



## FORECASTING OF TURKISH HOUSING PRICE INDEX: ARIMA, RANDOM FOREST, ARIMA-RANDOM FOREST

DOI: 10.17261/Pressacademia.2019.1134

PAP-IFC- V.10-2019(2)-p.7-11

Ebru Çağlayan Akay<sup>1</sup>, Kadriye Hilal Topal<sup>2</sup>, Saban Kizilarlan<sup>3</sup>, Hoseng Bulbul<sup>4</sup>

<sup>1</sup>Marmara University, Faculty of Economics, Department of Econometrics, Goztepe Campus, Istanbul, Turkey.

[ecaglayan@marmara.edu.tr](mailto:ecaglayan@marmara.edu.tr), ORCID: 0000-0002-9998-5334

<sup>2</sup>Nisantasi University, Nişantasi Vocational School, Program of Banking and Insurance, Maslak Campus, Istanbul, Turkey.

Marmara University, PhD. Candidate in Econometrics, Goztepe Campus, Istanbul, Turkey.

[hilal.topal@nisantasi.edu.tr](mailto:hilal.topal@nisantasi.edu.tr), ORCID: 0000-0001-5203-8017

<sup>3</sup>Van YuzuncuYil University, Faculty of Economics and Administrative Sciences, Department of Econometrics, Zeve Campus, Van, Turkey.

Marmara University, PhD. Candidate in Econometrics, Goztepe Campus, Istanbul, Turkey.

[saban.kizilarlan@marmara.edu.tr](mailto:saban.kizilarlan@marmara.edu.tr), ORCID: 0000-0003-1545-9597

<sup>4</sup>Marmara University, Faculty of Economics, Department of Econometrics, Goztepe Campus, Istanbul, Turkey.

[hoseng.bulbul@marmara.edu.tr](mailto:hoseng.bulbul@marmara.edu.tr), ORCID: 0000-0002-4541-8916

### To cite this document

ÇağlayanAkay, E., Topal K. H., Kizilarlan S., Bulbul H., (2019). Forecasting of Turkish housing price index: ARIMA, random forest, ARIMA-random forest. PressAcademia Procedia (PAP), V.10, p.7-11.

Permanent link to this document: <http://doi.org/10.17261/Pressacademia.2019.1134>

Copyright: Published by PressAcademia and limited licenced re-use rights only.

### ABSTRACT

**Purpose-** The aim of this study is to investigate whether the models developed by using Machine Learning methods are an alternative method for forecasting time series.

**Methodology-**Traditionally, the Autoregressive Integrated Moving Average (ARIMA) model has been one of the most widely used linear models in time series forecasting. In the study, we use Random Forest and Hybrid Random Forest-ARIMA models besides the ARIMA model and compare their forecasting performance for the Turkish Housing Price Index series.

**Findings-** The hybrid model was found to be more successful than other methods in forecasting the housing price index.

**Conclusion-** As a result, hybrid models that combine ARIMA and machine learning method can be used an alternative method in forecasting economic and financial data.

**Keywords:** ARIMA, random forest, ARIMA-random forest, hybrid, machine learning.

**JEL Codes:** C52, C58, C61

## TÜRKİYE KONUT FİYAT ENDEKSİ ÖNGÖRÜSÜ: ARIMA, RASSAL ORMAN VE ARIMA-RASSAL ORMAN

### ÖZET

**Amaç-** Bu çalışmanın amacı, Makine Öğrenmesi yöntemlerinden yararlanarak geliştirilen modellerin zaman serilerinin öngörüsünde alternatif bir yöntem olup olmadığının incelenmesidir.

**Yöntem-** Geleneksel olarak, Otoregresif Entegre Hareketli Ortalama (ARIMA) modeli, zaman serisi tahmininde en yaygın kullanılan doğrusal modellerden biridir. Çalışmada, ARIMA modellerinin yanı sıra Rassel Orman ve Hibrit Rassel Orman yöntemleri kullanılmış ve Türkiye Konut Fiyat Endeksi serisi için bu modellerin öngörü performansları karşılaştırılmıştır.

**Bulgular-** Hibrit modelin konut fiyat endeksini öngörmeye diğer yöntemlerden daha başarılı olduğu tespit edilmiştir.

**Sonuç-** Sonuç olarak, ARIMA ve Makine Öğrenmesi yöntemini birleştiren hibrit modellerin, ekonomik ve finansal verilerin öngörüsünde alternatif bir yöntem olarak kullanılabileceği tespit edilmiştir.

**Anahtar Kelimeler:** ARIMA, rassal orman, ARIMA-rassal orman, hibrit, makine öğrenmesi.

**JEL Kodları:** C52, C58, C6

### 1. GİRİŞ

Türkiye’de inşaat sektörü ekonomik büyümeye en çok katkı sağlayan sektörlerden biridir ve oldukça güçlü bir yapıya sahiptir. Sektöre yönelik yüksek yatırımlar arz-talep dengesinin değişmesine ve gelişmiş bir yatırım aracı olarak görülen konut piyasasında fiyatlar düzeyinde önemli artışların olmasına neden olmaktadır. Ayrıca konut kredilerindeki genişlemeyle, bankacılık sektöründe ve ekonomik büyüme üzerinde yarattığı önemli etkiler nedeniyle konut piyasasının geleceğini öngörmek önem kazanmaktadır. Finansal anlamda ele alınacak olursa, konut kredileri bankacılık sektörü için önemli bir kaynak oluşturmaktadır. TCMB, konut fiyatlarındaki değişimlerin takip edilmesi

amacıyla her yıl düzenli olarak konut fiyatları endeksi hesaplanmaktadır (TCMB;2018, 27 Ekim). Konut fiyatları ülke ekonomisinin değişiminde bir parametre olduğundan, bu fiyatların tahmini geleceğe yönelik planlamalarda önemli bir yere sahiptir. Finansal zaman serilerinin tahmin ve öngörüsünde doğrusal zaman serisi modellerinden biri olan ARIMA modelleri yaygın olarak kullanılmaktadır. ARIMA modelleri, serideki değişkenliği geçmiş dönem etkileri ve rassallıkla açıklayan ve yaygın olarak kullanılan modellerdir. Bu modellerin tercih edilme sebebi öngörü yöntemlerinin çalışması açısından uygun doğrusal modeller olmalarıdır. Doğrusal zaman serilerindeki başarılı performanslarına rağmen, doğrusal olmamanın çok yüksek olduğu zaman serilerinin modellenmesinde ARIMA gibi klasik yöntemler yeterli olmamaktadır. Bu nedenle, son zamanlarda bu tür zaman serileri için Makine Öğrenmesi yöntemleri sıklıkla kullanılmaya başlanmıştır. Literatürde en yaygın kullanılan Makine Öğrenmesi yöntemleri; Yapay Sinir Ağları (YSA), Destek Vektör Regresyonu (DVR) ve Rassal Orman (RO) olarak sıralanabilir. Bu yöntemlerden biri olan YSA, öngörü başarısı yüksek bir yöntem olmasına rağmen, küçük dağılımlı doğrusal süreçlerde doğrusal yöntemler kadar başarılı değildir. Ayrıca parametre sayısının yüksek olduğu veri setlerinde, bu yöntem lokal tahmin odaklı yapısı sebebiyle aşırı uyum gösterme eğilimindedir. Bunların yanı sıra, fonksiyonun en başarılı yakınsamayı sağladığı girdi-çıkı eşleştirilmesinde deneme-yanılma yöntemleri sıklıkla kullanıldığından ve veriyi öğrenmede daha fazla bilgi ve zaman gerektirdiğinden bu yöntem çok tercih edilmemektedir (Zhang vd., 1998, s.56). DVR ve RO ise klasik yöntemlerde olan normallik, doğrusallık, değişken bağımsızlığı gibi sıkı varsayımlar gerektirmediğinden ve YSA gibi lokal optimuma değil de global optimum çözüme odaklandığından daha başarılı sonuçlar vermektedir. Bu nedenle DVR ve RO yöntemleri literatürde daha çok tercih edilmektedir (Kumar ve Thenmozhi, 2014, s.285). Son zamanlarda Makine Öğrenmesi yöntemlerinin yanı sıra, ARIMA modelleri ile Makine Öğrenmesi yöntemlerinin birleştirilmesiyle elde edilen hibrit yaklaşım finansal verilerin ekonometrik modellemesinde alternatif olarak sunulmaktadır. Bu çalışmanın amacı, Türkiye Konut Fiyat Endeksinin öngörüsü için klasik ARIMA modelleri yanında makine öğrenmesi modellerinden RO ve hibrit ARIMA-RO modelleri kullanarak tahmin yapmak ve bu modellerin tahmin ve öngörü başarı performanslarını karşılaştırmaktır.

Çalışmanın birinci bölümünde çalışmayla ilgili genel bilgi verilmiştir. İkinci bölümde Hibrit Makine Öğrenmesi yöntemlerinin finansal verilerin analizinde kullanıldığı önceki çalışmalara değinilmiştir. Üçüncü bölümde kullanılan yöntemlerin teorik yapısı incelendikten sonra, dördüncü bölümde veri seti hakkında bilgi verilip araştırmanın hipotezleri ortaya konulmuştur. Beşinci bölümde ise ARIMA, RO ve hibrit ARIMA-RO modellerinden elde edilen sonuçlar tablo ve grafiklerle sunulmuştur. Son bölümde ise sonuçlar yorumlanmıştır.

## 2. LİTERATÜR

Bu çalışmada finansal bir veri olan konut fiyatları endeksinin öngörüsü için alternatif yöntemlerden yararlanılmıştır. Bu yöntemler klasik zaman serisi yöntemlerinden ARIMA modelinin yanı sıra Makine Öğrenmesi yöntemlerinden biri olan RO yöntemidir. Ayrıca bu yöntemlerin birlikte kullanılmasıyla geliştirilen hibrit yöntemden de yararlanılmıştır. Bu yöntemlerin bir arada kullanılması, alternatif yöntemlerin avantajlarından eş zamanlı olarak faydalanabilmeyi amaçlamaktadır. Sosyal bilimlerde bu yöntemlerin kullanıldığı çalışmalar, genellikle döviz kuru ve borsa endeksi gibi finansal serilerin öngörüsünü yapmak amaçlamaktadır. Finansal verilerin analizinde Makine Öğrenmesi yöntemleri yaygın olarak kullanılsa da, hibrit model yaklaşımının zaman serileri için kullanımı Zhang (2003) çalışmasıyla başlamıştır. Zhang bu çalışmada İngiltere'nin 1980-1993 haftalık döviz kuru serisine ARIMA, Yapay Sinir Ağları ve ARIMA-Yapay Sinir Ağları hibrit yöntemini uygulayarak sonuçlarını karşılaştırmış, ARIMA-Yapay Sinir Ağları yönteminin performansının diğerlerinden daha üstün olduğu sonucuna ulaşmıştır. Pai ve Lin (2005) çalışmasında 2003:10-2009:02 aralığında ABD günlük borsa endeksi serisi için ARIMA, Destek Vektör Regresyonu ve ARIMA- Destek Vektör Regresyonu yöntemlerini kullanmış ve ARIMA-Destek Vektör Regresyonu yöntemi daha başarılı bulunmuştur. Kumar ve Thenmozhi(2014) çalışmasında 2003:01-2009:12 arası için Hindistan'ın günlük borsa endeksi verisine ARIMA, Yapay Sinir Ağları, Destek Vektör Regresyonu, Rassal Orman yöntemlerinin yanısıra ARIMA-Yapay Sinir Ağları, ARIMA-Destek Vektör Regresyonu, ARIMA-Rassal Orman hibrit yöntemlerini uygulamış ve ARIMA-Destek Vektör Regresyonu yönteminin öngörü başarısının diğer yöntemlerden daha üstün olduğu bulgusuna ulaşmıştır.

## 3. METODOLOJİ

Çalışmada ARIMA modeli, Rassal Orman ve Hibrit ARIMA-Rassal Orman yöntemleri kullanılmıştır. Bu bölümde ilgili yöntemler hakkında teorik bilgiye yer verilmiştir.

**ARIMA Modeli:** Box-Jenkins ARIMA modelleri ekonomik zaman serilerinin davranışlarını daha iyi anlamak için modelleme ve öngöründe yaygın olarak kullanılan tek değişkenli zaman serisi modellerinden biridir. Bu modellerde bir değişkenin bugünkü değeri geçmiş gözlemlerin ve rassal şokların doğrusal bir fonksiyonu olduğu varsayılmaktadır. ARIMA(p,d,q) modellerini oluşturmak için otoregresif parametre AR(p), hareketli ortalama parametresi MA(q) ve bütünleşme parametresi I(d) bir araya getirilmektedir. Durağan bir zaman serisi aşağıdaki gibi modellenebilir;

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \dots - \theta_q \varepsilon_{t-q}$$

Burada p otoregresyon mertebesi, d bütünleşme mertebesi ve q hareketli ortalama mertebesi göstermektedir.  $y_t$  zamanındaki gerçek değer,  $\varepsilon_t$  hata terimi ve  $\phi_i$  ( $i = 1, 2, \dots, p$ ) ve  $\theta_j$  ( $j = 0, 1, 2, \dots, q$ ) model parametreleridir.  $\varepsilon_t$  hata teriminin sıfır ortalamalı ve  $\sigma^2$  sabit varyanslı bağımsız ve özdeş dağılıma sahip olduğu varsayılmaktadır. ARIMA modellerinin tahmininde model tanımlama, parametre tahmini ve ayırt edici kontrol aşamalarından oluşan Box-Jenkins (1976) metodolojisi kullanılmaktadır. Bu aşamalar en iyi model seçilene kadar tekrarlanır ve son aşamada seçilen model öngörü yapmak için kullanılır (Franses ve Van Dijk, 2003, s.21-25; Wang, 2008, s.37).

**Rassal Orman (RO):** Ho (1998) ve Breiman(2001) tarafından oluşturulan tahmin başarısı yüksek bir algoritmadır. Birden fazla karar ağacının aynı bağımlı değişken için oluşturulması ve ortak karar verilmesi işlemidir. Sınıfların çoğunluk seçimine (oyu) bakılarak karar verilir. N farklı regresyon ağaçları,  $\hat{C}_n(x)$ , n.rassal ağacın sınıf tahmini olmak üzere,  $\hat{C}_n^N(x) = \frac{1}{N} \{\hat{C}_n(x)\}_1^N$  olur. Rassal orman regresyonu ise veri kümesinin birden çok parçaya bölünmesi ve her biri için regresyon ile elde edilen tahmin sonuçlarının ortalamasının alınması yöntemiyle

karar verilen bir kolektif öğrenme algoritmasıdır. Bu algoritmanın parametre sayısı (mtry) ve ağaç sayısı (N) olmak üzere iki önemli parametresi bulunmaktadır (Hastie, vd., 2009).

**Hibrit Yöntem - ARIMA-Rassal Orman:** Zaman serileri ile çalışıldığında, serilerin doğrusallık yapısı önem arz etmektedir. Doğrusal yapıdaki serilerin modellenmesinde ARIMA modelleri başarılı sonuçlar verse de, bu durum doğrusal yapıda olmayan seriler için geçerli değildir. Diğer taraftan doğrusal olmayan serilerin modellenmesinde kullanılan yöntemlerin de doğrusal yapıdaki serilerin modellenmesinde her zaman başarılı olmadığı tespit edilmiştir. Dolayısıyla her durumda iyi sonuç veren genel bir yöntem yoktur. Doğrusallık boyutunu dikkate almak adına, bir zaman serisinin doğrusal ve doğrusal olmayan iki alt bileşenden oluştuğu kabul edilebilir. Bu noktadan hareketle, seriyi modellerken bu alt bileşenlerin tümünü modelleyebilmek için farklı yöntemlerin kombine edilerek kullanılması fikri ortaya çıkmıştır. İncelenen değişkenin,  $y_t = X_t + Z_t$  yapısında olduğu varsayalım. Burada  $X_t$  doğrusal bileşeni,  $Z_t$  ise doğrusal olmayan bileşeni temsil etmektedir. Doğrusal bileşenin modellenmesinde ARIMA yöntemi kullanılabilir. Seri ARIMA modeli kullanılarak modellendiğinde elde edilen tahmini değerler,  $\hat{X}_t$ , serinin doğrusal bileşenine ait tahmini değerler olacaktır. Dolayısıyla ARIMA modelinden elde edilen kalıntılar artık yalnızca doğrusal olmayan bileşeni içerecektir  $e_t^{ARIMA} = y_t - \hat{X}_t$ . Bu nedenle kalıntılar, serinin doğrusal olmayan bileşenini modellemek için kullanılabilir. Kalıntıların modellenmesinde ise doğrusallığın olmadığı durumda başarılı olan yöntemler kullanılacaktır. Kalıntılar, Rassal Orman yöntemi kullanılarak modellenebilir:  $e_t^{ARIMA} = f(e_{t-1}^{ARIMA}, e_{t-2}^{ARIMA}, L, e_{t-n}^{ARIMA})$ . Burada  $f$  fonksiyonu Rassal Orman tarafından kullanılan modeli temsil etmektedir. Bu yöntemle elde edilen tahmini değerler,  $\hat{Z}_t$ , doğrusal olmayan bileşene ait olacaktır. Son aşamada ilgilenilen serinin tahmini değerlerini elde etmek için iki yöntemden elde edilen tahmini değerlerin toplamı alınarak hibrit yöntemle ait tahmini değerler elde edilmiş olacaktır:  $\hat{y}_t = \hat{X}_t + \hat{Z}_t$ . Burada kullanılan ARIMA – Rassal Orman (ARIMA-RO) yöntemi ile doğrusallık durumunda ARIMA modellerinin, doğrusal olmama durumu için ise RO yönteminin avantajlarından faydalanılmıştır (Zhang, 2003; Kumar ve Thenmozhi, 2014).

#### 4. VERİ SETİ

Çalışmada kullanılan veri seti T.C.M.B. Elektronik Veri Dağıtım sisteminden elde edilmiştir. Konut fiyatları endeksi (KFE) serisi 2010:01-2018:10 dönemi için analiz edilmiş ve ölçeklendirme amacıyla serinin doğal logaritması alınmıştır. ARIMA modellerinin analizinde EViews programı, RO ve ARIMA-RO modellerinin analizinde R yazılımı kullanılmıştır. ARIMA modellerinde durağan serilerle çalışılmaktadır, bu nedenle öncelikle serinin durağan olup olmadığını araştırmak amacıyla ADF ve PP birim kök testi uygulanmıştır. ADF (düzey: -2.425[0.3620]; ilk fark: -4.824[0.0001]) ve PP (düzey: -2.552[0.3020]; ilk fark: -4.663 [0.0002]) birim kök testi sonuçlarına göre logaritmik konut fiyat endeksi serisinin durağan olmadığı görülmüştür. Bu nedenle çalışmada logaritmik fark serisiyle çalışılmıştır. Analizlerde veri, öğrenme veri seti ve test veri seti olarak iki parçaya ayrılmıştır. 2010:01-2017:12 dönemi öğrenme veri seti olarak ve 2018:01-2018:10 dönemi test veri seti olarak 90:10 oranında bölünmüştür. Öğrenme veri seti parametre tahminleri için, test veri seti öngörü için kullanılmıştır.

#### 5. AMPİRİK BULGULAR

Analizde öğrenme veri seti kullanılarak ARIMA için 9, makine öğrenmesi yöntemleri için 6 dönem gecikmeli modeller oluşturulmuştur. Modeller ARIMA, RO ve ARIMA-RO olmak üzere 3 farklı yöntem kullanılarak tahmin edilmiş ve sonrasında örneklem dışı statik öngörü yapılarak tahmin geçerlilikleri test edilmiştir. Öngörü tahminlerinde yaygın olarak kullanılan ARIMA modelleri doğrusal modellerdir ve serinin doğrusal yapısının tahmininde kullanılmaktadır. Bu nedenle çalışmada doğrusal yapı ARIMA modelleriyle elde edilmiştir. Tablo 1’de ARIMA model tahmin sonuçlarına yer verilmektedir.

**Tablo 1: ARIMA(9,1,0) Model Tahmini**

| ARIMA(9,1,0)   | $\phi_0$              | $\phi_1$                                       | $\phi_2$               | $\phi_3$                              | $\phi_4$            | $\phi_5$                                 |
|--|-----------------------|--|------------------------|---------------------------------------|---------------------|--|
| Katsayılar   | 0.0077***<br>(0.0015) | 0.6293***<br>(0.0731)                          | -0.2374***<br>(0.0964) | 0.2653***<br>(0.1072)                 | -0.2389<br>(0.1065) | 0.4170<br>(0.0970)                       |
| F-ist.:<br>19.537 [0.0000]***  | R-Kare:<br>0.52       | Breusch-Godfrey LM Testi:<br>14.62 [0.2625]*** |                        | ARCH LM Testi:<br>18.0235 [0.1150]*** |                     | Jarque-Bera Testi:<br>0.0696 [0.9657]*** |
| Notlar:<br>(i) ARIMA model tahmininde öncelikle korelogramdan yararlanarak Otokorelasyon Fonksiyonu (ACF) ve Kısmi Otokorelasyon Fonksiyonu (PACF) yardımıyla AR(p) ve MA(q) gecikmelerine karar verilmiştir. (ii) Alternatif ARIMA modellerinin seçiminde minimum AIC, SC değerlerine sahip ARIMA(9,1,0) modeli en uygun model olarak belirlenmiştir. (iii) Parantez içindeki değerler standart hatalar, köşeli parantez içindeki değerler ise olasılık değerlerini göstermektedir. |                       |  |                        |                                       |                     |  |

ARIMA(9,1,0) modeli uygun model olarak belirlenmiştir ve gerekli tüm varsayımları sağlamıştır. Bu modelin tahminiyle serinin doğrusal yapısı tahmin edilmiştir. Konut fiyatları serisi doğrusal ve/veya doğrusal olmayan yapıyı optimize edebilmek için RO ve ARIMA-RO yöntemiyle tahmin edilmiştir. RO yönteminde her bir ağaçta yer alacak rasgele değişken sayısı (mtry) ve toplam ağaç sayısı (N) olmak üzere iki parametre bulunmaktadır. Genellikle uygulamada RO regresyon için değişken sayısının 1/3’ünü almak önerilmektedir. Mtry bir ayarlama parametresidir ve yanlış seçimi tahmin performansında azaltıcı etki yaratabildiğinden seçiminde hassas davranılmalıdır (Tibshirani vd., 2008:592). Bu nedenle çalışmada hem RO hem de ARIMA-RO yöntemleri için mtry [1:5] arası için 10-fold CV yöntemiyle seçilmiştir. ARIMA-RO yöntemi, ARIMA model kalıntılarının RO yöntemiyle tahminlenip elde edilen tahmini kalıntıların ARIMA tahmin sonuçlarına eklenmesiyle uygulanmıştır. Burada ARIMA tahminleri modelin doğrusal kısmını, RO ile tahminlenen ARIMA tahmin kalıntıları ise modelin doğrusal olmayan kısmını oluşturmaktadır. Modelin doğrusal olmayan kısmı için ARIMA modelinden elde edilen kalıntılar tekrar Rassal Orman yöntemiyle tahmin edilmiştir. Son aşamada ise doğrusal ve doğrusal olmayan tahmin sonuçları toplanarak konut fiyatları endeksinin nihai tahmin değerleri elde edilmiştir. Modelin RO ve ARIMA-RO tahmini mtry parametreleri hatalara dayalı RMSE(0.004489;0.002738), MAE(0.003528; 0.002255) ve kriterleri minimum yapan, R<sup>2</sup> (0.999755; 0.172456) kriterini maksimum yapan değerler dikkate alındığında sırasıyla [5 ve 2] olarak elde edilmiştir. RO ve ARIMA-RO için önemli bir parametre de ağaç sayısıdır. Ağaç sayısının seçiminde genellikle çok sayıda ağaç alındığında aşırı uyum sorunu yaşanabileceğinden hem RO hem de ARIMA-RO yöntemleri için 500’e kadar çeşitli ağaç sayıları alınarak model tahminlenip, hatalara dayalı RMSE, MAE ölçütlerini minimum, R<sup>2</sup>’yi maksimum yapan ağaç sayısı belirlenmiştir. RO ve ARIMA-

RO ağaç sayısı seçiminde RMSE(0.004461;0.004544) , MAE(0.003498; 0.003539) ve R<sup>2</sup> (0.999758;0.999750) ölçütlerine bakıldığında RO ve ARIMA-RO tahmini için optimum ağaç sayısı sırasıyla [480 ve 350] olarak belirlenmiştir.RO ve ARIMA-RO hiper parametre tahmin sonuçları Tablo 2’de sunulmuştur.

**Tablo 2: Hiper-parametre Tahmin Sonuçları : Rassel Orman ve ARIMA –Rassel Orman**

| Metod    | Hiper-parametreler |     |
|----------|--------------------|-----|
| RO       | mtry               | 5   |
|          | ntree              | 480 |
| Arima-RO | mtry               | 2   |
|          | ntree              | 350 |

Hiperparametreler modelde karmaşıklığı gidermekte ve varyans-sapma dengesini ayarlayarak modelin iyi fitlenmesini sağlamaktadır (Pedregosa,2016, sy.739). Bu yüzden seçimi oldukça önemlidir. Yapılan hiperparametre optimizasyonu sonucu karşılaştırma istatistiklerine bakılarak rassel değişken sayısı(mtry) modeldeki toplam değişken sayısı olan 6’dan; optimum ağaç sayısı(ntrees) maksimum ağaç sayısı olan 500’den küçük olarak belirlenmiştir. Elde edilen hiperparametreler RO için [5;480] ve ARIMA-RO için [2;350]’dir.Tahmin başarısı, tahmin değerlerinin tahminde kullanılan veriye ne kadar yakınsadığını ölçmektedir. Bir şekilde parametrelerin optimum seçilip seçilmediğinin, en iyi modelin kurulup kurulmadığının bilgisini vermektedir. Öngörü başarısı ise kurulan bu modelin farklı bir veri setindeki başarısını ölçmektedir. Çalışmada 2010:01-2017:12 dönemi öğrenme veri seti olarak ve 2018:01-2018:10 dönemi test veri seti olarak alınmıştır. Tahmin sonuçlarını elde etme amacıyla öğrenme veri setine klasik ARIMA, Makine Öğrenmesi yöntemlerinden Rassel Orman ve hibrit yöntemlerden ARIMA-Rassel Orman olmak üzere 3 yöntem uygulanmıştır.Test veri seti statik öngörü için kullanılarak öngörü sonuçları elde edilmiştir. Tahmin ve öngörü başarısı karşılaştırma istatistikleri Tablo 3 ‘de sunulmuştur.

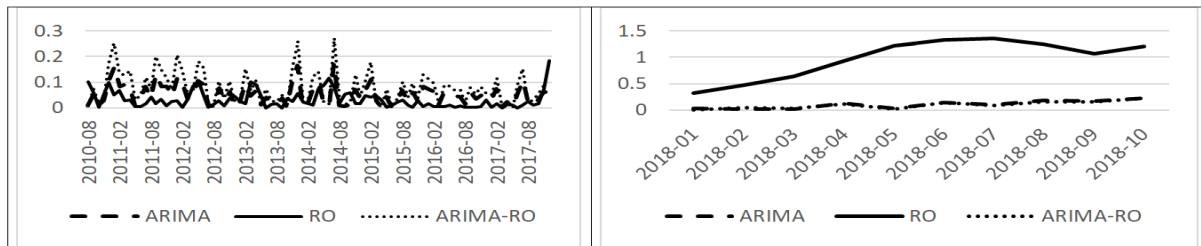
**Tablo 3: Tahmin ve Öngörü Performansları : ARIMA, Rassel Orman ve ARIMA –Rassel Orman**

| Metod    | Tahmin  |         | Öngörü  |         |
|----------|---------|---------|---------|---------|
|          | RMSE    | MAE     | RMSE    | MAE     |
| ARIMA-RO | 0.00421 | 0.00336 | 0.00573 | 0.00473 |
| RO       | 0.00207 | 0.00140 | 0.04900 | 0.04602 |
| ARIMA    | 0.00269 | 0.00216 | 0.00600 | 0.00490 |
| #N       | 90      |         | 10      |         |
| Toplam N | 100     |         | 100     |         |
| % N      | 90%     |         | 10%     |         |

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}, MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

Tahmin için kullanılan öğrenme veri setinde RO başarılı olsa da ARIMA-Rassel Orman Regresyonu örneklem dışı öngöründe daha iyi sonuçlar vermiştir. Tahmin performansı açısından değerlendirildiğinde ARIMA yöntemi Rassel Orman Regresyonundan sonra en iyi 2. performansa sahip yöntem olurken, konut fiyatları endeksi tahmininde ARIMA-RO yöntemi en kötü performansa sahip yöntem olmuştur.Grafik yöntemiyle karşılaştırma amacıyla nispi hatalar hesaplanmıştır. Nispi hatalar, her bir gerçek değere karşılık gelen mutlak tahmin hatasının yüzdesi olarak hesaplanmaktadır. Nispi hatalar yardımıyla üç modelin tahmin ve öngörü performansları karşılaştırılmış ve sonuçlar Grafik 1’de verilmiştir.  $Nispi\ Hata = \frac{\sum_{t=1}^n |y_t - \hat{y}_t|}{y_t} \times 100$  formülüyle hesaplanmıştır.

**Grafik 1: Nispi Hatalar ARIMA, Rassel Orman ve ARIMA –Rassel Orman**



Tahmin grafiği incelendiğinde Rassel Orman nispi hatalarının dağılımının sıfıra diğer iki yöntemden daha yakın olduğu görülmektedir. Nispi hatalara bakıldığında Rassel Orman yönteminin Türkiye’nin konut fiyatları endeksinin tahmininde en başarılı yöntem olduğu görülmektedir.Öngörü başarı performansı karşılaştırmasında MSE ve MAE ölçümlerine bakılarak, önerilen ARIMA-RO hibrit yönteminin açıkça diğer yöntemden daha üstün olduğunu görülmektedir. Öngörü başarıları karşılaştırmasında ARIMA yöntemi ARIMA-Rassel Orman Yönteminden sonra en iyi ikinci performansa sahip yöntem olurken, Türkiye konut fiyatları endeksinin öngörüsünde RO yöntemi en kötü performansa sahip yöntem olmuştur.

Gerçek değerlere karşılık gelen mutlak öngörü hatalarının yüzdesi alınarak nispi hatalar hesaplanarak elde edilen öngörü grafiği incelendiğinde ise ARIMA- Rassel Orman yönteminin nispi hata dağılımının sıfıra daha yakın olduğu görülmektedir. Öngörü performansı

karşılaştırmasında Nispi hatalara bakıldığında ARIMA-Rassal Orman yönteminin en başarılı yöntem olduğu görülmektedir. ARIMA yöntemi de ufak farkla ikinci en başarılı yöntemdir. Nispi hatalara göre öngörü performansı en düşük yöntemin ise Rassal Orman yöntemi olduğu görülmektedir. Elde edilen bulgular klasik ekonometrik yöntemler ile makine öğrenmesi yöntemlerinin birleştirilmesiyle elde edilen hibrit modelinin konut fiyat endeksinin öngörüsünde daha iyi performansa sahip olduğunu göstermektedir.

## 6.SONUÇ

Bu çalışmada Türkiye'nin konut fiyatları endeksinin tahmini amacıyla dinamik model yapısı kullanılmıştır. Çalışmanın hipotezleri ile klasik ekonometrik model tahmini için kullanılan ARIMA yöntemine karşılık Makine Öğrenmesi yöntemlerinden RO ve ARIMA-RO yöntemlerinin tahmin ve öngörü başarısında bir artışa sebep olup olmadığı araştırılmıştır. Çalışmadan elde edilen bulgular, ekonometrik ve finansal verilerin tahmin ve öngörüsünde her ne kadar doğrusal modellerin kullanımı yaygın olsa da Türkiye'nin konut fiyatları endeksi ile elde edilen dinamik modelin doğrusal olmaması nedeniyle ekonometrik modellemede eğrisel olarak ele alınması gerektiğini söylemektedir. Ayrıca ARIMA-RO gibi hem doğrusal hem de doğrusal olmama durumunu ele alan hibrit model yaklaşımının öngörü başarısını artırarak gerçeğe daha yakın tahminler verdiğini göstermektedir. Son dönemlerde Makine Öğrenmesi yöntemlerinin ekonometrik analizlerdeki kullanımı her ne kadar tartışma konusu olsa, yapılan çalışmalardan elde edilen olumlu sonuçlar makine öğrenmesi yöntemlerinin ekonometrik analizlerde giderek daha çok yer almasına olanak sağlamaktadır.

## KAYNAKLAR

- Franses, P. H., & Van Dijk, D. (2003). *Non-linear Time Series Models in Empirical Finance*. Cambridge University Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. New York: Springer-Verlag.
- Kumar, M., & Thenmozhi, M. (2014). Forecasting Stock Index Returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-Random Forest Hybrid Models. *Int. J. Banking, Accounting and Finance*, 284 – 308.
- Pai, P-F., & Lin, C-S. (2005). A Hybrid ARIMA and Support Vector Machines Model in Stock Price Forecasting. *Omega*, 497 – 505.
- Pedregosa, F. (2016). Hyperparameter Optimization With Approximate Gradient. *JMLR: W&CP vol.48*, 737-746.
- TCMB(2019, 27 Ekim). KonutFiyatEndeksi. <https://www.tcmb.gov.tr/wps/wcm/connect/b4628fa9-11a7-4426-ae6-dae67fc56200/KFE-Metaveri.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-b4628fa9-11a7-4426-ae6-dae67fc56200-mDXEz4N>
- Wang, P. (2008). *Financial Econometrics*. Routledge.
- Zhang, G. P. (2003). Time Series Forecasting using A Hybrid ARIMA and Neural Network Model. *Neurocomputing*, 159 – 175.
- Zhang, G., Patuwo, B. E., & Hu M. Y. (1998). Forecasting with Artificial Neural Networks: The State of The Art. *International Journal of Forecasting*, 35–62.