Stationary
Akpınar	0.052	0.010	0.100	Stationary
Aksoy	0.350	0.010	0.100	Stationary
Aktepe	0.196	0.010	0.100	Stationary
Alaybey	0.432	0.010	0.052	Stationary
Aliahya Merkez	0.362	0.010	0.100	Stationary
Alsancak	0.447	0.010	0.100	Stationary
Altınşehir	0.270	0.010	0.100	Stationary
Altıntaş	0.644	0.010	0.056	Stationary
Altıparmak	0.583	0.010	0.058	Stationary
Anittepe	0.198	0.010	0.100	Stationary
Arnavutköy	0.966	0.005	0.060	Stationary
Aşağı Dikmen	0.412	0.010	0.059	Stationary
Aşağı Eğlence	0.406	0.010	0.100	Stationary
Aşağı Öveçler	0.303	0.010	0.100	Stationary
Ata	0.654	0.010	0.100	Stationary
Ataevler	0.604	0.015	0.084	Stationary
Atakent	0.601	0.010	0.073	Stationary
Atapark	0.094	0.010	0.100	Stationary
Atilla	0.682	0.010	0.097	Stationary
Aydınlı	0.010	0.010	0.100	Stationary
Ayrancı	0.583	0.010	0.100	Stationary
Ayvalı	0.562	0.010	0.100	Stationary
Aziziye	0.451	0.010	0.100	Stationary
Bademlik	0.060	0.010	0.100	Stationary
Bağlarbaşı	0.283	0.010	0.100	Stationary
Bahariye	0.628	0.010	0.059	Stationary
Bahçelievler	0.184	0.010	0.066	Stationary
Bahriye Üçok	0.219	0.010	0.100	Stationary
Balat	0.697	0.022	0.060	Stationary

**Tablo B.7 Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 2**

<b>Mahalleler</b>	<b>ADF Test Olasılığı</b>	<b>PP Test Olasılığı</b>	<b>KPSS Test Olasılığı</b>	<b>Sonuçlar</b>
Balgat	0.948	0.010	0.100	Stationary
Balmumcu	0.799	0.010	0.050	Stationary
Barbaros	0.529	0.010	0.100	Stationary
Barış	0.673	0.010	0.053	Stationary
Basınevleri	0.027	0.010	0.100	Stationary
Bebek	0.528	0.050	0.057	Stationary
Beşevler	0.330	0.010	0.096	Stationary
Beytepe	0.623	0.010	0.098	Stationary
Birlik	0.495	0.010	0.100	Stationary
Bostancı	0.172	0.027	0.100	Stationary
Bostanlı	0.500	0.010	0.100	Stationary
Büyükesat	0.021	0.010	0.100	Stationary
Caddebostan	0.014	0.146	0.100	Stationary
Caferağa	0.256	0.010	0.100	Stationary
Cebeci	0.493	0.010	0.100	Stationary
Cedit	0.567	0.010	0.062	Stationary
Cevizlidere	0.580	0.010	0.100	Stationary
Cihannüma	0.305	0.010	0.100	Stationary
Cumhuriyet	0.963	0.010	0.065	Stationary
Çaldıran	0.448	0.010	0.100	Stationary
Çalseki	0.058	0.010	0.100	Stationary
Çamlıca	0.380	0.010	0.100	Stationary
Çamlıtepe	0.721	0.010	0.056	Stationary
Çankaya	0.529	0.010	0.060	Stationary
Çayyolu	0.545	0.010	0.060	Stationary
Çekirge	0.879	0.010	0.054	Stationary
Çirpan	0.955	0.010	0.055	Stationary
Çiçekli	0.276	0.010	0.100	Stationary
Çiğdem	0.918	0.010	0.073	Stationary
Çukurambar	0.123	0.010	0.100	Stationary
Dedebaşı	0.533	0.010	0.100	Stationary
Değirmendere Merkez	0.543	0.010	0.100	Stationary
Demirci	0.927	0.010	0.050	Stationary
Demirköprü	0.230	0.016	0.100	Stationary
Demirtaş Cumhuriyet	0.220	0.010	0.100	Stationary
Demirtaşpaşa	0.272	0.010	0.100	Stationary
Dikilitaş	0.139	0.010	0.100	Stationary
Dikkaldırım	0.473	0.010	0.100	Stationary
Doburca	0.724	0.010	0.053	Stationary
Donanmacı	0.634	0.010	0.056	Stationary
Dumlupınar	0.849	0.036	0.056	Stationary
Eğitim	0.629	0.024	0.050	Stationary
Emek	0.377	0.010	0.100	Stationary
Emrah	0.413	0.010	0.100	Stationary
Erenköy	0.013	0.045	0.100	Stationary

**Tablo B.8 Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 3**

<b>Mahalleler</b>	<b>ADF Test Olasılığı</b>	<b>PP Test Olasılığı</b>	<b>KPSS Test Olasılığı</b>	<b>Sonuçlar</b>
Ertuğrul	0.513	0.010	0.100	Stationary
Esertepe	0.087	0.010	0.100	Stationary
Etiler	0.833	0.025	0.059	Stationary
Etlük	0.287	0.010	0.100	Stationary
Fenerbahçe	0.300	0.010	0.100	Stationary
Feneryolu	0.441	0.012	0.100	Stationary
Fethiye	0.490	0.010	0.054	Stationary
Fikirtepe	0.913	0.043	0.056	Stationary
Fikri Altay	0.355	0.010	0.100	Stationary
Gayrettepe	0.021	0.010	0.100	Stationary
Gaziosmanpaşa	0.442	0.010	0.100	Stationary
Goncalar	0.738	0.010	0.052	Stationary
Görükle	0.986	0.044	0.050	Stationary
Göztepe	0.132	0.041	0.100	Stationary
Güçlükaya	0.673	0.010	0.099	Stationary
Gümüşdere	0.047	0.010	0.100	Stationary
Güvenevler	0.235	0.010	0.100	Stationary
Güzelyalı	0.461	0.010	0.100	Stationary
Harbiye	0.573	0.010	0.100	Stationary
Hasanpaşa	0.144	0.019	0.100	Stationary
Hilal	0.966	0.010	0.060	Stationary
Hocahasan	0.475	0.010	0.056	Stationary
Huzur	0.167	0.010	0.100	Stationary
Hüdavendigâr	0.773	0.010	0.060	Stationary
İbrahimpaşa	0.981	0.010	0.060	Stationary
İhsaniye	0.417	0.010	0.060	Stationary
İlkadım	0.017	0.010	0.100	Stationary
İlkbahar	0.622	0.010	0.100	Stationary
İlker	0.207	0.010	0.100	Stationary
İncirli	0.143	0.010	0.100	Stationary
İşçi Blokları	0.501	0.010	0.100	Stationary
Kadıköy	0.349	0.010	0.100	Stationary
Kalaba	0.144	0.010	0.100	Stationary
Kamil Ocak	0.044	0.010	0.100	Stationary
Kanuni	0.509	0.010	0.057	Stationary
Karapınar	0.227	0.010	0.100	Stationary
Kavacık Subayevleri	0.454	0.010	0.100	Stationary
Kavaklıdere	0.256	0.010	0.100	Stationary
Keklik Pınarı	0.447	0.010	0.100	Stationary
Kılıç Reis	0.842	0.010	0.056	Stationary
Kırkkonaklar	0.280	0.010	0.100	Stationary
Kızılay	0.071	0.010	0.100	Stationary
Kirazlı	0.908	0.041	0.055	Stationary
Konak	0.764	0.043	0.060	Stationary
Konaklar	0.919	0.010	0.100	Stationary

**Tablo B.9 Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 4**

<b>Mahalleler</b>	<b>ADF Test Olasılığı</b>	<b>PP Test Olasılığı</b>	<b>KPSS Test Olasılığı</b>	<b>Sonuçlar</b>
Koşuyolu	0.395	0.010	0.100	Stationary
Kozyatağı	0.020	0.020	0.100	Stationary
Köşk	0.468	0.010	0.100	Stationary
Kuruçeşme	0.721	0.010	0.068	Stationary
Kuşcağız	0.521	0.010	0.100	Stationary
Küçükesat	0.062	0.010	0.100	Stationary
Kükürtlü	0.267	0.010	0.100	Stationary
Kültür	0.548	0.010	0.056	Stationary
Levazım	0.038	0.249	0.055	Stationary
Levent	0.613	0.010	0.085	Stationary
Maksem	0.990	0.010	0.050	Stationary
Maltepe	0.381	0.010	0.100	Stationary
Mavişehir	0.555	0.010	0.100	Stationary
Mecidiye	0.359	0.010	0.100	Stationary
Merdivenköy	0.212	0.010	0.100	Stationary
Merkez	0.043	0.010	0.100	Stationary
Mithatpaşa	0.435	0.010	0.100	Stationary
Muradiye	0.034	0.010	0.100	Stationary
Mutlukent	0.033	0.010	0.100	Stationary
Namık Kemal	0.129	0.010	0.100	Stationary
Nergiz	0.428	0.010	0.100	Stationary
Nispetiye	0.606	0.010	0.100	Stationary
Oran	0.608	0.010	0.063	Stationary
Ortaköy	0.010	0.010	0.100	Stationary
Osmanağa	0.225	0.010	0.062	Stationary
Osmangazi	0.220	0.010	0.100	Stationary
Ovaakça Merkez	0.537	0.010	0.074	Stationary
Örnekköy	0.368	0.010	0.100	Stationary
Öveçler	0.116	0.010	0.100	Stationary
Özlüce	0.658	0.010	0.054	Stationary
Pınarbaşı	0.239	0.010	0.100	Stationary
Prof. Dr. Ahmet Taner Kışlalı	0.592	0.023	0.051	Stationary
Rasimpaşa	0.159	0.010	0.100	Stationary
Sahrayı Cedit	0.359	0.042	0.100	Stationary
Sancak	0.254	0.010	0.100	Stationary
Sancaktepe	0.737	0.010	0.100	Stationary
Seyranbağları	0.414	0.010	0.100	Stationary
Sinanpaşa	0.946	0.010	0.050	Stationary
Sokullu Mehmet Paşa	0.306	0.010	0.100	Stationary
Suadiye	0.042	0.064	0.100	Stationary
Şefkat	0.550	0.010	0.080	Stationary
Şehit Kubilay	0.554	0.010	0.100	Stationary
Şemikler	0.282	0.010	0.100	Stationary
Şenlik	0.493	0.010	0.100	Stationary
Tepebaşı	0.362	0.010	0.100	Stationary

**Tablo B.10** Birinci Mertebeden ADF, PP ve KPSS Birim Kök Sonuçları 5

<b>Mahalleler</b>	<b>ADF Test Olasılığı</b>	<b>PP Test Olasılığı</b>	<b>KPSS Test Olasılığı</b>	<b>Sonuçlar</b>
Türkali	0.407	0.010	0.100	Stationary
Ufuktepe	0.518	0.010	0.055	Stationary
Ulus	0.586	0.040	0.060	Stationary
Uyanış	0.644	0.010	0.092	Stationary
Üçevler	0.827	0.010	0.056	Stationary
Ümit	0.371	0.010	0.100	Stationary
Vişnezade	0.949	0.024	0.050	Stationary
Yahyakaptan	0.387	0.010	0.100	Stationary
Yakacık	0.404	0.010	0.100	Stationary
Yalı	0.548	0.010	0.100	Stationary
Yaşamkent	0.372	0.012	0.100	Stationary
Yayla	0.353	0.010	0.100	Stationary
Yenişehir	0.903	0.010	0.059	Stationary
Yıldız	0.242	0.010	0.100	Stationary
Yıldizevler	0.128	0.010	0.064	Stationary
Yukarı Dikmen	0.408	0.010	0.100	Stationary
Yunuseli	0.483	0.010	0.100	Stationary
Yükseltepe	0.134	0.010	0.093	Stationary
Yüzüncüyl	0.837	0.015	0.050	Stationary
Zafertepe	0.953	0.010	0.080	Stationary
Zühtüpaşa	0.628	0.010	0.100	Stationary

### C.1 AR Model Parametre Belirleme R Kodları

```

library(readxl)
library(forecast)

Sale_Indices <- read_excel("C:/Users/Bertil/Desktop/Sale_Indices.xlsx")
View(Sale_Indices)
Sale_Indices <- ts(Sale_Indices, start = c(2010,01), frequency = 12)
Sale_Indices_D1 <- diff(Sale_Indices, differences = 1)

# Defining variables
l <- length(Sale_Indices_D1[1,])
namedistrict <- matrix(, nrow = l, ncol = 1)
coef <- matrix(, nrow = l, ncol = 10)
aic <- matrix(, nrow = l, ncol = 1)
bic <- matrix(, nrow = l, ncol = 1)
aicc <- matrix(, nrow = l, ncol = 1)
boxcox_lambda <- matrix(, nrow = l, ncol = 1)
pierce_box <- matrix(, nrow = l, ncol = 1)
ljung_box <- matrix(, nrow = l, ncol = 1)
forecast <- matrix(, nrow = 24, ncol = 1)

for(i in 1:l)
{
  lambda = BoxCox.lambda(Sale_Indices_D1[i])
  fit <- auto.arima(Sale_Indices_D1[i], ic = "bic", lambda = lambda, allowmean =
FALSE, d = 0, max.p = 10, max.q = FALSE, max.P = FALSE, max.Q = FALSE, max.D
= 0)
  namedistrict[i,1] <- names(Sale_Indices_D1[1,i])
  print(fit$coef)
  coef[i,1] <- fit$coef[1]
  coef[i,2] <- fit$coef[2]
  coef[i,3] <- fit$coef[3]
  coef[i,4] <- fit$coef[4]
  coef[i,5] <- fit$coef[5]
  coef[i,6] <- fit$coef[6]
  coef[i,7] <- fit$coef[7]
  coef[i,8] <- fit$coef[8]
  coef[i,9] <- fit$coef[9]
  coef[i,10] <- fit$coef[10]
}

```

```

fcast <- forecast(fit, n.ahead = 12)
forecast[,i] <- fcast$mean

aic[i,1] <- fit$aic
bic[i,1] <- fit$bic
aicc[i,1] <- fit$aicc

#lb <- checkresiduals(fit, df = NULL, plot = FALSE)
bt_pb <- Box.test(fit$residuals, lag = 1, type = "Box-Pierce")
bt_lb <- Box.test(fit$residuals, lag = 1, type = "Ljung-Box")

pierce_box[i,1] <- bt_pb$p.value
ljung_box[i,1] <- bt_lb$p.value
boxcox_lambda[i,1] <- lambda

rm(fit)
rm(lambda)
rm(fcast)
rm(bt_pb)
rm(bt_lb)
}

colnames(coef) <- c("AR1", "AR2", "AR3", "AR4", "AR5", "AR6", "AR7", "AR8",
"AR9", "AR10")
arima_results <- data.frame(namedistrict, coef, aic, bic, aicc, boxcox_lambda,
pierce_box, ljung_box)
colnames(forecast) <- namedistrict
forecast <- ts(forecast, start = c(2017,01), frequency = 12)
write.csv(arima_results, file = "arima_results.csv")
write.csv(forecast, file = "forecast.csv")
rm(l)
rm(i)
rm(namedistrict)
rm(forecast)
rm(coef)
rm(arima_results)
rm(aic)
rm(bic)
rm(aicc)
rm(ljung_box)
rm(pierce_box)
rm(boxcox_lambda)

```

## AR TAHMİNLENEN MODEL PARAMETRELERİ

Tablo D.1 AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 1

Mahalleler	AR1	AR2	AR3	AR4	AR5	AR6	AR7	Pierce Box	Ljung Box
19 Mayıs	1.65	-0.40	-0.25	0.00	0.00	0.00	0.00	0.63	0.62
23 Nisan	1.34	-0.34	0.00	0.00	0.00	0.00	0.00	0.06	0.05
28 Haziran	1.80	-0.80	0.00	0.00	0.00	0.00	0.00	0.16	0.15
Abbasaga	1.78	-0.78	0.00	0.00	0.00	0.00	0.00	0.24	0.24
Acibadem	1.88	-0.88	0.00	0.00	0.00	0.00	0.00	0.35	0.34
Adnan Menderes	1.53	-0.53	0.00	0.00	0.00	0.00	0.00	0.46	0.45
Ahlatlibel	2.10	-1.46	0.24	0.41	-0.29	0.00	0.00	0.91	0.91
Ahmetpasa	1.56	-0.37	-0.19	0.00	0.00	0.00	0.00	0.59	0.58
Akarca	1.31	-0.07	-0.08	-0.17	0.00	0.00	0.00	0.81	0.80
Akat	1.98	-0.98	0.00	0.00	0.00	0.00	0.00	0.04	0.03
Akin Simav	1.65	-0.65	0.00	0.00	0.00	0.00	0.00	0.80	0.80
Akpınar	2.20	-1.92	1.09	-0.37	0.00	0.00	0.00	0.51	0.50
Aksoy	1.57	-0.69	0.57	-0.45	0.00	0.00	0.00	0.91	0.91
Aktepe	1.61	-0.61	0.00	0.00	0.00	0.00	0.00	0.47	0.46
Alaybey	1.84	-0.84	0.00	0.00	0.00	0.00	0.00	0.18	0.17
Alikahya Merkez	1.38	-0.38	0.00	0.00	0.00	0.00	0.00	0.71	0.71
Alsancak	1.48	-0.48	0.00	0.00	0.00	0.00	0.00	0.88	0.87
Altınşehir	1.88	-0.88	0.00	0.00	0.00	0.00	0.00	0.98	0.98
Altıntaş	1.66	-0.63	0.15	-0.18	0.00	0.00	0.00	0.83	0.83
Altıparmak	1.56	-0.56	0.00	0.00	0.00	0.00	0.00	0.53	0.52
Anittepe	1.74	-0.74	0.00	0.00	0.00	0.00	0.00	0.85	0.85
Arnavutköy	1.98	-0.98	0.00	0.00	0.00	0.00	0.00	0.30	0.29
Asağı Dikmen	1.78	-0.78	0.00	0.00	0.00	0.00	0.00	0.75	0.74
Asağı Eglence	1.60	-0.60	0.00	0.00	0.00	0.00	0.00	0.86	0.86
Asağı Öveçler	1.83	-0.83	0.00	0.00	0.00	0.00	0.00	0.53	0.52
Ata	1.80	-0.80	0.00	0.00	0.00	0.00	0.00	0.98	0.97
Ataevler	1.84	-0.84	0.00	0.00	0.00	0.00	0.00	0.78	0.78
Atakent	1.31	-0.38	0.40	-0.18	0.07	-0.22	0.00	0.86	0.85
Atapark	1.77	-0.77	0.00	0.00	0.00	0.00	0.00	0.81	0.81
Atilla	1.25	-0.02	0.14	-0.37	0.00	0.00	0.00	0.63	0.63
Aydınlar	1.80	-0.80	0.00	0.00	0.00	0.00	0.00	0.28	0.28
Ayrancı	1.64	-0.64	0.00	0.00	0.00	0.00	0.00	0.73	0.73
Ayvalı	1.59	-0.59	0.00	0.00	0.00	0.00	0.00	0.16	0.15
Aziziye	1.78	-0.78	0.00	0.00	0.00	0.00	0.00	0.87	0.87
Bademlik	1.65	-0.65	0.00	0.00	0.00	0.00	0.00	0.16	0.15
Bağlarbaşı	1.68	-0.68	0.00	0.00	0.00	0.00	0.00	0.34	0.33
Bahariye	1.43	-0.20	0.02	-0.25	0.00	0.00	0.00	0.65	0.65
Bahçelievler	1.82	-0.82	0.00	0.00	0.00	0.00	0.00	0.27	0.26
Bahriye Üçok	1.85	-0.85	0.00	0.00	0.00	0.00	0.00	0.74	0.74
Balat	1.65	-0.39	-0.26	0.00	0.00	0.00	0.00	0.62	0.62



**Tablo D.2 AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 2**

Mahalleler	AR1	AR2	AR3	AR4	AR5	AR6	AR7	Pierce Box	Ljung Box
Balgat	1.54	-0.54	0.00	0.00	0.00	0.00	0.00	0.87	0.87
Balmumcu	1.26	0.38	-0.64	0.00	0.00	0.00	0.00	0.11	0.11
Barbaros	1.78	-0.78	0.00	0.00	0.00	0.00	0.00	0.96	0.96
Baris	1.76	-0.76	0.00	0.00	0.00	0.00	0.00	0.26	0.25
Basinevleri	1.72	-0.72	0.00	0.00	0.00	0.00	0.00	0.82	0.81
Bebek	2.19	-1.41	0.22	0.00	0.00	0.00	0.00	0.98	0.98
Besevler	1.77	-0.80	0.33	-0.66	0.62	-0.25	0.00	0.85	0.85
Beytepe	1.79	-0.79	0.00	0.00	0.00	0.00	0.00	0.69	0.68
Birlik	1.72	-0.72	0.00	0.00	0.00	0.00	0.00	0.84	0.84
Bostanci	1.88	-0.88	0.00	0.00	0.00	0.00	0.00	0.78	0.78
Bostanlı	1.84	-0.84	0.00	0.00	0.00	0.00	0.00	0.39	0.38
Büyükesat	1.65	-0.65	0.00	0.00	0.00	0.00	0.00	0.61	0.60
Caddebostan	2.28	-1.98	1.21	-0.75	0.23	0.00	0.00	0.96	0.96
Caferaga	1.83	-0.83	0.00	0.00	0.00	0.00	0.00	0.43	0.42
Cebeci	1.56	-0.56	0.00	0.00	0.00	0.00	0.00	0.75	0.74
Cedit	1.72	-0.72	0.00	0.00	0.00	0.00	0.00	0.18	0.17
Cevizlidere	1.76	-0.76	0.00	0.00	0.00	0.00	0.00	0.98	0.98
Cihannüma	1.28	-0.23	0.00	0.19	-0.03	-0.22	0.00	0.81	0.80
Cumhuriyet	1.41	-0.23	0.15	-0.33	0.00	0.00	0.00	0.61	0.61
Çaldıran	1.52	-0.34	-0.56	0.82	-0.44	0.00	0.00	0.48	0.48
Çalseki	1.53	-0.32	-0.20	0.00	0.00	0.00	0.00	0.58	0.57
Çamlıca	1.76	-0.76	0.00	0.00	0.00	0.00	0.00	0.50	0.50
Çamlıtepe	1.71	-0.71	0.00	0.00	0.00	0.00	0.00	0.52	0.51
Çankaya	1.76	-0.76	0.00	0.00	0.00	0.00	0.00	0.70	0.69
Çayyolu	1.65	-0.65	0.00	0.00	0.00	0.00	0.00	0.68	0.67
Çekirge	1.67	-0.67	0.00	0.00	0.00	0.00	0.00	0.59	0.58
Çirpan	1.59	-0.33	-0.26	0.00	0.00	0.00	0.00	0.92	0.92
Çiçekli	1.53	-0.36	-0.18	0.00	0.00	0.00	0.00	0.91	0.91
Çiğdem	1.56	-0.56	0.00	0.00	0.00	0.00	0.00	0.18	0.17
Çukurambar	2.10	-1.38	-0.05	0.77	-0.44	0.00	0.00	0.83	0.83
Dedebasi	1.83	-0.83	0.00	0.00	0.00	0.00	0.00	0.10	0.09
Degirmendere Merkez	1.70	-0.89	0.19	0.00	0.00	0.00	0.00	0.83	0.82
Demirci	1.38	-0.01	-0.19	-0.19	0.00	0.00	0.00	0.79	0.79
Demirköprü	1.90	-0.90	0.00	0.00	0.00	0.00	0.00	0.46	0.45
Demirtas Cumhuriyet	2.00	-1.25	-0.08	0.84	-0.51	0.00	0.00	0.69	0.68
Demirtaspasa	1.07	0.38	-0.19	-0.26	0.00	0.00	0.00	0.73	0.73
Dikilitas	2.01	-1.46	0.69	-0.25	0.00	0.00	0.00	0.80	0.80
Dikkaldırım	1.71	-0.71	0.00	0.00	0.00	0.00	0.00	0.94	0.93
Doburca	1.41	0.04	-0.45	0.00	0.00	0.00	0.00	0.41	0.40
Donanmaci	1.93	-1.32	0.66	-0.27	0.00	0.00	0.00	0.77	0.76

**Tablo D.3 AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 3**

Mahalleler	AR1	AR2	AR3	AR4	AR5	AR6	AR7	Pierce Box	Ljung Box
Dumlupınar	1.77	-0.57	-0.19	0.00	0.00	0.00	0.00	0.35	0.34
Eğitim	1.94	-0.94	0.00	0.00	0.00	0.00	0.00	1.00	1.00
Emek	1.73	-0.73	0.00	0.00	0.00	0.00	0.00	0.93	0.93
Emrah	1.39	-0.25	0.11	-0.25	0.00	0.00	0.00	0.81	0.81
Erenköy	1.92	-0.92	0.00	0.00	0.00	0.00	0.00	0.41	0.40
Ertugrul	1.79	-0.79	0.00	0.00	0.00	0.00	0.00	0.46	0.45
Esertepe	1.96	-1.41	0.71	-0.26	0.00	0.00	0.00	0.57	0.56
Etiler	1.93	-0.93	0.00	0.00	0.00	0.00	0.00	0.97	0.97
Etlük	1.76	-0.76	0.00	0.00	0.00	0.00	0.00	0.19	0.18
Fenerbahçe	1.92	-0.92	0.00	0.00	0.00	0.00	0.00	0.65	0.65
Feneryolu	2.31	-2.09	1.16	-0.39	0.00	0.00	0.00	0.84	0.84
Fethiye	1.84	-0.84	0.00	0.00	0.00	0.00	0.00	0.78	0.77
Fikirtepe	1.93	-0.93	0.00	0.00	0.00	0.00	0.00	0.52	0.51
Fikri Altay	1.45	-0.05	-0.40	0.00	0.00	0.00	0.00	0.42	0.41
Gayrettepe	1.81	-0.81	0.00	0.00	0.00	0.00	0.00	0.76	0.76
Gaziosmanpaşa	1.63	-0.63	0.00	0.00	0.00	0.00	0.00	0.63	0.62
Goncalar	1.27	-0.08	-0.04	-0.01	0.08	-0.21	0.00	0.45	0.44
Görükle	2.06	-1.77	1.30	-0.59	0.00	0.00	0.00	0.34	0.33
Göztepe	1.87	-0.87	0.00	0.00	0.00	0.00	0.00	0.77	0.77
Güçlükaya	1.57	-0.39	-0.18	0.00	0.00	0.00	0.00	0.67	0.66
Gümüşdere	1.25	-0.06	0.01	-0.20	0.00	0.00	0.00	0.68	0.67
Güvenevler	1.12	0.29	-0.41	0.00	0.00	0.00	0.00	0.65	0.65
Güzelyalı	1.78	-0.78	0.00	0.00	0.00	0.00	0.00	0.49	0.48
Harbiye	1.79	-0.79	0.00	0.00	0.00	0.00	0.00	0.77	0.77
Hasanpaşa	1.86	-0.86	0.00	0.00	0.00	0.00	0.00	0.41	0.40
Hilal	1.83	-1.25	0.51	0.50	-0.94	0.35	0.00	0.54	0.53
Hocahasan	1.25	0.17	-0.42	0.00	0.00	0.00	0.00	0.70	0.69
Huzur	1.57	-0.40	-0.17	0.00	0.00	0.00	0.00	0.55	0.55
Hüdevandigar	1.56	-0.48	0.14	-0.22	0.00	0.00	0.00	0.70	0.70
İbrahimpaşa	1.70	-0.70	0.00	0.00	0.00	0.00	0.00	0.82	0.81
İhsaniye	1.86	-0.86	0.00	0.00	0.00	0.00	0.00	0.32	0.31
İlkadim	1.71	-0.71	0.00	0.00	0.00	0.00	0.00	0.94	0.94
İlkbahar	2.09	-1.50	0.06	1.12	-1.02	0.24	0.00	0.91	0.91
İlker	1.74	-1.06	0.52	-0.19	0.00	0.00	0.00	0.74	0.74
İncirli	1.69	-0.69	0.00	0.00	0.00	0.00	0.00	0.82	0.81
İşçi Blokları	1.32	0.18	-0.50	0.00	0.00	0.00	0.00	0.31	0.31
Kadıköy	1.39	-0.39	0.00	0.00	0.00	0.00	0.00	0.22	0.21
Kalaba	1.72	-0.72	0.00	0.00	0.00	0.00	0.00	0.88	0.88
Kamil Ocak	1.32	-0.13	-0.19	0.00	0.00	0.00	0.00	0.31	0.30
Kanuni	1.61	-0.60	0.20	-0.22	0.00	0.00	0.00	0.99	0.99

**Tablo D.4 AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 4**

Mahalleler	AR1	AR2	AR3	AR4	AR5	AR6	AR7	Pierce Box	Ljung Box
Karapınar	1.45	-0.45	0.00	0.00	0.00	0.00	0.00	0.18	0.17
Kavacık Subayevleri	1.49	-0.49	0.00	0.00	0.00	0.00	0.00	0.42	0.42
Kavaklıdere	1.61	-0.61	0.00	0.00	0.00	0.00	0.00	0.41	0.40
Keklik Pinari	1.77	-0.77	0.00	0.00	0.00	0.00	0.00	0.81	0.80
Kiliç Reis	1.39	-0.31	-0.07	0.38	-0.39	0.00	0.00	0.96	0.96
Kirkkonaklar	1.62	-0.62	0.00	0.00	0.00	0.00	0.00	0.96	0.96
Kızılay	1.81	-1.19	0.58	-0.19	0.00	0.00	0.00	0.59	0.59
Kirazlı	2.10	-1.27	0.17	0.00	0.00	0.00	0.00	0.98	0.98
Konak	1.77	-0.58	-0.19	0.00	0.00	0.00	0.00	0.78	0.78
Konaklar	1.33	-0.04	-0.30	0.00	0.00	0.00	0.00	0.31	0.30
Kosuyolu	1.58	-0.55	0.23	-0.27	0.00	0.00	0.00	0.79	0.78
Kozyatagi	1.88	-0.88	0.00	0.00	0.00	0.00	0.00	0.20	0.19
Kösk	1.63	-0.63	0.00	0.00	0.00	0.00	0.00	0.77	0.76
Kuruçesme	1.97	-0.97	0.00	0.00	0.00	0.00	0.00	0.61	0.60
Kuscagiz	1.76	-0.76	0.00	0.00	0.00	0.00	0.00	0.59	0.59
Küçükesat	1.53	-0.53	0.00	0.00	0.00	0.00	0.00	0.74	0.73
Kükürtlü	1.80	-1.14	0.58	-0.24	0.00	0.00	0.00	0.62	0.61
Kültür	1.60	-0.33	-0.28	0.00	0.00	0.00	0.00	0.79	0.79
Levazım	1.94	-0.94	0.00	0.00	0.00	0.00	0.00	0.27	0.26
Levent	1.94	-0.94	0.00	0.00	0.00	0.00	0.00	0.88	0.88
Maksem	1.57	-0.57	0.00	0.00	0.00	0.00	0.00	0.18	0.18
Maltepe	1.52	-0.52	0.00	0.00	0.00	0.00	0.00	0.82	0.82
Mavisehir	1.80	-0.80	0.00	0.00	0.00	0.00	0.00	0.20	0.20
Mecidiye	1.53	-0.43	0.10	-0.21	0.00	0.00	0.00	0.76	0.75
Merdivenköy	1.84	-0.84	0.00	0.00	0.00	0.00	0.00	0.63	0.62
Merkez	1.82	-1.51	1.41	-1.30	1.05	-0.74	0.26	0.65	0.64
Mithatpasa	1.80	-1.36	1.01	-0.44	0.00	0.00	0.00	0.67	0.66
Muradiye	1.70	-0.70	0.00	0.00	0.00	0.00	0.00	0.34	0.33
Mutlukent	1.69	-0.69	0.00	0.00	0.00	0.00	0.00	0.88	0.88
Namik Kemal	1.48	-0.48	0.00	0.00	0.00	0.00	0.00	0.65	0.64
Nergiz	1.92	-0.92	0.00	0.00	0.00	0.00	0.00	0.82	0.81
Nispetiye	1.47	0.16	-0.97	0.52	-0.18	0.00	0.00	0.71	0.71
Oran	2.14	-1.82	1.07	-0.39	0.00	0.00	0.00	0.77	0.77
Ortaköy	1.75	-0.75	0.00	0.00	0.00	0.00	0.00	0.77	0.76
Osmanaga	1.59	-0.26	-0.33	0.00	0.00	0.00	0.00	0.55	0.54
Osmanгази	1.61	-0.42	-0.19	0.00	0.00	0.00	0.00	0.62	0.62
Ovaakça Merkez	1.39	-0.39	0.00	0.00	0.00	0.00	0.00	0.21	0.20
Örnekköy	1.85	-0.85	0.00	0.00	0.00	0.00	0.00	0.39	0.38
Öveçler	1.67	-0.67	0.00	0.00	0.00	0.00	0.00	0.95	0.95
Özlüce	1.88	-0.88	0.00	0.00	0.00	0.00	0.00	0.74	0.73

**Tablo D.5 AR Model Parametreleri ve Pierce ve Ljung Box Test Olasılıkları 5**

Mahalleler	AR1	AR2	AR3	AR4	AR5	AR6	AR7	Pierce Box	Ljung Box
Pınarbasi	1.34	-0.29	0.36	-0.41	0.00	0.00	0.00	0.88	0.88
Prof. Dr. Ahmet Taner Kislali	2.16	-1.59	0.40	0.22	-0.20	0.00	0.00	0.97	0.97
Rasimpasa	1.89	-0.89	0.00	0.00	0.00	0.00	0.00	0.20	0.19
Sahrayi Cedit	1.89	-0.89	0.00	0.00	0.00	0.00	0.00	0.80	0.80
Sancak	1.49	-0.49	0.00	0.00	0.00	0.00	0.00	0.27	0.26
Sancaktepe	1.55	-0.37	-0.18	0.00	0.00	0.00	0.00	0.39	0.38
Seyranbaglari	1.69	-0.69	0.00	0.00	0.00	0.00	0.00	0.64	0.64
Sinanpasa	1.36	-0.02	-0.10	-0.24	0.00	0.00	0.00	0.56	0.55
Sokullu Mehmet Pasa	1.78	-0.78	0.00	0.00	0.00	0.00	0.00	0.81	0.81
Suadiye	1.90	-0.90	0.00	0.00	0.00	0.00	0.00	0.43	0.42
Sefkat	1.63	-0.63	0.00	0.00	0.00	0.00	0.00	0.14	0.14
Sehit Kubilay	1.54	-0.54	0.00	0.00	0.00	0.00	0.00	0.07	0.06
Semikler	1.91	-0.91	0.00	0.00	0.00	0.00	0.00	0.39	0.38
Senlik	1.42	-0.42	0.00	0.00	0.00	0.00	0.00	0.22	0.21
Tepebasi	1.70	-1.37	1.28	-1.07	0.86	-0.40	0.00	0.94	0.94
Türkali	1.85	-0.85	0.00	0.00	0.00	0.00	0.00	0.66	0.66
Ufuktepe	1.67	-0.67	0.00	0.00	0.00	0.00	0.00	0.25	0.24
Ulus	1.71	-0.44	-0.27	0.00	0.00	0.00	0.00	0.91	0.91
Uyanis	1.74	-0.74	0.00	0.00	0.00	0.00	0.00	0.97	0.97
Üçevler	1.82	-0.82	0.00	0.00	0.00	0.00	0.00	0.66	0.65
Ümit	1.76	-1.22	0.82	-0.36	0.00	0.00	0.00	0.71	0.70
Visnezade	2.42	-2.63	1.91	-0.69	0.00	0.00	0.00	0.06	0.06
Yahyakaptan	1.61	-0.70	0.57	-0.96	0.56	0.01	0.14	0.98	0.98
Yakacik	1.18	0.37	-0.55	0.00	0.00	0.00	0.00	0.82	0.82
Yali	1.74	-0.74	0.00	0.00	0.00	0.00	0.00	0.14	0.13
Yasamkent	1.82	-0.82	0.00	0.00	0.00	0.00	0.00	0.65	0.64
Yayla	1.70	-0.70	0.00	0.00	0.00	0.00	0.00	0.60	0.60
Yenisehir	1.61	-0.61	0.00	0.00	0.00	0.00	0.00	0.69	0.68
Yildiz	1.54	-0.54	0.00	0.00	0.00	0.00	0.00	0.24	0.23
Yildizevler	1.60	-0.60	0.00	0.00	0.00	0.00	0.00	0.96	0.96
Yukari Dikmen	1.17	-0.30	1.02	-0.89	0.00	0.00	0.00	0.86	0.86
Yunuseli	2.21	-1.83	1.00	-0.38	0.00	0.00	0.00	0.44	0.43
Yükseltepe	1.47	-0.47	0.00	0.00	0.00	0.00	0.00	0.43	0.42
Yüzüncüyil	1.61	-0.26	-0.34	0.00	0.00	0.00	0.00	0.77	0.77
Zafertepe	1.47	-0.09	-0.38	0.00	0.00	0.00	0.00	0.44	0.43
Zühtüpaşa	1.47	0.21	-1.16	0.47	0.00	0.00	0.00	0.48	0.47

### E.1 Hiyerarşik Kümeleme ve VCV Kodları

```

% pdist() uzaklık/benzerlik parametreleri:
% euclidean
% mahalanobis
% cityblock
% minkowski
% chebychev
% cosine
% correlation
% single hierchical clustering
% linkage() hiyerarşik kümeleme yöntemleri:
% single
% average
% centroid
% complete
% median
% ward
% weighted
d = pdist(sale_indices,'euclidean');
D = squareform(d);
figure(1)
for i = 2:10
    z = linkage(D,'single');
    T = cluster(z,i);
    results(1:length(sale_indices), i-1) = T;
    [Ts,inds]=sort(T);
    subplot(3,3,i-1)
    imagesc(D(inds,inds))
end
figure(2)
dendrogram(z,0,'colorthreshold','default');
title('Single Hierchical Clustering')
for i = 1:9
[Y,eigvals] = cmdscale(D);
figure(i+2)
gscatter(Y(:,1),Y(:,2),results(:,i))
title('Multidimensional Scaling')
xlabel('1st Principal Component')
ylabel('2nd Principal Component') end

```

## E.2 K-Ortalamlar Kümeleme ve VCV Kodları

```
d = pdist(sale_indices,'euclidean');
D = squareform(d);
figure(1)
for i = 2:10
    T = kmeans(D,i);
    [Ts,inds]=sort(T);
    results(1:length(sale_indices), i-1) = T;
    subplot(3,3,i-1)
    imagesc(D(inds,inds))
end

for i = 1:9
    [Y,eigvals] = cmdscale(D);
    figure(i+1)
    gscatter(Y(:,1),Y(:,2),results(:,i))
    title('Multidimensional Scaling')
    xlabel('1st Principal Component')
    ylabel('2nd Principal Component')
end
```

## E.3 SOM Kümeleme ve VCV Kodları

```
dimension1 = 10;
dimension2 = 10;
net = selforgmap([dimension1 dimension2]);
[net,tr] = train(net,sale_indices);
outputs = net(sale_indices);
view(net)
```

## E.4 Bulanık C-Ortalamlar Kümeleme ve VCV Kodları

```
d = pdist(sale_indices,'euclidean');
D = squareform(d);
figure(1)
for i = 2:10
    [centers,U] = fcm(D,i);
    [M,T]=max(U)
    results(1:length(sale_indices), i-1) = T';
    [Ts,inds]=sort(T);
    subplot(3,3,i-1)
    imagesc(D(inds,inds))
end

for i = 1:9
    [Y,eigvals] = cmdscale(D);
    figure(i+1)
    gscatter(Y(:,1),Y(:,2),results(:,i))
    title('Multidimensional Scaling')
    xlabel('1st Principal Component')
    ylabel('2nd Principal Component')
end
```

## OM VE OMBBK ÖNGÖRÜ PERFORMANSLARI

Table G.1 Ocak-Aralık 2017 Dönemi Öngörü Performansları 1

Mahalleler	OMBBK			OM			Kazanan Model		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
19 Mayıs	13.75	6.60	14.95	9.77	4.59	10.59	OM	OM	OM
23 Nisan	3.69	2.75	4.08	5.94	4.52	6.59	OMBBK	OMBBK	OMBBK
28 Haziran	11.62	9.43	13.43	12.45	10.19	14.41	OMBBK	OMBBK	OMBBK
Abbasaga	7.84	3.59	8.93	8.75	4.03	9.91	OMBBK	OMBBK	OMBBK
Acibadem	3.03	1.57	4.11	2.89	1.51	3.77	OM	OM	OM
Adnan Menderes	7.09	5.09	7.56	9.91	7.29	10.70	OMBBK	OMBBK	OMBBK
Ahlattibel	2.00	1.64	2.80	2.17	1.76	2.55	OMBBK	OMBBK	OM
Ahmetpaşa	5.33	3.72	6.60	6.11	4.30	7.63	OMBBK	OMBBK	OMBBK
Akarca	8.07	4.87	10.39	8.53	5.18	10.94	OMBBK	OMBBK	OMBBK
Akat	14.61	4.31	16.39	12.82	3.39	16.75	OM	OM	OMBBK
Akin Simav	2.77	1.70	3.29	4.41	2.78	5.59	OMBBK	OMBBK	OMBBK
Akpinar	0.77	0.65	0.90	0.94	0.80	1.08	OMBBK	OMBBK	OMBBK
Aksoy	24.72	10.96	27.07	21.24	9.25	23.06	OM	OM	OM
Aktepe	2.39	1.40	2.77	3.85	2.29	4.48	OMBBK	OMBBK	OMBBK
Alaybey	16.34	7.53	19.17	16.32	7.52	19.22	OM	OM	OMBBK
Aliahya Merkez	18.73	13.30	20.50	19.58	14.03	21.58	OMBBK	OMBBK	OMBBK
Alsancak	2.58	1.79	2.97	4.34	3.16	4.82	OMBBK	OMBBK	OMBBK
Altınşehir	3.08	1.45	3.65	6.55	2.98	6.86	OMBBK	OMBBK	OMBBK
Altıntaş	7.81	4.11	10.08	3.99	2.04	5.15	OM	OM	OM
Altıparmak	15.29	12.37	17.40	18.53	15.50	21.07	OMBBK	OMBBK	OMBBK
Anittepe	21.88	14.22	23.98	22.24	14.49	24.42	OMBBK	OMBBK	OMBBK
Arnavutköy	18.76	11.53	19.40	31.67	21.71	33.80	OMBBK	OMBBK	OMBBK
Asağı Dikmen	4.64	2.94	5.23	5.63	3.60	6.43	OMBBK	OMBBK	OMBBK
Asağı Eglence	12.59	10.54	14.75	14.60	12.50	17.00	OMBBK	OMBBK	OMBBK
Asağı Öveçler	3.13	2.60	3.67	3.85	3.23	4.54	OMBBK	OMBBK	OMBBK
Ata	1.57	1.41	1.73	2.16	1.95	2.37	OMBBK	OMBBK	OMBBK
Ataevler	26.03	14.87	28.84	27.19	15.67	30.22	OMBBK	OMBBK	OMBBK
Atakent	17.02	7.16	18.06	9.14	3.75	10.13	OM	OM	OM
Atapark	3.50	2.37	4.19	4.43	3.03	5.31	OMBBK	OMBBK	OMBBK
Atilla	9.97	5.30	12.20	5.51	2.83	7.11	OM	OM	OM
Aydınlar	1.69	1.05	2.00	1.82	1.13	2.07	OMBBK	OMBBK	OMBBK
Ayrancı	1.83	1.13	2.29	1.95	1.24	2.15	OMBBK	OMBBK	OM
Ayvalı	2.86	2.21	3.37	4.05	3.16	4.52	OMBBK	OMBBK	OMBBK
Aziziye	1.11	0.97	1.35	1.54	1.35	1.85	OMBBK	OMBBK	OMBBK
Bademlik	5.29	3.32	5.80	6.83	4.35	7.55	OMBBK	OMBBK	OMBBK
Bağlarbaşı	4.54	3.29	5.26	5.59	4.08	6.46	OMBBK	OMBBK	OMBBK
Bahariye	7.95	4.07	9.27	1.57	0.78	1.82	OM	OM	OM
Bahçelievler	8.41	3.86	9.84	7.88	3.60	9.34	OM	OM	OM
Bahriye Üçok	4.48	1.98	5.15	4.42	1.96	5.16	OM	OM	OMBBK
Balat	10.27	6.22	11.57	6.54	3.85	7.29	OM	OM	OM



Table G.2 Ocak-Aralık 2017 Dönemi Öngörü Performansları 2

Mahalleler	OMBBK			OM			Kazanan Model		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Balgat	17.22	11.32	18.95	17.30	11.38	19.04	OMBBK	OMBBK	OMBBK
Balmumcu	38.60	13.93	42.11	47.11	17.64	51.27	OMBBK	OMBBK	OMBBK
Barbaros	1.30	1.11	1.92	1.91	1.65	2.71	OMBBK	OMBBK	OMBBK
Baris	6.41	3.43	7.98	7.53	4.07	9.59	OMBBK	OMBBK	OMBBK
Basinevleri	1.93	1.27	2.66	3.65	2.46	4.82	OMBBK	OMBBK	OMBBK
Bebek	37.95	9.80	44.83	9.15	2.15	10.76	OM	OM	OM
Besevler	19.96	11.20	22.27	17.26	9.51	19.16	OM	OM	OM
Beytepe	2.89	2.72	3.08	2.45	2.32	2.57	OM	OM	OM
Birlik	2.73	1.69	3.45	1.86	1.17	2.19	OM	OM	OM
Bostanci	4.07	1.54	5.37	3.50	1.32	4.33	OM	OM	OM
Bostanlı	12.19	5.48	14.90	11.78	5.28	14.45	OM	OM	OM
Büyükesat	26.34	17.59	28.38	22.85	14.85	24.58	OM	OM	OM
Caddebostan	9.61	3.53	11.23	16.12	6.17	19.75	OMBBK	OMBBK	OMBBK
Caferaga	12.23	3.92	13.85	12.42	3.98	14.06	OMBBK	OMBBK	OMBBK
Cebeci	13.73	9.66	15.18	16.45	11.84	18.10	OMBBK	OMBBK	OMBBK
Cedit	19.50	10.67	22.35	21.17	11.72	24.29	OMBBK	OMBBK	OMBBK
Cevizlidere	2.30	1.45	2.52	3.27	2.07	3.53	OMBBK	OMBBK	OMBBK
Cihannüma	2.72	1.56	3.14	5.87	3.30	6.07	OMBBK	OMBBK	OMBBK
Cumhuriyet	14.99	7.23	16.68	3.31	1.53	4.17	OM	OM	OM
Çaldıran	8.13	6.05	8.76	9.18	6.89	9.96	OMBBK	OMBBK	OMBBK
Çalseki	2.64	1.76	3.19	3.94	2.66	4.68	OMBBK	OMBBK	OMBBK
Çamlıca	13.50	8.33	15.46	14.74	9.20	16.95	OMBBK	OMBBK	OMBBK
Çamlıtepe	6.95	3.54	7.51	3.80	1.98	4.79	OM	OM	OM
Çankaya	8.23	4.29	9.13	10.23	5.39	11.43	OMBBK	OMBBK	OMBBK
Çayyolu	11.85	8.51	14.93	12.80	9.30	16.16	OMBBK	OMBBK	OMBBK
Çekirge	13.32	8.96	15.03	16.80	11.64	18.95	OMBBK	OMBBK	OMBBK
Çirpan	6.84	4.57	8.92	5.30	3.48	6.76	OM	OM	OM
Çiçekli	9.83	7.11	10.75	10.76	7.85	11.84	OMBBK	OMBBK	OMBBK
Çigdem	13.55	8.94	14.14	14.46	9.61	15.29	OMBBK	OMBBK	OMBBK
Çukurambar	1.16	0.97	1.35	2.12	1.74	2.53	OMBBK	OMBBK	OMBBK
Dedebasi	10.50	5.63	11.86	11.01	5.93	12.51	OMBBK	OMBBK	OMBBK
Degirmendere Merkez	15.43	11.79	16.77	17.01	13.21	18.73	OMBBK	OMBBK	OMBBK
Demirci	7.66	4.28	8.68	2.15	1.17	2.27	OM	OM	OM
Demirköprü	14.33	6.72	17.94	13.73	6.42	17.30	OM	OM	OM
Demirtas Cumhuriyet	10.95	9.20	12.96	10.44	8.71	12.22	OM	OM	OM
Demirtaspasa	5.05	3.25	5.70	7.36	4.81	7.97	OMBBK	OMBBK	OMBBK
Dikilitas	15.96	7.86	19.19	19.31	9.72	22.86	OMBBK	OMBBK	OMBBK
Dikkaldırım	9.90	5.60	11.55	13.94	8.12	16.17	OMBBK	OMBBK	OMBBK
Doburca	6.91	3.78	8.99	3.47	1.85	4.25	OM	OM	OM
Donanmaci	24.72	12.10	28.19	24.31	11.88	27.93	OM	OM	OM
Dumlupınar	7.14	2.24	9.37	18.48	5.53	22.59	OMBBK	OMBBK	OMBBK
Eğitim	3.77	1.20	4.61	20.93	6.16	24.51	OMBBK	OMBBK	OMBBK
Emek	6.57	3.64	6.86	8.54	4.78	8.94	OMBBK	OMBBK	OMBBK
Emrah	7.80	5.49	8.93	6.89	4.81	7.99	OM	OM	OM
Erenköy	5.52	1.84	5.67	6.30	2.09	6.44	OMBBK	OMBBK	OMBBK

**Table G.3 Ocak-Aralık 2017 Dönemi Öngörü Performansları 3**

Mahalleler	OMBBK			OM			Kazanan Model		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Ertugrul	7.08	3.38	8.80	7.82	3.75	9.88	OMBBK	OMBBK	OMBBK
Esertepe	4.70	2.93	5.03	7.47	4.74	7.93	OMBBK	OMBBK	OMBBK
Etiler	15.47	6.67	17.90	25.48	10.37	29.74	OMBBK	OMBBK	OMBBK
Etlik	10.31	6.13	11.19	11.48	6.88	12.54	OMBBK	OMBBK	OMBBK
Fenerbahçe	15.33	5.28	16.08	20.08	6.77	21.46	OMBBK	OMBBK	OMBBK
Feneryolu	6.96	2.60	7.52	4.09	1.56	4.44	OM	OM	OM
Fethiye	8.30	4.91	9.58	7.71	4.54	9.00	OM	OM	OM
Fikirtepe	4.61	1.23	6.20	3.89	1.05	4.44	OM	OM	OM
Fikri Altay	13.11	7.09	15.75	12.62	6.81	15.30	OM	OM	OM
Gayrettepe	17.32	7.55	18.42	19.06	8.38	20.35	OMBBK	OMBBK	OMBBK
Gaziosmanpasa	3.92	2.24	4.24	6.30	3.70	7.99	OMBBK	OMBBK	OMBBK
Goncalar	4.75	2.43	6.11	1.43	0.73	1.81	OM	OM	OM
Görükle	14.49	8.61	17.31	7.45	4.19	8.54	OM	OM	OM
Göztepe	4.34	1.64	5.30	6.72	2.58	8.32	OMBBK	OMBBK	OMBBK
Güçlükaya	6.28	4.14	8.28	7.72	5.17	10.13	OMBBK	OMBBK	OMBBK
Gümüşdere	0.58	0.44	0.69	1.92	1.46	2.15	OMBBK	OMBBK	OMBBK
Güvenevler	1.68	1.07	1.97	2.09	1.34	2.73	OMBBK	OMBBK	OMBBK
Güzelyali	7.56	3.84	10.24	8.72	4.47	11.74	OMBBK	OMBBK	OMBBK
Harbiye	3.87	2.24	4.81	4.69	2.73	5.81	OMBBK	OMBBK	OMBBK
Hasanpasa	13.72	5.08	14.76	14.41	5.35	15.58	OMBBK	OMBBK	OMBBK
Hilal	4.03	3.89	4.22	5.82	5.72	5.94	OMBBK	OMBBK	OMBBK
Hocahasan	13.15	9.95	15.65	14.16	10.82	16.88	OMBBK	OMBBK	OMBBK
Huzur	4.01	2.24	4.94	2.38	1.35	2.88	OM	OM	OM
Hüdavendigâr	8.59	4.93	11.12	7.34	4.17	9.54	OM	OM	OM
İbrahimpasa	8.14	5.75	9.72	10.79	7.82	12.72	OMBBK	OMBBK	OMBBK
İhsaniye	9.08	4.92	10.16	7.69	4.13	8.70	OM	OM	OM
İlkadim	5.01	3.51	6.16	6.44	4.57	8.02	OMBBK	OMBBK	OMBBK
İlkbahar	2.73	2.54	3.39	6.41	5.76	7.05	OMBBK	OMBBK	OMBBK
İlker	12.10	7.69	12.98	12.85	8.21	13.94	OMBBK	OMBBK	OMBBK
İncirli	0.98	0.63	1.13	1.49	0.97	1.70	OMBBK	OMBBK	OMBBK
İşçi Blokları	1.48	0.81	1.99	1.75	0.95	2.28	OMBBK	OMBBK	OMBBK
Kadıköy	9.10	6.09	11.51	12.98	9.03	16.03	OMBBK	OMBBK	OMBBK
Kalaba	2.78	1.71	3.20	3.69	2.29	4.25	OMBBK	OMBBK	OMBBK
Kamil Ocak	5.10	3.96	5.68	6.15	4.83	6.84	OMBBK	OMBBK	OMBBK
Kanuni	6.65	3.57	7.15	6.29	3.37	6.76	OM	OM	OM
Karapınar	0.76	0.45	1.01	0.86	0.51	1.16	OMBBK	OMBBK	OMBBK
Kavacık Subayevleri	6.60	4.69	7.45	9.41	6.86	10.54	OMBBK	OMBBK	OMBBK
Kavaklıdere	7.09	4.11	7.83	2.70	1.61	3.01	OM	OM	OM
Keklik Pınarı	8.57	5.73	9.67	9.33	6.28	10.57	OMBBK	OMBBK	OMBBK
Kiliç Reis	13.75	7.42	16.36	9.05	4.72	10.56	OM	OM	OM
Kirkkonaklar	3.77	2.30	4.17	2.05	1.27	2.46	OM	OM	OM
Kızılay	1.48	0.91	1.94	5.66	3.58	6.58	OMBBK	OMBBK	OMBBK
Kirazlı	3.94	2.69	4.25	8.06	5.31	9.03	OMBBK	OMBBK	OMBBK
Konak	13.14	6.71	16.03	4.10	1.96	5.17	OM	OM	OM
Konaklar	11.36	6.53	12.56	14.60	8.58	16.17	OMBBK	OMBBK	OMBBK

**Table G.4 Ocak-Aralık 2017 Dönemi Öngörü Performansları 4**

Mahalleler	OMBBK			OM			Kazanan Model		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Kosuyolu	2.50	0.99	3.25	2.64	1.04	3.76	OMBBK	OMBBK	OMBBK
Kozyatagi	19.77	7.15	22.88	17.90	6.42	20.80	OM	OM	OM
Kösk	9.58	7.46	10.55	9.73	7.60	10.79	OMBBK	OMBBK	OMBBK
Kuruçesme	8.27	3.52	10.01	23.64	9.22	24.78	OMBBK	OMBBK	OMBBK
Kuscagiz	3.42	2.27	4.07	4.36	2.92	5.18	OMBBK	OMBBK	OMBBK
Küçükkesat	7.75	4.71	8.85	12.54	7.92	14.38	OMBBK	OMBBK	OMBBK
Kükürtlü	14.79	8.99	17.80	19.73	12.49	23.14	OMBBK	OMBBK	OMBBK
Kültür	10.91	6.19	13.85	7.07	3.88	9.31	OM	OM	OM
Levazim	21.31	5.19	25.57	12.30	2.95	14.47	OM	OM	OM
Levent	18.91	8.64	19.99	17.63	8.01	18.58	OM	OM	OM
Maksem	6.82	5.01	8.72	10.31	7.86	12.68	OMBBK	OMBBK	OMBBK
Maltepe	11.62	7.10	12.27	11.13	6.78	11.94	OM	OM	OM
Mavisehir	15.99	7.50	18.01	17.44	8.26	19.71	OMBBK	OMBBK	OMBBK
Mecidiye	1.49	0.82	1.83	4.29	2.28	5.33	OMBBK	OMBBK	OMBBK
Merdivenköy	31.68	11.52	34.13	32.78	11.97	35.41	OMBBK	OMBBK	OMBBK
Merkez	1.36	1.06	1.72	2.80	2.22	3.31	OMBBK	OMBBK	OMBBK
Mithatpasa	5.35	4.18	6.38	8.27	6.66	9.47	OMBBK	OMBBK	OMBBK
Muradiye	15.29	7.10	17.27	16.08	7.50	18.23	OMBBK	OMBBK	OMBBK
Mutlukent	7.56	4.07	8.59	11.02	6.07	12.50	OMBBK	OMBBK	OMBBK
Namik Kemal	0.31	0.25	0.35	1.46	1.21	1.64	OMBBK	OMBBK	OMBBK
Nergiz	13.03	5.24	14.42	11.50	4.59	12.68	OM	OM	OM
Nispetiye	9.49	4.10	12.11	12.68	5.36	16.04	OMBBK	OMBBK	OMBBK
Oran	20.07	10.89	21.45	20.54	11.10	22.09	OMBBK	OMBBK	OMBBK
Ortaköy	16.34	7.60	18.71	17.99	8.46	20.68	OMBBK	OMBBK	OMBBK
Osmanaga	13.62	5.94	16.09	12.77	5.54	15.12	OM	OM	OM
Osmangazi	0.72	0.52	0.86	0.83	0.60	1.00	OMBBK	OMBBK	OMBBK
Ovaakça Merkez	15.52	11.39	17.19	16.84	12.51	18.71	OMBBK	OMBBK	OMBBK
Örnekköy	5.92	2.92	7.83	5.10	2.50	6.89	OM	OM	OM
Öveçler	7.80	4.86	9.30	4.50	2.89	5.64	OM	OM	OM
Özlüce	4.24	1.99	4.48	5.91	2.79	6.10	OMBBK	OMBBK	OMBBK
Pinarbasi	0.88	0.60	1.05	1.18	0.80	1.31	OMBBK	OMBBK	OMBBK
Prof. Dr. Ahmet Taner Kislali	12.70	13.84	14.74	15.68	17.85	18.17	OMBBK	OMBBK	OMBBK
Rasimpasa	7.18	2.12	8.53	8.52	2.50	10.00	OMBBK	OMBBK	OMBBK
Sahrayi Cedit	11.51	4.44	15.04	12.91	5.02	16.78	OMBBK	OMBBK	OMBBK
Sancak	2.50	1.40	3.01	1.66	0.94	2.03	OM	OM	OM
Sancaktepe	2.19	1.33	2.76	3.58	2.20	4.62	OMBBK	OMBBK	OMBBK
Seyranbaglari	14.71	9.30	17.06	15.18	9.63	17.65	OMBBK	OMBBK	OMBBK
Sinanpasa	3.00	1.56	3.62	6.26	3.03	7.74	OMBBK	OMBBK	OMBBK
Sokullu Mehmet Pasa	5.43	3.08	7.73	4.58	2.63	6.48	OM	OM	OM
Suadiye	9.95	3.50	10.48	10.35	3.64	10.89	OMBBK	OMBBK	OMBBK
Sefkat	2.71	1.56	3.18	7.25	4.32	8.20	OMBBK	OMBBK	OMBBK
Sehit Kubilay	3.93	2.88	4.63	5.63	4.21	6.93	OMBBK	OMBBK	OMBBK
Semikler	16.00	7.08	18.53	13.61	5.94	15.78	OM	OM	OM
Senlik	1.40	0.85	1.67	2.56	1.59	3.42	OMBBK	OMBBK	OMBBK
Tepebasi	4.28	2.74	5.26	15.33	10.79	17.37	OMBBK	OMBBK	OMBBK

**Table G.5 Ocak-Aralık 2017 Dönemi Öngörü Performansları 5**

Mahalleler	OMBBK			OM			Kazanan Model		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE
Türkali	5.09	2.20	5.71	6.28	2.69	7.01	OMBBK	OMBBK	OMBBK
Ufuktepe	4.07	2.80	4.58	4.73	3.28	5.40	OMBBK	OMBBK	OMBBK
Ulus	4.67	1.78	5.37	4.64	1.72	5.87	OM	OM	OMBBK
Uyanis	3.31	1.76	3.71	5.39	2.90	5.97	OMBBK	OMBBK	OMBBK
Üçevler	9.52	5.41	12.03	10.19	5.83	12.86	OMBBK	OMBBK	OMBBK
Ümit	12.46	7.77	14.58	15.45	9.88	17.99	OMBBK	OMBBK	OMBBK
Visnezade	9.06	4.24	11.32	10.69	5.16	15.21	OMBBK	OMBBK	OMBBK
Yahyakaptan	18.50	8.85	21.40	13.60	6.31	15.76	OM	OM	OM
Yakacik	3.09	1.85	3.98	5.69	3.50	7.29	OMBBK	OMBBK	OMBBK
Yali	13.44	5.82	15.79	16.06	7.06	18.84	OMBBK	OMBBK	OMBBK
Yasamkent	2.16	1.19	2.28	2.97	1.65	3.19	OMBBK	OMBBK	OMBBK
Yayla	1.30	0.77	1.51	1.68	1.00	2.58	OMBBK	OMBBK	OMBBK
Yenisehir	17.46	10.03	20.31	22.99	13.77	26.46	OMBBK	OMBBK	OMBBK
Yıldiz	8.22	4.48	10.66	11.80	6.64	15.10	OMBBK	OMBBK	OMBBK
Yıldizevler	13.84	9.85	14.66	14.55	10.43	15.47	OMBBK	OMBBK	OMBBK
Yukari Dikmen	6.19	3.37	7.93	4.12	2.18	4.86	OM	OM	OM
Yunuseli	11.58	5.46	13.35	6.29	2.88	6.90	OM	OM	OM
Yükseltepe	3.08	2.45	3.24	1.80	1.46	1.97	OM	OM	OM
Yüzüncüyil	5.99	3.14	7.98	2.83	1.46	3.08	OM	OM	OM
Zafertepe	8.01	4.22	9.19	4.68	2.41	5.49	OM	OM	OM
Zühtüpaşa	1.71	1.52	1.86	1.31	1.18	1.53	OM	OM	OM

## TEZDEN ÜRETİLMİŞ YAYINLAR

---

İletişim Bilgisi: vatansevermetin@gmail.com

### Makaleler

1. M. Vatansever, İ. Demir ve A. Hepşen, "Cluster and Forecasting Analysis of the Residential Market in Turkey: An Autoregressive Model-Based Fuzzy Clustering Approach", International Journal of Housing Markets and Analysis, cilt 13, sayı 2, 2020.
2. M. Vatansever ve İ. Demir, "A Segmentation Study of Non-Performing Loans Rates in Turkish Credit Market", International Business Research, cilt 10, sayı 11, ss. 29-41, 2017.

