

Finansal Baloncukların Tespiti ve Tahminlenmesi*

Merve MERT SARITAŞ¹, Mert URAL²

Özet

Finansal balonların öngörülmesi, uygulamalı finans literatüründe önemli bir araştırma konusu olarak kabul görmektedir. Bu çalışmanın amacı, 2002M01-2024M06 dönemi boyunca BİST 100, BİST Banka ve BİST Sınai endeksindeki finansal balonları ekonometrik ve makine öğrenimi yöntemlerini birleştirerek tespit etmek ve öngörmektir. Bu doğrultuda, ilk olarak balon dönemlerinin tespitinde Genelleştirilmiş Supremum Artırılmış Dickey-Fuller (GSADF) testi kullanılmıştır. Bu test, zaman serilerindeki yapısal kırılmaları ve volatilité değişimlerini dikkate alarak balonların tespitinde daha güçlü bir yöntem sunmaktadır. İkinci olarak, balon oluşumunu makroekonomik göstergelerle ilişkilendirerek tahmin etmek için Rastgele Orman Ağaçları (Random Forest Algorithm) yöntemi kullanılmıştır. Rastgele Orman (Random Forest) algoritması ile balon oluşum olasılığını etkileyen finansal ve makroekonomik faktörlerin göreceli önemi belirlenmiştir. Analiz sonuçları, Türkiye'de BİST 100, BİST Banka ve BİST Sınai endeksinde fiyat balonlarının oluşumunda para arzı, döviz kuru, tüketici fiyat endeksi, FED faiz oranı, Bileşik Öncü Göstergeler Endeksinin belirleyici olduğunu göstermektedir. Bu doğrultuda, balon risklerini azaltmak ve finansal istikrarı sağlamak için para politikası ve makro ihtiyati tedbirlerin etkin kullanımının önemini ortaya koymaktadır. Aynı zamanda makine öğrenmesi yöntemlerinin finansal balonların erken tespiti ve piyasa risklerinin yönetimi açısından önemli bir potansiyel taşıdığını göstermektedir. Modelin performansını değerlendiren sonuçlar, önerilen yaklaşımın yüksek tahmin gücüne sahip olduğunu ve finansal balonların öngörülmesinde etkili bir alternatif olabileceğini göstermektedir.

Anahtar kelimeler: GSADF, Finansal Baloncuklar, Makine Öğrenimi Algoritmaları, Rastgele Orman Ağaçları, Erken Uyarı
Jel Kodu: C53, C58, G17, E60

Identifying and Forecasting of Financial Bubbles

Abstract

The forecasting of financial bubbles has been recognised as an important research topic in the applied finance literature. The objective of this study is to identify and predict financial bubbles in the BIST 100, BIST Bank, and BIST Industrials indices over the period 2002M01-2024M06 by combining econometric and machine learning methods. Accordingly, the Generalised Supremum Augmented Dickey-Fuller (GSADF) test is first used to identify bubble periods. This test provides a more robust method for detecting bubbles by taking into account structural breaks and volatility changes in time series. Subsequently, the Random Forest Algorithm, a powerful machine learning method, is used to establish a predictive model for bubble occurrences by establishing correlations with a set of macroeconomic indicators. Moreover, the Random Forest algorithm determines the relative importance of financial and macroeconomic factors that influence the probability of bubble formation. Empirical findings from the analysis highlight that broad money, exchange rates, CPI, FED interest rate, and Composite Leading Indicators Index are determinants of price bubble formation in BIST 100, BIST Bank, and BIST Industrials indices in Turkey. Accordingly, it shows the importance of effective use of monetary policy and macroprudential measures to mitigate bubble risks and ensure financial stability. It also shows that machine learning methods have significant potential for early detection of financial bubbles and management of market risks. The evaluation of model performance reveals a high degree of predictive accuracy, reinforcing the efficacy of the proposed hybrid approach as a viable alternative in the realm of financial bubble forecasting.

Keywords: GSADF, Financial Bubbles, Machine Learning Algorithms, Random Forest Algorithm, Early Warning
Jel Codes: C53, C58, G17, E60

*Bu çalışma, 6. Uluslararası Ekonomi, Finans ve Enerji Kongresi EFE'2024'te sunulan "Sağ Kuyruklu ADF ve Makine Öğrenmesi Yöntemlerine Dayalı Olarak Finansal Baloncukların Tespiti ve Tahmini" başlıklı bildiriden türetilmiştir.

ATIF ÖNERİSİ (APA): Mert Saritaş, M., & Ural, M. (2025). Identifying and Forecasting of Financial Bubbles Based on Right-Tailed ADF and Machine Learning Methods. *İzmir İktisat Dergisi*. 40(1). 272-293. Doi: 10.24988/ije.1554683

¹ Arş. Gör., Çanakkale Onsekiz Mart Üniversitesi, Siyasal Bilgiler Fakültesi, İktisat Bölümü, Merkez / Çanakkale, Türkiye, **EMAIL:** merve.mertsaritas@comu.edu.tr **ORCID:** 0009-0009-4549-1679

² Prof. Dr., Dokuz Eylül Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İktisat Bölümü, Buca / İzmir, Türkiye, **EMAIL:** mert.ural@deu.edu.tr **ORCID:** 0000-0003-3252-846X

1. GİRİŞ

Balon kelime anlamı olarak, rasyonellikten uzak sürü davranışı içeren pozitif veya negatif yönlü patlayıcı davranışlar olarak tanımlanabilir. Finansal piyasalardaki balon ise, bir ekonomik varlığın piyasa değerinin temel değerinden sürekli ve sistematik bir biçimde sapması olarak tanımlanmaktadır (Santoni, 1987). Brunnermeier'e (2016) göre balonlar, bir varlığın fiyatının temel değerinden daha yüksek olması durumunda ortaya çıkmaktadır. Temel değer belirlenimci arasında tahmini getiriler, varlığın tahmini nihai değeri ve iskonto oranı yer almaktadır. Dolayısıyla, bir varlığın cari fiyatı yüksek olduğunda ve temel faktörler bu fiyatı desteklemediğinde, bir balonun var olduğu söylenmektedir (Stiglitz, 1990).

Literatürde, bir finansal balon, varlık fiyatının içsel değerini aştığı durum olarak tanımlanmakta olup, erken uyarı sinyalleri olarak değerlendirilmeleri, politika yapımcılar ve düzenleyiciler için kritik öneme sahiptir. Finansal balonların öngörülmesi, uygulamalı finans literatüründe önemli bir araştırma konusu olarak kabul görmektedir. Varlık fiyatlarının temel değerini aşması durumunda ortaya çıkan balonlar, finansal sistemde rolü olan politika yapımcılar ve finansal düzenleyiciler için önemli riskler oluşturmaktadır (Case ve Shiller, 2003). Bu nedenle borsa verilerindeki anormalliklerin tespiti, finansal krizlerin başlangıcını belirlemek ve varlık fiyat balonlarını izlemek için önemli bir rol oynamaktadır. Bir varlık fiyat balonu, genellikle varlık fiyatlarının temel değerini aşması durumu olarak tanımlanır ve balonlar, piyasa fiyatları ile gerçek değerler arasındaki anormal fark olarak da düşünülebilir. Bu nedenle, erken uyarı sinyallerinden biri olarak balonların öngörülmesi, gelecekteki krizlere karşı önleyici tedbirler almakla sorumlu olan politika yapımcılar ve düzenleyiciler için hayati önem taşımaktadır. Bu nedenle, balon oluşumundaki ana faktörleri anlamak ve bunları erken aşamalarında tahmin etmek için artan sayıda çalışma yapılmaktadır.

Fiyatlarda meydana gelen balon dönemleri Phillips, Shi ve Yu (2015) çalışmasında geliştirilen özyinelemeli bir sağ kuyruklu Genelleştirilmiş Supremum Artırılmış Dickey Fuller Testi (Generalized Sup-Augmented Dickey Fuller-GSADF) tahminlenerek belirlenmektedir. Çalışmada kullanılan bir diğer yöntem olan makine öğrenmesi ise bilgisayarlara açıkça programlanmadan öğrenme yeteneği vermektedir. Makine öğrenmesi yöntemlerinden olan rastgele Orman algoritması ise büyük veri setlerindeki karmaşık örüntüleri tespit edebilme ve genelleme yeteneği sayesinde finansal balonların tahmininde güçlü bir yöntem olarak karşımıza çıkmaktadır.

Bu çalışmanın amacı, 2002-2024 dönemi boyunca BİST 100, BİST Banka ve BİST Sınai endeksindeki finansal balonları ekonometrik ve makine öğrenimi yöntemlerini birleştirerek tespit etmek ve öngörmektir. Çalışmanın ilk aşamasında, söz konusu BİST endekslerindeki balonların tespiti için sağ kuyruklu birim kök testleri kullanılmıştır. Bu testler, zaman serilerindeki yapısal kırılmaları ve oynaklık değişimlerini dikkate alarak balonların varlığını tespit etmede oldukça etkin bir yöntem sunmaktadır (Shi vd., 2011). Bu testler, balonların varlığını doğrulamak için temel bir adım olarak kabul edilmektedir, çünkü bu testler, fiyat serilerinde patlayıcı bir davranışın varlığını göstermekte ve böylece bir balonun varlığına dair ampirik kanıt sağlamaktadır.

Borsa endekslerinin, makroekonomik göstergeler, sosyo-politik değişimler, yatırımcı tercihleri, çeşitli beklenti ve güven endeksleri gibi çok sayıda farklı faktörden etkilenmesi, endeks tahminlemesini oldukça güçleştirmektedir (Liu vd., 2012). Günümüzde tahminleme modellerinde makine öğrenme algoritmaları oldukça sık kullanılmakta ve gayet başarılı sonuçlar elde edilmektedir. Buna istinaden çalışmanın ikinci aşamasında ise, balon oluşumunu makroekonomik göstergelerle ilişkilendirerek tahmin etmek için Rastgele Orman (Random Forest) Algoritması kullanılmıştır. Tahminleme modeli için hedef değişkeni etkileyen parametrelerden bir öznelik kümesi oluşturulmuş ve makine öğrenme algoritmaları ile hedef veri tahmin edilmeye çalışılmıştır. BIST 100, BIST Banka ve BIST Sınai endeksinin hem makroekonomik hem de finansal göstergelerden etkilendiği düşünülerek, bu göstergelerden oluşan bir öznelik kümesi oluşturularak makine

öğrenmesi algoritmaları ile hedef değişken olan endekslerdeki balonlar tahmin edilmeye çalışılmıştır.

Çalışmanın ikinci bölümünde literatür araştırması yapılmış, makalede kullanılacak olan makine öğrenmesi tabanlı balon tahmini yönteminin literatürdeki yerini ve önemi açıkça gösterilmiştir. Üçüncü bölümünde GSADF ve Rastgele Orman Algoritması hakkında temel bilgiler verilmiştir. Dördüncü bölümde belirlenen endeks fiyatlarında GSADF ve Rastgele orman yönteminin analiz bulguları açıklanmış, model ölçüm performanslarına detaylı bir şekilde değinilmiştir. Beşinci bölümde ise sonuç ve değerlendirmeye yer verilmiş ve çalışmanın kapsamını ve temel araştırma sorularını daha iyi anlamayı sağlayan kapsamlı bir bakış açısı da sunulmuştur. Balonların tespiti ve analizi, finansal istikrarın sağlanması açısından kritik bir öneme sahip olduğundan, bu çalışmada, Borsa İstanbul hisse senedi piyasasındaki balonların öngörülmesi amacıyla, gerçek zamanlı balon tespit testi ve rastgele orman ağaçları (RFA) temelli iki aşamalı bir makine öğrenmesi modeli önerilmektedir.

Bu çalışmanın literatüre katkısı, borsa endeks fiyatlarında meydana gelen balonları tespit ederek, balon oluşumlarına etki eden makroekonomik faktörlerin ilişkisini açıklamak ve bu ilişkiler çerçevesinde balon dönemlerinin makine öğrenimi yöntemleriyle tahmin edilmesini sağlamaktır. Bu bağlamda, makroekonomik değişkenlerin dinamik etkileşimleri incelenerek, piyasa davranışlarının daha iyi anlaşılması hedeflenmektedir.

2. LİTERATÜR ARAŞTIRMASI

Borsa verilerindeki anormalliklerin tespit edilmesi, finansal krizlerin genellikle varlık fiyatlarındaki bir patlama ve çöküş döngüsünü, yani varlık fiyat balonlarını takip etmesi nedeniyle çok önemli hale gelmiştir. Balon genellikle varlık fiyatlarının temel değerini aştığı bir durum olarak tanımlanır. Gilles ve Leroy (1992) balonları piyasa fiyatları ile gerçek değerler arasındaki anormal fark olarak tanımlamıştır. Teorik bir bakış açısıyla, herhangi bir varlığın gerçek veya temel değeri, gelecekteki tüm nakit akışlarının gerekli getiri oranıyla iskonto edilmiş bugünkü değerlerinin toplamına eşittir. Tarih, balonların öngörülmesinin, finansal sistemin refahından sorumlu olan politika yapıcılar ve finansal düzenleyiciler için her zaman hayati önem taşıdığını göstermektedir. Bundan sonra, literatürde giderek artan sayıda çalışma, krizlere yol açan anormal borsa davranışlarını tespit etmek için erken uyarı sistemleri geliştirmeye çalışmıştır.

Literatürde Borsa İstanbul endeks fiyatlarını etkileyen faktörlerin belirlenmesi ve tahmin edilmesi amacıyla yapılan pek çok çalışma bulunmakla beraber, söz konusu çalışmalarda farklı teknikler kullanılmıştır. Chang ve arkadaşları (2014), 1990-2003 döneminde BRICS ülkelerinde çoklu finansal balon oluşumlarının varlığını SADF ve GSADF testlerini kullanarak kapsamlı bir şekilde incelemişlerdir. Bu çalışma, gelişmekte olan piyasalarda finansal balonların yaygınlığına dair önemli bir kanıt sunmaktadır. Phillips vd. (2015a), 1871'den 2010'a kadar S&P 500 endeksine hem SADF hem de GSADF testlerini uygulamışlar ve GSADF'nin iki balon dönemini başarıyla tespit ettiğini ortaya koymuşlardır: 1873 Paniği (Ekim 1879'dan Nisan 1880'e kadar) ve dot-com balonu (Temmuz 1997'den Ağustos 2001'e kadar). Popüler bir gerçek zamanlı balon tespit testi olan GSADF Phillips, Shi ve Yu tarafından 2015 yılında geliştirilmiştir. O zamandan beri krizler için bir erken uyarı aracı olarak yaygın bir şekilde kullanılmaktadır. Test, Augmented Dickey-Fuller (ADF) model spesifikasyonuna dayanmakta ve birden fazla balonu tespit etmede ve bunların başlangıç ve bitiş tarihlerini belirlemede olağanüstü bir performans elde eden özyinelemeli bir algoritma kullanmaktadır (Phillips vd., 2015b). Genelleştirilmiş Sup ADF (GSADF) testi, fiyat balonlarını tanımlamak ve finansal zaman serisi verilerindeki patlayıcı davranışları tespit etmek için güçlü bir araçtır. Bu test, zaman içinde birden fazla balonun tanımlanmasını sağlayarak hisse senedi fiyatlarının dinamiklerini anlama noktasında çok önemli bir rol oynamaktadır. GSADF testinin, çeşitli hisse senedi piyasalarında uygulanması geçmiş ve gelecekteki fiyat düzeltmelerine işaret edebilecek

fiyat balonlarını tespit etmedeki etkinliği vurgulanmıştır (Zeren ve Yılandı, 2019). Bu özellik, özellikle balon dinamiklerini anlayarak yatırım stratejilerini şekillendirebileceğinden Borsa İstanbul'da yıldız pazar ve ana pazarda işlem gören büyük şirketlere ait hisselerden oluşan XU100, XBANK ve XUSIN endeksleri için büyük önem taşımaktadır.

Korkmaz vd. (2016), Borsa İstanbul 100 Endeksi getiri oynaklığı üzerinde altın, döviz kuru ve mevduat faiz oranı gibi alternatif yatırım enstrümanlarındaki spekülasyon fiyat artışlarının etkilerini incelemiştir. Ocak 2002-Mayıs 2016 dönemini kapsayan veri setini kullandıkları çalışmada, spekülasyon fiyat artışlarının tespiti için SADF ve GSADF testleri uygulanmış olup, dolar kurunda gözlemlenen fiyat balonlarının Borsa İstanbul 100 Endeksi oynaklığını artırdığı, ancak altın fiyatlarındaki balonların ise Borsa İstanbul 100 Endeksi oynaklığını azalttığı tespit edilmiştir. Akkaya (2018), Borsa İstanbul 100 Endeksi'ndeki fiyat balonlarının belirleyicilerini Ocak 2002-Mart 2017 dönemi için incelemiştir. RTADF testi kullanılarak belirlenen balon dönemlerinde, yurtdışı yerleşiklerin hisse senedi portföyü ile Borsa İstanbul 100 Endeksi balonları arasında istatistiksel olarak anlamlı bir ilişki tespit edilmiştir. Ayrıca, reel döviz kuru, ihracat, yurtdışı yerleşiklerin hisse senedi portföyü ve ABD doları aylık faiz oranından balonlara doğru nedensellik ilişkisi gözlemlenmiştir.

Yukarıda özetlenen bulgulara dayanarak, GSADF testinin piyasa balonlarının varlığını tespit etmek için etkili bir yaklaşım olduğu açıktır. Dolayısıyla, bu çalışmada XU100, XBANK ve XUSIN endeks fiyat balonlarının varlığını tespit etmek için GSADF testi kullanılmıştır.

Makine öğrenimi algoritmalarının geleneksel zaman serisi analiz yöntemleriyle entegrasyonu, finansal balonların tespit ve tahmin süreçlerine yeni bir boyut kazandırmıştır. Özellikle, makine öğrenimi modellerinin GSADF testi gibi yapısal kırılma noktalarını tespit eden yöntemlerle birlikte kullanılması, daha güçlü ve kapsamlı analizler yapma imkânı sunmaktadır. Örneğin, makine öğreniminin GSADF ile entegrasyonu, geçmiş fiyat verilerini analiz eden ve kabarcık oluşumunu gösteren kalıpları belirleyen öngörücü modellerin geliştirilmesini kolaylaştırabilir. Ji vd. (2021) tarafından gerçekleştirilen çalışmada, GSADF modelinin Geri Yayılım Ağı (BP) ile birleştirilerek XU100 gibi finansal endekslerdeki balon oluşumlarının daha doğru bir şekilde tahmin edildiği gösterilmiştir. Bu yaklaşım, geçmiş fiyat verilerindeki karmaşık ve doğrusal olmayan ilişkileri yakalayarak, geleneksel istatistiksel yöntemlerin sınırlamalarını aşmaktadır. Makine öğrenimi algoritmalarının GSADF testi ile kullanımı, finansal balonların başlangıç ve bitiş zamanlarını daha kesin olarak belirlemek ve bu süreçlerin dinamiklerini daha iyi anlamak için değerli bir araç haline gelmektedir.

Finansal piyasaların karmaşık ve doğrusal olmayan yapısı, hisse senedi fiyat tahminlerini zorlu bir hale getirmektedir. Bu bağlamda, Rastgele Orman algoritması, büyük veri kümelerini işleme ve karmaşık ilişkileri yakalama yeteneği sayesinde hisse senedi fiyat tahmininde önemli bir potansiyel sunmaktadır. Çoklu karar ağacının bir araya getirilmesiyle oluşturulan bu yöntem, aşırı uyum sorununu en aza indirerek daha genellenebilir tahminler elde etmeyi mümkün kılmaktadır (Sun & Zhao, 2020). Literatürdeki karşılaştırmalı çalışmalar, Rastgele Orman algoritmasının, doğrusal regresyon ve destek vektör makineleri gibi geleneksel yöntemlere kıyasla hisse senedi fiyat tahminlerinde daha başarılı olduğunu göstermektedir (Nikou vd., 2019; Rouf vd., 2021). Rastgele Orman algoritmasının performansı, hiper parametre optimizasyonu ile daha da iyileştirilebilir. Özellikle, GSADF testi ile belirlenen balon dönemlerini içeren veri kümelerinde hiper parametre optimizasyonu, modelin piyasa dinamiklerine olan duyarlılığını artırarak daha doğru tahminler elde edilmesini sağlayabilir (Hoque & Aljamaan, 2021). Bu sayede, Rastgele Orman algoritması, finansal piyasalardaki yapısal kırılmaları daha iyi modelleyerek yatırım kararlarının alınmasında önemli bir araç haline gelebilir. Tran vd., (2023) 2001'den 2021'e kadar Vietnam borsasındaki finansal balonları tespit ve tahmin etmek için, sağ kuyruklu birim kök testini içeren PSY prosedürünü kullanmışlardır. Daha sonra gerçek zamanlı finansal balon olaylarını tahmin etmek için makine öğrenimi

algoritmaları kullanılmışlar ve 2006-2007 ve 2017-2018 dönemlerinde Vietnam borsasında finansal balonların varlığını ortaya koymuştur. Ek olarak, ampirik kanıtlar, Vietnam borsasındaki finansal balonları tahmin etmede rastgele orman ve yapay sinir ağı algoritmalarının geleneksel istatistiksel yöntemlere göre üstün performansını desteklemiştir.

Makroekonomide, hisse senedi piyasasında finansal balonları tespit etmek için makine öğrenmesinin kullanımı nispeten yeni bir yaklaşımdır. Finansal krizler konusunda yapılan çoğu çalışma, finansal balonları tahmin etmek yerine bankacılık, menkul kıymet piyasaları ve kamu borcu gibi alanlarda genel ekonomik krizleri tahmin etmeye odaklanmıştır. Bugüne kadar (Başoğlu Kabran & Ünlü, 2021) S&P 500 hisse senedi piyasasındaki balonları, gerçek zamanlı balon tespit testi ve destek vektör makinesi (DVM) kullanan iki aşamalı bir makine öğrenimi yaklaşımıyla tahmin etmiştir. İki adımdan oluşan analizde önce sağ kuyruklu birim kök testi kullanılarak S&P 500 endeksindeki balonların tespit edilmiş, daha sonra ise balonları makroekonomik göstergelere göre tahmin etmek için DVM kullanılmıştır. Sonuçlar, yüksek tahmin gücüne sahip önerilen yaklaşımın balon tahmininde uygun bir alternatif olabileceğini belirtmişlerdir.

Makroekonomik değişkenlerin gelişmekte olan ülkelerdeki hisse senedi piyasaları üzerindeki etkileri de araştırmacılar arasında ilgi görmüştür. Finansal balonlar üzerinde etkili olan makroekonomik faktörleri inceleyen Sağlam Bezgin ve Başar (2020) çalışmalarında, Türkiye’de BIST 100 endeksi üzerinden 1997-2018 yılları arasında finansal balon oluşumlarını ve bu oluşumları etkileyen makroekonomik faktörleri incelemişlerdir. GSADF testi ile Ocak 2000, Aralık 2007 ve Şubat 2018 yıllarında önemli finansal balonların oluştuğu tespit edilmiştir. Bu balonların oluşumunda faiz oranı, M2 para arzı ve kredi hacmi gibi makroekonomik değişkenlerin etkili olduğu, Hatemi-J asimetrik nedensellik testi ile ortaya konmuştur. Bir diğer çalışmada, para politikasının varlık fiyat balonları üzerindeki etkisi deneysel ortamlarda araştırılmış ve faiz oranı ayarlamaları gibi proaktif para politikası önlemlerinin, yatırımcı beklentilerini ve davranışlarını etkileyerek büyük balonların oluşumunu azaltabileceği bulunmuştur (Hennequin & Hommes, 2023). Bu içgörü, stratejik müdahalelerin fiyatları istikrara kavuşturabileceğini gösterdiğinden, XU100, XBANK ve XUSIN gibi endeksleri izleyen finansal kurumlar ve politika yapımcılar için özellikle önemlidir.

Bu araştırma, finansal balonların tespiti ve tahmini alanında çığır açan makine öğrenimi tekniklerinin metodolojik üstünlüklerinden yararlanarak, mevcut literatüre önemli katkı sağlamaktadır. Geleneksel istatistiksel yöntemlerin aksine, makine öğrenmesi yöntemleri çeşitli avantajlar sunmaktadır. Ayrıca, mevcut araştırmaların makro göstergeler ile hisse senedi getirileri arasındaki doğrusal ilişkileri incelemiş olması nedeniyle, bu çalışma, ilgili makroekonomik değişkenlerin BIST endeks getirileri üzerindeki doğrusal olmayan ve karmaşık kalıplarını parametrik olmayan bir biçimde ele alarak literatüre değerli bir katkıda bulunmaktadır. Çalışmamız, bu çalışmayı birkaç boyutta geliştirmektedir. İlk olarak, GSADF testi ile birlikte Rastgele orman algoritmalarını birlikte uygulamaktadır. İkinci olarak, bu çalışma günlük veriler yerine, daha geniş bir aralıkta 2002:01 ile 2024:06 arasında aylık veriler kullanmaktadır. Son olarak, çalışmamız, sepet kur, TÜFE, ÜFE, çekirdek TÜFE, para arzı, FED faiz oranı, uluslararası rezervler, bileşik öncü göstergeler endeksi, Türkiye için jeopolitik risk endeksi, mevduat ve kredi faizleri gibi makroekonomik değişkenleri içererek makro bir bakış açısıyla analiz yürütmektedir. Bu sayede, çalışmamız, finansal balonların oluşum mekanizmalarına dair daha derinlemesine bir anlayış sunarak, politika yapımcılara ve yatırımcılara önemli bilgi ve içgörüler sağlamaktadır.

3. YÖNTEM

3.1 GSADF Testi

Phillips vd. (2015a), SADF yönteminin sınırlamalarını aşmak için, PSY prosedürü olarak da adlandırılan Genelleştirilmiş Sup ADF (GSADF) testini geliştirmiştir. GSADF testi, örnek dizilerde patlayıcı kalıpları tespit etmeyi amaçlayan yuvarlanan pencere SADF testine dayalı, sağ kuyruklu ADF testinin yinelemeli bir uygulamasıdır. SADF'ye kıyasla, GSADF yuvarlanan pencereler açısından daha esnektir ve bu da fiyat patlaması davranışını incelemek ve piyasa balonlarının varlığını doğrulamak için değerli bir araç haline getirmektedir. Borsa İstanbul endeks fiyat serilerindeki balonları tespit edebilmek amacıyla Genelleştirilmiş Sup ADF (GSADF) testi kullanılmıştır. Genelleştirilmiş Supremum Dickey Fuller (GSADF) testi, toplam gözlem sayısının alt örneklemi için özyinelemeli olacak şekilde tekrarlanan ADF testi regresyonlarına dayanmaktadır (Ural, 2022). GSADF test istatistiğinin elde edilmesi için, ADF test regresyonu aşağıdaki şekilde tahmin edilmektedir:

$$\Delta y_t = \hat{\alpha}_{r_1, r_2} + \hat{\beta}_{r_1, r_2} y_{t-1} + \sum_{i=1}^k \hat{\psi}_{r_1, r_2}^i \Delta y_{t-i} + \hat{\varepsilon}_t \quad (1)$$

Burada y_t , t zamanındaki ilgili borsa endeks fiyatlarını, Δ ise birinci fark operatörünü ifade etmektedir. Olası bir otokorelasyonu önlemek için $i = 1, 2, \dots, k$, Δy 'nin gecikmeleri denklemin sağ tarafına dahil edilmiştir. r_1 ve r_2 alt simgeleri, alt örneklemin başlangıç ve bitiş noktalarını göstermektedir (Phillips vd., 2015a).

Birim kökün sıfır hipotezi ($H_0: \beta_{r_1, r_2} = 0$) patlayıcı davranış alternatifine ($H_1: \beta_{r_1, r_2} > 0$) karşı $ADF_{r_1}^{r_2} = \hat{\beta}_{r_1, r_2} / s.e.(\hat{\beta}_{r_1, r_2})$ istatistiği kullanılarak test edilmektedir. Burada s.e. eğim katsayısının standart hatasını göstermektedir.

GSADF test istatistiğini elde etmek için hem r_1 hem de r_2 'yi değiştirerek tüm alt örneklem için $ADF_{r_1}^{r_2}$ test istatistikleri hesaplanmaktadır:

$$GSADF(r_0) = \sup_{r_2 \in [r_0, 1], r_1 \in [0, r_2 - r_0]} ADF_{r_1}^{r_2} \quad (2)$$

Burada r_0 , $r_0 = 0.01 + 1.8/\sqrt{T}$ şeklinde elde edilen başlangıç alt örneklem boyutunu, T ise gözlem sayısını göstermektedir, yani ilk gözlem 0 ile $r_2 - r_0$ arasında değişirken, son gözlem r_0 ile 1 arasında değişmektedir. Kritik değerleri elde etmek için Phillips ve Shi'nin (2020) önerisini izleyerek bootstrap simülasyonları kullanılmıştır. GSADF test istatistiği, her bir gözlem için 1000 adet Bootstrap ve Monte Carlo simülasyonu kullanılarak elde edilmiş olup, %95 güven düzeyinde kritik değerler dikkate alınmıştır. Sıfır hipotezinin reddedilmesi, seride en az bir balon olduğunu göstermektedir.

Balon varlığı tespiti sonrasında ise balon oluşum dönemleri geriye dönük SADF (Backwards sup ADF-BSADF) istatistik dizileri yardımıyla belirlenmektedir. BSADF istatistik dizileri geriye doğru (backwards) genişleyen yapıdaki örneklem için sağ kuyruklu ADF testleri kullanılarak elde edilmektedir (Ural, 2022). BSADF istatistiği ve buna bağlı olarak GSADF istatistiği sırasıyla aşağıdaki gibi yazılabilir (Caspi, 2017):

$$\{BSADF_{r_2}(r_0)\}_{r_2 \in [r_0, 1]} \quad (3)$$

$$GSADF(r_0) = \sup_{r_2 \in [r_0, 1]} \{BSADF_{r_2}(r_0)\} \quad (4)$$

Oluşturulan BSADF dizileri ile Monte Carlo simülasyonu ile hesaplanan sağ kuyruklu kritik değerler karşılaştırılarak baloncukların oluşum tarihleri tespit edilmektedir.

3.2 Rastgele Orman (Random Forest Algorithm)

Rastgele Orman algoritmasının birçok çeşidi vardır. Ancak günümüzde kullanılan en yaygın olanı Breiman (2001) tarafından geliştirilen rastgele orman (random forest) algoritması, karar ağacı (decision tree) algoritmalarından oluşan bir topluluk öğrenmesi (ensemble learning) yöntemi olarak tanımlanmaktadır. Denetimli bir makine öğrenimi algoritması olan rastgele orman birden fazla sınıflandırıcıyı birleştiren bir teknik olan topluluk öğrenimini kullanır. Temel olarak rastgele orman yönteminin amacı, tek bir karar ağacı üretmek yerine her biri farklı eğitim setlerinde ele alınmış çok sayıda ağacın kararlarını birleştirmektir (Çağlayan Akay, 2020). Rastgele Orman algoritması, makine öğrenimi alanında, özellikle sınıflandırma ve regresyon problemlerinde sıklıkla tercih edilen güçlü bir yöntemdir. Avantajlarına bakıldığında, algoritmanın başarısı hem örneklem hem de özellik seçimi aşamalarında uyguladığı rastgelelik ilkesine dayanmaktadır. Bu sayede, aşırı öğrenme problemi minimize edilir ve modelin genelleme yeteneği artırılır. Ayrıca, algoritmanın çoklu karar ağaçlarından oluşması, modelin gürültülü verilere karşı daha dayanıklı olmasını sağlar. Özellik önemi analizinde de kullanılabilen bu yöntem, yüksek boyutlu veri setlerinde bile başarılı sonuçlar verir. Dahası, algoritmanın ölçeklenebilirliği, büyük veri kümelerindeki uygulamaları için oldukça avantajlıdır. Son olarak, Rastgele Orman algoritmasının hem sayısal hem de kategorik verilerle çalışabilme özelliği, uygulamaların çeşitliliğini artırır. Bu özellikler sayesinde, Rastgele Orman algoritması, birçok farklı alanda başarılı bir şekilde kullanılmaktadır.

Rastgele Orman algoritmasının sunduğu avantajlara karşın, bazı dezavantajları da bulunmaktadır. Algoritmanın en belirgin dezavantajı, yüksek hesaplama maliyetidir. Çok sayıda karar ağacından oluşan bir ormanın eğitilmesi ve yeni bir veri noktası için tahmin yapılması, özellikle büyük veri setlerinde zaman alıcı olabilir. Aşırı uyum (overfitting) problemlerine karşı tamamen bağımsız olmamak da Rastgele Orman algoritmasının dezavantajları arasında sayılabilir. Özellikle küçük veri setlerinde veya aşırı karmaşık bir modelde, algoritma eğitim verilerine aşırı uyum sağlayarak yeni verilere genelleme yeteneğini kaybedebilir. Fakat ağaç sayısı yeterli ölçüde ise aşırı uyum sorunu çözümlenebilecektir.

Rastgele orman sınıflandırıcısında model, verilen veri kümesinden öğrenmekte ve ardından yeni verileri eğitime dayalı olarak sınıflara veya gruplara ayırmaktadır. Evet veya hayır, 0 veya 1, “Balon var” veya “Balon Yok” gibi sınıf etiketlerini çıktı olarak tahmin etmektedir. Sınıflandırma modellemesinde, değerlendirme ölçütleri Doğru Pozitifler (DP), Yanlış Pozitifler (YP), Doğru Negatifler (DN) ve Yanlış Negatifler (YN) üzerinden hesaplanmaktadır ve sınıflandırma modelinin performansını değerlendirmek için birbirinden farklı ölçütler kullanılmaktadır.

4. UYGULAMA - AMPİRİK BULGULAR

4.1 Balonların Tespiti

Tablo 1: Dışsal Değişkenlerin Kategorileri

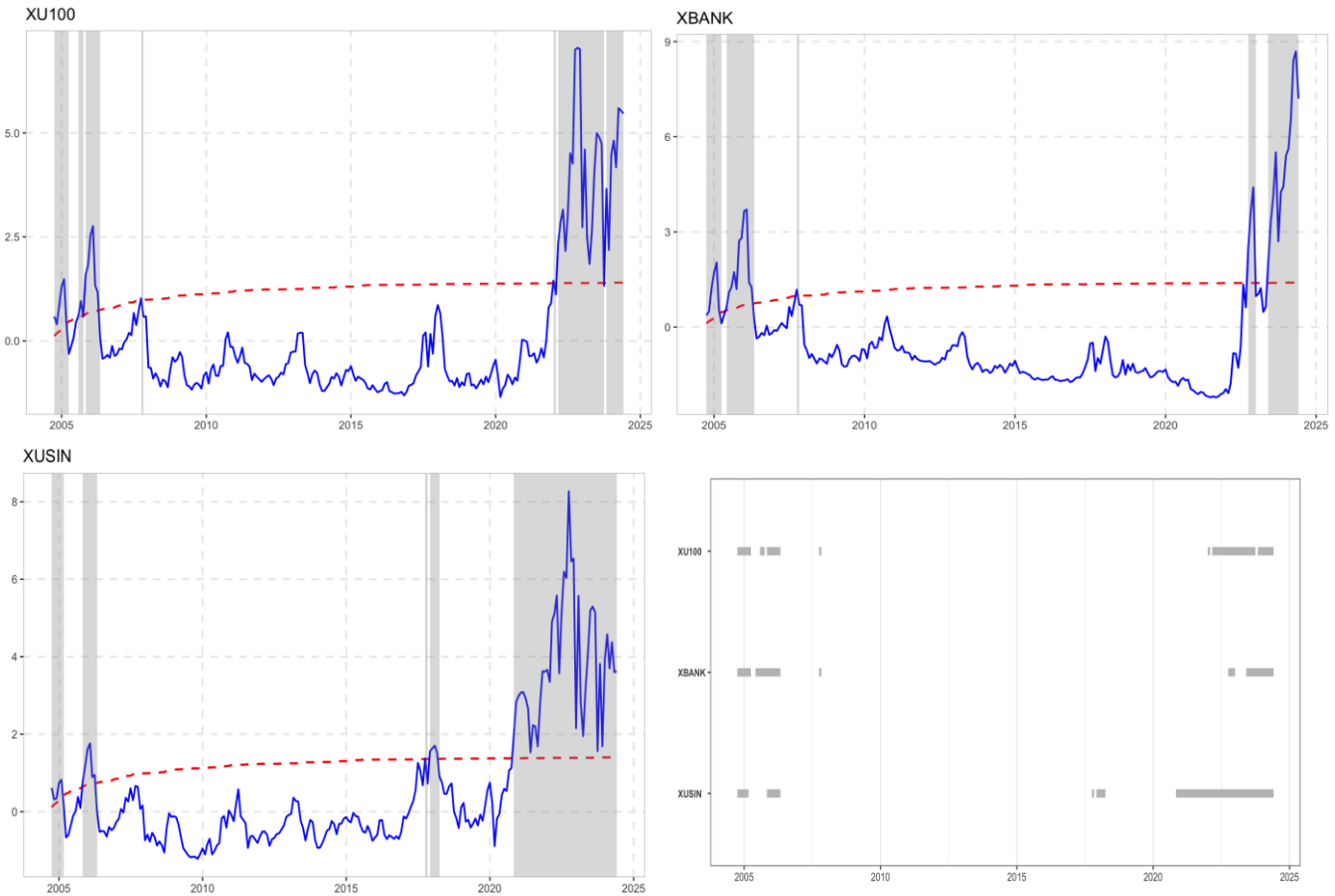
<i>Dışsal Değişkenlerin Kategorileri</i>	
<i>Makroekonomik Göstergeler</i>	<i>Finansal Göstergeler</i>
Sepet Kur	Mevduat Faizi
Uluslararası Rezervler	Kredi Faizi
M2 Para Arzı	GPRC_TUR
BOGE	FEDFUNDS
TÜFE	
Cektüfe	
ÜFE	

Çalışmanın ilk aşamasında Borsa İstanbul 100 endeksi (BIST100), Bankalar (XBANK) ve Sınai (XUSIN) fiyatlarında 2002 yılından itibaren spekülasyon baloncuklarının varlığı ve balon oluşum tarihleri, Phillips, Shi ve Yu (2015) çalışmasında geliştirilen sağ kuyruklu birim kök testi olan GSADF ile tespit etmek amaçlanmıştır. Çalışmada 2002M01-2024M06 tarihleri arasındaki aylık veriler incelenmiş ve veriler investing.com veri tabanından alınmıştır.

Tablo 1’de yer alan Tüketici fiyatları endeksi, Çekirdek TÜFE, üretici fiyatları endeksi, uluslararası rezervler, kredi faizi, TL üzerinden açılan mevduatlara uygulanan faiz EVDS’den alınmıştır. Türkiye Cumhuriyet Merkez Bankası (TCMB) ile Ekonomik Kalkınma ve İşbirliği Örgütü (OECD) ortak çalışması sonucunda oluşturulan ve 2002 tarihinden bu yana aylık olarak yayınlanan bileşik öncü göstergeler endeksi (BOGE) ise yine EVDS’den temin edilmiştir. Sepet Kur, USD/TL Döviz Kuru ve EUR/TL Kuru verileri ile ortalama alınarak yazar tarafından oluşturulmuştur. M2 para arzı International Financial Statistics (IFS), ABD politika faiz oranı Federal Reserve Economic Data veri tabanları kullanılarak elde edilmiş, ilgili ülkeye ait jeopolitik riskleri ölçen Jeopolitik Risk Endeksi (GPRC) ise <https://www.policyuncertainty.com/gpr.html> sitesinden Türkiye için alınmıştır.

BIST 100, BIST Sınai ve BIST Banka Fiyat Endeksi'ne benzer şekilde tüm dışsal değişkenler aylık olarak kullanılmıştır. Tablo 2, GSADF test istatistiklerini ve bootstrap kritik değerlerini göstermektedir. Tablo 2’de yer alan test istatistikleri incelendiğinde %1 önem seviyesinde BIST Pay endekslerinde finansal balon bulunmadığını belirten H_0 hipotezi reddedilmektedir. GSADF test sonuçları genel olarak değerlendirildiğinde XU100, XBANK ve XUSIN endekslerinde spekülasyon finansal balonlarının varlığı tespit edilmiştir.

Şekil 1: Endeksler için GSADF test sonuçları ve balon dönemlerinin toplu gösterimi



Açıklama: Gölgeli alanlar balon dönemlerini temsil etmektedir.

Aynı zamanda GSADF testini kullanarak endekslerdeki fiyat balonu başlangıç ve bitiş tarihleri tespit edilmiştir. Şekil 1 her endeks için balon dönemlerini grafik halinde gösterilmektedir.

Tablo 2: GSADF Test İstatistikleri ve Kritik Değerler

Değişken	Test İstatistiği	Kritik Değerler			Fiyat Balonu
		%1	%5	%10	
XU100	7,05	2,47	2,06	1,82	Var
XBANK	8,69	2,08	1,36	1,1	Var
XUSIN	8,27	2,47	2,06	1,82	Var

4.2 Rastgele Orman Sonuçları

İncelenen dönemde ilgili endeks fiyatlarında balon olup olmadığını öngörebilmek için veri seti öncelikle eğitim ve test verisi olarak rastgele biçimde ikiye bölünmüştür. Eğitim verisi, tüm veri setinin rastgele seçilmiş %80'i olarak belirlenirken diğer %20'si ise test verisi olarak ayrılmıştır. Bağımlı değişkenin Borsa endekslerinde balon olup olmadığını "Balon Var" ve "Balon Yok" şeklinde iki kategori ile temsil ettiği sınıflandırma ağacı Breiman vd. (1984) tarafından ortaya konulan yöntemsel çerçeveye uygun bir şekilde oluşturulmuştur.

4.2.1 Rastgele Orman Performans Ölçütleri (Başarısı)

Makine öğrenmesinde, değerlendirme ölçütleri, modellerin performansını anlamak ve karşılaştırmak açısından kritik bir öneme sahiptir. Rastgele Orman Algoritmasının performansını değerlendirmek için birbirinden farklı ölçütler kullanılmıştır. Bunlar "Doğruluk Oranı", "Kesinlik", "Duyarlılık", "Özgüllük", "F1 Puanı" ve "AUC Değeri"dir. Örneklemimizi her bir endeks için öğrenme örneklemini ve test örneklemini olarak iki kategoriye ayırdığımızda, rastgele orman algoritması birçok ağaç modeli oluşturmakta ve fiyat balonunun oluşumuna dair bir tahminde bulunmaktadır; ardından ise tahmin edilen sonuç ile test örneklemini karşılaştırmaktadır. Dolayısıyla bu ölçütler analizde rastgele orman modelinin tahmin kalitesini temsil etmektedir. Bu ölçütlerin formülasyonları şu şekilde verilmektedir (Uğuz, 2019):

$$\begin{aligned} \text{Doğruluk} &= \frac{DP + DN}{DP + DN + YP + YN} & \text{Kesinlik} &= \frac{DP}{DP + YP} & \text{Duyarlılık} &= \frac{DP}{DP + YN} \\ \text{Özgüllük} &= \frac{DN}{DN + YP} & \text{F1 puanı} &= 2 \cdot \frac{\text{Duyarlılık} \cdot \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \end{aligned}$$

Tablo 3: Tahmine dayalı performans ölçütleri

Ölçüt	XU100	XBANK	XUSIN
<i>Doğruluk</i>	0,963	0,926	0,926
<i>Kesinlik</i>	0,818	0,800	0,900
<i>Duyarlılık</i>	1,000	0,571	0,750
<i>Özgüllük</i>	0,956	0,979	0,976
<i>F1 Puanı</i>	0,900	0,667	0,818
<i>AUC</i>	0,974	0,985	0,986

Bu ölçütler kullanılarak oluşturulan Tablo 3, her bir endeks için rastgele orman modelinin tahmin gücünü göstermektedir. Aynı zamanda rastgele orman algoritmasının Borsa İstanbul endeksindeki fiyat balonlarına katkıda bulunan potansiyel faktörlerin iyi performans gösterdiğini ortaya koymaktadır.

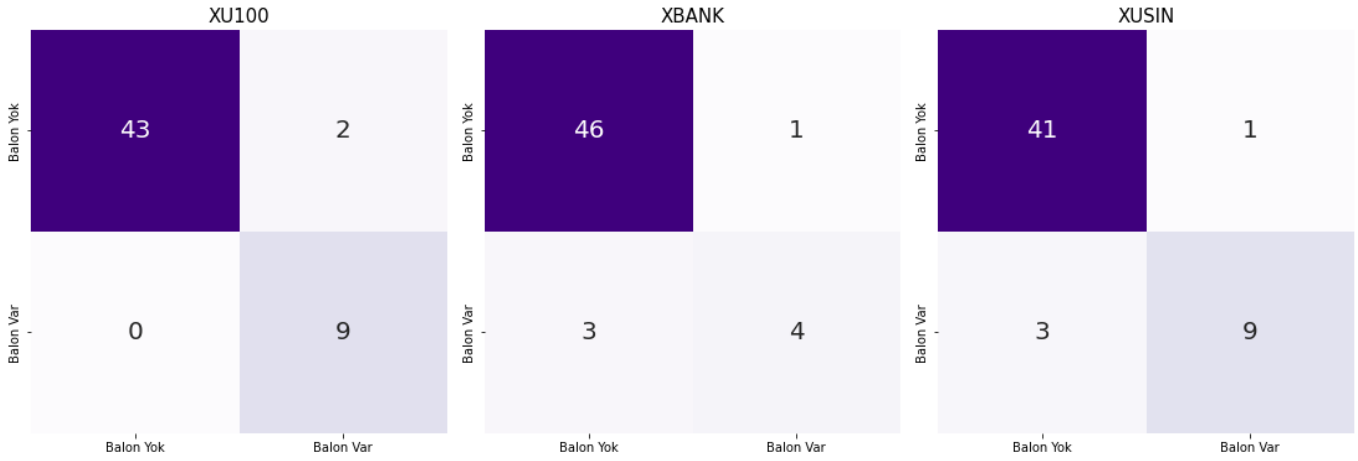
Hata matrisi, bir sınıflandırıcının iki farklı sınıf etiketlerini ne ölçüde sınıflandırabildiği ile ilgili performansını gösteren bir analiz aracıdır. Doğru pozitif (DP) ve doğru negatif (DN) doğru

tahminlere karşılık gelirken, yanlış pozitif (YP) ve yanlış negatif (YN) yanlış tahminlere karşılık gelir (Metz, 1978).

Tablo 4: BIST Endeksleri erken uyarısı için karışıklık matrisi

Tahmin Edilen	0: Balon Yok	1: Balon Var
Gerçek		
0: Balon Yok	Doğru Negatif (DN)	Yanlış Pozitif (YP)
1: Balon Var	Yanlış Negatif (YN)	Doğru Pozitif (DP)

Şekil 2: Balon varlığını tahminleyen karışıklık matrisi sonuçları



Şekil 2’de yer alan XU100 için karışıklık diğer bir adıyla hata matrisini incelediğimizde modelin balon olmadığını doğru tahmin etme başarısının 43 doğru negatif değeriyle çok yüksek olduğunu görmekteyiz. 9 doğru pozitif değeriyle modelin balon olduğunu doğru tahmin etme yeteneğinin yine oldukça başarılı ve yanlış pozitif sayısının 2 değeriyle oldukça düşük olduğunu, bu durumun ise modelin balon yok iken hata yapma oranının düşük olduğu belirtmektedir. Yanlış negatif ise hiç yoktur, yani modelin balon var iken bunu kaçırmamış olduğunu belirtmek mümkündür. Bu sonuçlara göre modelimizin, "Balon Var" sınıflarını yanlış pozitiflerden çok az etkilemiş olduğu, ancak hiç yanlış negatif bulunmadığını görmekteyiz. Yani, BIST 100 endeksinde fiyat balonunun var olduğu durumda modelin doğru tespit etme oranının yüksek olduğu görülmektedir. Genel olarak, karışıklık matrisi modelimizin dengeli bir performans sergilediğini göstermektedir. XBANK için bakıldığında, model, 46 DN değeri ile "Balon Yok" olan durumları büyük oranda doğru bir şekilde tespit etmiştir. Bu, modelin YP oranının oldukça düşük olduğunu göstermektedir. Ancak, 4 DP değeri ile "Balon Var" olan durumları da bir dereceye kadar tespit edebilmiş, üç kez "Balon Var" olan durumu kaçırap "Balon Yok" olarak tahmin etmiştir. Bu durum ise BIST Banka endeksinde modelin balon olan durumlardaki duyarlılığının (sensitivity) biraz daha iyileştirilebileceğini ifade etmektedir. BIST Sınai endeks fiyat balonlarını dikkate aldığımızda, modelin 41 DN ile "Balon Yok" olan durumları oldukça iyi bir şekilde tespit ettiği, yani balon olmayan durumlarda yüksek bir doğruluğa sahip olduğu belirtilebilir. Yine modelin 9 DP ile "Balon Var" olan durumları tespit etmede de başarılı, ancak 3 YN ile bazı balonlu durumları kaçırdığı söylenebilir. YP sayısının çok düşük olması balon olmayan durumlarda yanlış tahminlerin az olduğunu göstermektedir. Yani XUSIN için modelimiz genel olarak hem "Balon Yok" hem de "Balon Var" durumlarında başarılı iken YN sayısının bir miktar iyileştirilebileceği söylenebilecektir. Bu da modelin balonları daha hassas bir şekilde tespit edebilmesini sağlayabilir.

Sınıflandırma algoritmalarının performans değerlendirmesinde yaygın olarak kullanılmakta olan doğruluk oranı (Accuracy), bir sınıflandırıcı modelin genel performansını ölçen temel bir metrik olup, doğru tahminlerin toplam tahminlere oranını ifade etmektedir. Daha yüksek doğruluk oranı,

daha yüksek bir tahmin gücü seviyesine işaret eder. İncelenen rastgele orman sınıflandırma ağacının başarı oranı XU100 için %96,3, XBANK ve XUSIN için ise %92,6 olarak hesaplanmıştır. Bu oranlar modelin genel performansının yüksek olduğunu ve endeks fiyatlarının balon olup olmadığını belirlemede büyük ölçüde doğru tahminler sağladığını göstermektedir. Ancak, doğruluk oranının yüksekliği, modelin diğer performans metriklerini dikkate alarak daha kapsamlı bir değerlendirmeye ihtiyaç duyduğunu belirtmektedir.

Kesinlik (Precision), sınıflandırıcının pozitif bir sınıf olarak bildirdiği grupta gerçekten pozitif olduğu ortaya çıkan kayıtların oranını belirlemektedir. Kesinlik ne kadar yüksek olursa, model tarafından yapılan Yanlış Pozitiflerin sayısı yani gerçekte balon olmadığı halde balon olduğunu belirten tahminlerin sayısı o kadar düşük olmaktadır. Doğru olarak sınıflandırılan endeks fiyatlarındaki balon sayısını bulmak için kesinlik ölçütü kullanılmaktadır. Kesinlik ölçütü XU100 için 0,818 iken XBANK ve XUSIN için sırasıyla 0,8 ve 0,9'dur.

Duyarlılık (Hassasiyet, Hatırlama) ölçütü, doğru şekilde sınıflandırılan pozitif kalıpların oranını ölçmek için kullanılmaktadır. Yüksek duyarlılığa sahip sınıflandırıcılarda, yanlış olarak sınıflandırılmış negatif örnek sayısının az olması beklendiğinden eğer sınıflandırıcının mükemmel sonuçlar vermesi isteniyorsa söz konusu sınıflandırıcı tüm doğru pozitifleri seçmekte ve hiçbir doğru pozitif kaçırılmamalıdır. XU100 endeksi için, duyarlılık ölçütü 1 olarak hesaplanmıştır. Tahminleyici model, gerçek pozitif örneklerin yani balon var olan fiyatların hepsini doğru bir şekilde tahmin edebilmiştir. Sınıflandırıcıda endeks fiyatında balon olduğu halde balonun var olmadığını belirten örnek yoktur, yani pozitif sınıftaki tüm örnekler doğru şekilde sınıflandırılmıştır. Bu durumda XU100 için tahminleyici model oldukça etkili görünmektedir.

Özgüllük, bir sınıflandırıcının reddedilmesi gereken tüm durumları reddetme yeteneği olarak tanımlanmaktadır. Mükemmel bir sınıflandırıcı tüm doğru negatifleri reddetmekte ve beklenmeyen sonuçlar vermemektedir. Uygulamada ise genellikle herhangi bir sınıflandırıcı, reddedilmesi gereken bazı durumları seçmekte ve bu nedenle bazı yanlış pozitif değerlere sahip olmaktadır. Performans sonuçlarına göre, sınıflandırıcının XU100, XBANK ve XUSIN için özgüllük değeri sırasıyla 0,956, 0,979 ve 0,976'dır. Özgüllük ölçütüne göre tahminleyici sınıflandırıcı model negatif sınıfları doğru tahmin etmede yüksek bir başarı göstermektedir. Yani gerçekte balon olmadığı halde sınıflandırıcının balon olduğunu belirten yanlış pozitif örnek oranı düşük olup, modelin yüksek bir başarı göstermiş olduğunu belirtmektedir.

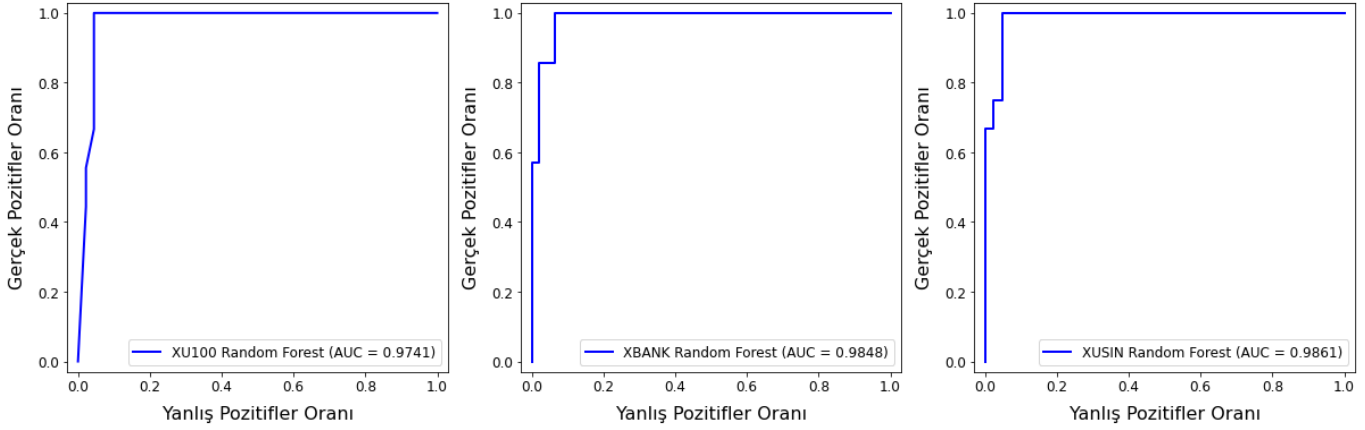
F1 puanı, duyarlılık ve kesinlik değerleri için harmonik ortalama hesaplanmaktadır. F1 puanının en iyi durumda 1 ve en kötü durumda 0 değeri aldığı göz önünde bulundurulduğunda tüm endeksler için F1 skorunun ortalamasının üzerinde iyi olduğu belirtilebilecektir.

Bir ROC eğrisi (Receiver Operating Characteristic Curve-Alıcı İşletim Karakteristik Eğrisi), tüm sınıflandırıcı modelinin performansını duyarlılık ve özgüllüğü dikkate alarak gösteren bir grafikdir. ROC eğrisi, modelin farklı sınıflandırma eşiklerinde doğru pozitiflerin oranına karşı yanlış pozitiflerin oranını çizmektedir. Bu eğri sol üst köşeye ne kadar yakınsa o kadar iyi performans gösterdiği ifade edilmektedir. Aşağıda endekslere ait şekiller incelendiğinde ROC eğrisinin sol üst köşede yer aldığı görülmektedir. İncelenen tahminleyicinin iyi performansa sahip bir sınıflandırıcı olduğu yorumu yapılabilecektir.

AUC değeri ise ROC Eğrisinin Altındaki kalan alanı ifade eden bir ölçüttür. AUC ne kadar yüksekse modelin de sınıfları doğru tahmin etme de o kadar iyi olduğu belirtilmektedir. Şekil 3'te görüldüğü üzere XUSIN, XBANK ve XU100 için AUC değerleri sırasıyla 0.9861, 0.9848 ve 0.9741 olup 1'e çok yakın olduğundan endeks fiyatlarında balon olduğu ve balon olmadığını belirten sınıfların sınıflandırıcı tarafından büyük oranda doğru sınıflandırılabilirdiği sonucunu ortaya koymaktadır.

Sonuç olarak Rastgele orman algoritması ile incelenen endeks fiyatlarının her birinde eğitim örneklemleri ile test örneklemleri üzerindeki balon tahminleri karşılaştırıldığında yüksek tahmin performansı sağlamaktadır.

Şekil 3: Rastgele Orman sınıflandırıcısının ROC eğrileri



Bir sonraki adımda, endeks fiyat balonları olasılığını tahmin etmede makroekonomik ve finansal göstergelerin göreceli önemini araştırılmıştır. Bu göstergelerin fiyat balonlarını tahmin etmedeki göreceli önemi Tablo 5’te gösterilmektedir. Bu ölçütlerin, ele alınan her endeks için fiyat balonlarını tahmin etmedeki göreceli önemlerinin ölçülmesinde ortak öngörücüler olarak kullanılması da dikkat çekicidir. Tablo 5, her bir endeks fiyatı için balon olasılığını tahmin etmede makroekonomik ve finansal gösterge öneminin bir ölçüsü olarak Gini safsızlığını göstermektedir. Gini safsızlığı, her bir ağaçtaki herhangi bir düğümde yanlış sınıflandırma olasılığını ölçer. Örnek vermek gerekirse, sepet kur için XU100 sütundaki 0,191 sayısı, bu tahminci modelden çıkarıldığında Gini safsızlık ölçüsünün %19,1 oranında azaldığı anlamına gelmektedir.

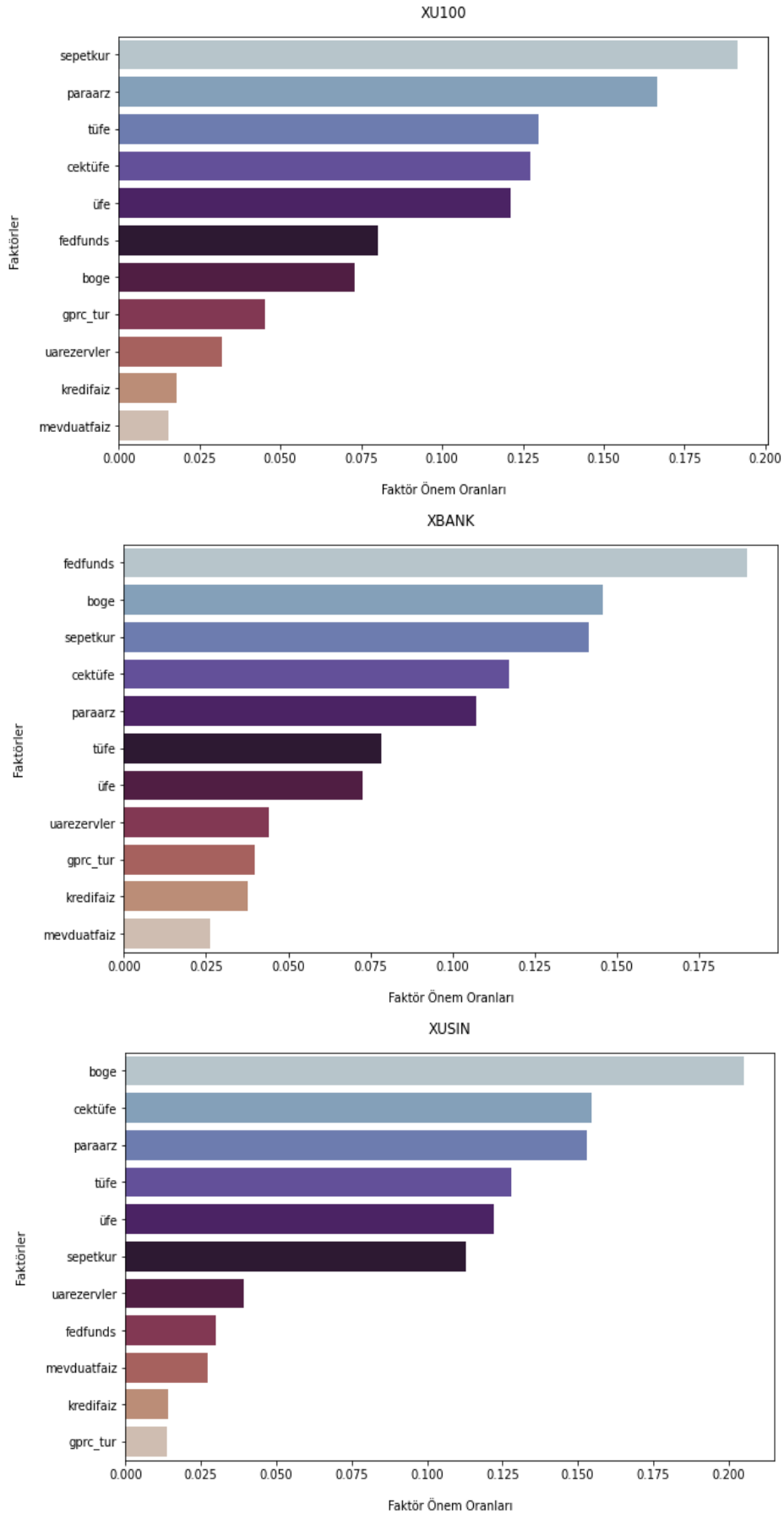
Tablo 5: Değişkenlerin Önem Düzeyleri

Değişkenler	XU100	XBANK	XUSIN
sepet kur	0,191	0,142	0,113
paraarz	0,167	0,107	0,153
tüfe	0,130	0,078	0,128
cektüfe	0,127	0,117	0,154
üfe	0,121	0,073	0,122
fedfunds	0,080	0,189	0,030
boge	0,073	0,146	0,205
gprc_tur	0,045	0,040	0,014
uarezervler	0,032	0,044	0,039
kredifaiz	0,018	0,038	0,014
mevduatfaiz	0,015	0,026	0,027

Açıklama: Tablo içerisindeki sayılar, Gini safsızlık ölçüsündeki ortalama azalmayı (%) temsil etmektedir. Kalın şekilde belirtilmiş sayılar, her bir endeks için en yüksek ağaç tahminleyicilerini göstermektedir.

Tablo 5 ve Şekil 4, endekslerde ortaya çıkan balon riskini anlamada en etkili olan makroekonomik değişkenleri sıralayarak, modelin nasıl çalıştığına dair önemli ipuçları sağlamaktadır. Grafikte görüldüğü üzere, sepet kur ve para arzı XU100 üzerindeki balon oluşumuna en çok etki eden faktörler olarak ön plandadır. Bu, döviz kurlarının ve likiditenin piyasada oynadığı kritik rolü yansıtmaktadır.

Şekil 4: Balon tahminlemede makroekonomik ve finansal faktörlerin önem oranları



Enflasyon göstergeleri (TÜFE, Çekirdek TÜFE ve ÜFE) de önemli faktörler arasında yer aldığından fiyat artışlarının balon riskini artırabileceğini göstermektedir. FED Faiz Oranları ve mevduat, kredi faizleri gibi diğer makroekonomik göstergeler daha düşük ağırlığa sahip olduğu belirtilebilir ancak yine de piyasa dinamikleri üzerinde belirli bir etkiye sahiptir.

Grafikte açıkça görüldüğü üzere, Fed Funds Rate (ABD Merkez Bankası'nın faiz oranı), XBANK endeksi üzerinde en büyük etkiye sahip faktördür. ABD'deki para politikasındaki değişiklikler, sermaye akışları ile Türkiye'deki bankacılık sektörünü doğrudan etkileyerek, döviz kuru, faiz oranları ve kredi büyümesi gibi değişkenler üzerinde önemli etkiler yaratır. TÜFE ve çekirdek enflasyon gibi enflasyon göstergeleri ile kredi büyümesini temsil eden faktörler de XBANK endeksini etkileyen önemli değişkenler arasındadır. Yüksek enflasyon ve hızlı kredi büyümesi, finansal istikrarı tehdit ederek bankacılık sektörü üzerinde olumsuz etkiler yaratabilir.

FED faiz oranları gibi küresel faktörlerin yanı sıra, BOGE ve GPRC gibi yerel faktörlerin de XUSIN endeksi üzerindeki etkisi belirgindir. BOGE endeksi, Türkiye ekonomisinin gelecekteki performansı hakkında önemli ipuçları verirken, GPRC_TUR ise jeopolitik risklerin ekonomik aktivite üzerindeki etkilerini yansıtmaktadır. TÜFE ve ÜFE gibi enflasyon göstergelerinin de balon oluşumunda önemli bir rol oynadığı görülmektedir. Yüksek enflasyon, reel getirileri düşürerek yatırımcıları riskli varlıklara yöneltebilir ve böylece balon riskini artırabilir. Kredi faizi, mevduat faizi ve uluslararası rezervler gibi diğer değişkenlerin de balon oluşumunda belirli bir etkisi bulunmakla birlikte, bu etkilerin sepet kuru ve para arzı kadar güçlü olmadığı görülmektedir.

Rastgele orman modeli sonuçları, BIST 100, BIST Banka ve BIST Sınai endeks fiyatı için fiyat balonlarını tahmin etmede nispeten daha önemli tahmincilerin sepet kur, para arzı, bileşik öncü göstergeler endeksi ve TÜFE olduğunu savunmaktadır. Başka bir deyişle, bu ölçütler ilgili endeks fiyat balonlarının doğru sınıflandırılmasına daha fazla katkıda bulunmaktadır. Tablo 5 ayrıca mevduat ve kredi faizlerinin balonları tahmin etmede daha az etkili olduğuna dair dikkate değer kanıtlar sunmaktadır. Bu sonuçlar aynı zamanda makroekonomik göstergelerin balonları tahminlemede finansal değişkenlerden daha iyi olduğunu göstermektedir.

Kısmi Bağımlılık Grafiği

Kısmi bağımlılık grafiği (Partial Dependence Plot, PDP), Friedman (2001) tarafından geliştirilen ve bir modelde belirli bir girdinin hedef değişken üzerindeki marjinal etkisini incelemeye yarayan bir analiz yöntemidir. PDP, makine öğrenimi modellerinde, özellikle rastgele orman gibi "kara kutu" modellerde, bir özelliğin çıktı üzerindeki etkisini görselleştirerek anlaşılmasını kolaylaştırmaktadır. Bu tür algoritmalarda, model tahmincileri ile model çıktısı arasındaki ilişkilerin doğrudan yorumlanması zor olduğundan, kısmi bağımlılık grafiği, bu ilişkilerin görsel olarak temsil edilmesiyle içgörüler sağlamak ve modelin davranışı hakkında daha derinlemesine bilgi sunmaktadır.

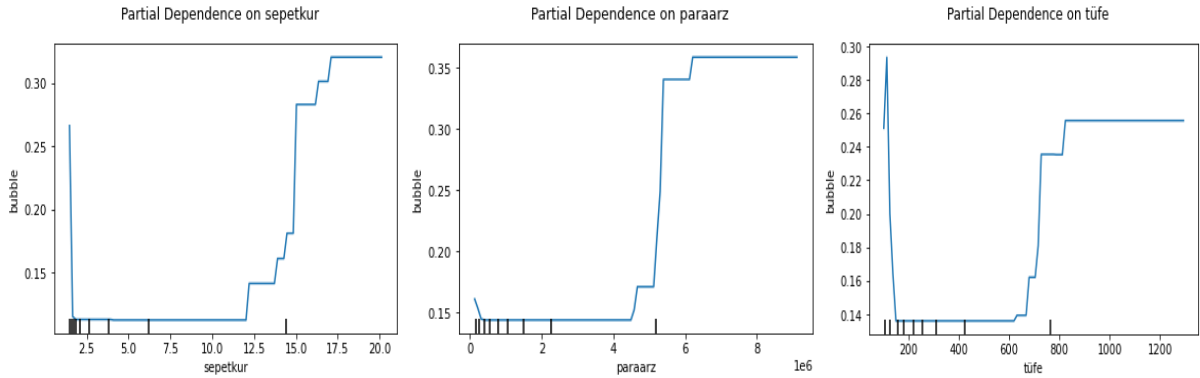
Fiyat balonlarını tahmin etmede değişkenlerin göreceli önemini tartışmadan önce, bu değişkenlerin her bir endekse ait fiyat balonu olasılığı üzerindeki etkisini göstermek için en önemli üç tahmin edici değişkenin kısmi bağımlılık grafikleri (PDP) gösterilmiştir. Şekil 5, 6 ve 7 tahminci değişkenler ile endeks fiyat balonu olasılığı arasındaki ilişkiyi ortaya koyan kısmi bağımlılık grafiklerini göstermektedir. Bu şekiller endeks fiyat balonlarının erken uyarı göstergeleri için önemli çıkarımlar sağlamaktadır. XU100 endeks fiyatında balon olma ihtimalinin sepet kur, para arzı ve TÜFE'ye olan kısmi bağımlılığı Şekil 5'te gösterilmektedir. İlk olarak, sepet kurun 17,5 seviyesine kadar yükselmesinin BIST 100 balon olasılığını artırdığı görülmektedir. Ancak, sepet kurdaki söz konusu artışın bu eşik seviyesinden sonra BIST 100 endeks fiyatındaki balonlar için ek bir kanıt sağlamadığını göstermektedir. BIST 100 endeksindeki birçok büyük şirket, ihracat ağırlıklı

çalışmaktadır. Sepet kurun artması, yani TL'nin değer kaybetmesi, bu şirketlerin döviz cinsinden gelirlerinin TL karşılığını artırmakta ve böylece bu şirketlerin kârlılık oranları yükselmektedir. Özellikle otomotiv, demir-çelik, savunma sanayi ve tekstil gibi ihracata dayalı sektörler sepet kurdaki artıştan olumlu etkilenebilir ve bu durum, endeks üzerinde yukarı yönlü bir baskı yaratabilir. İthal girdi maliyetlerine bağlı olan sektörler için sepet kurun artması olumsuz bir etki yaratır. Bu da XU100 endeksi üzerinde aşağı yönlü bir baskıya neden olabilir. 17,5 seviyesi sonrası ise fiyat balonu olasılığı için herhangi bir ek kanıt sağlayamayacağını da belirtilebilecektir.

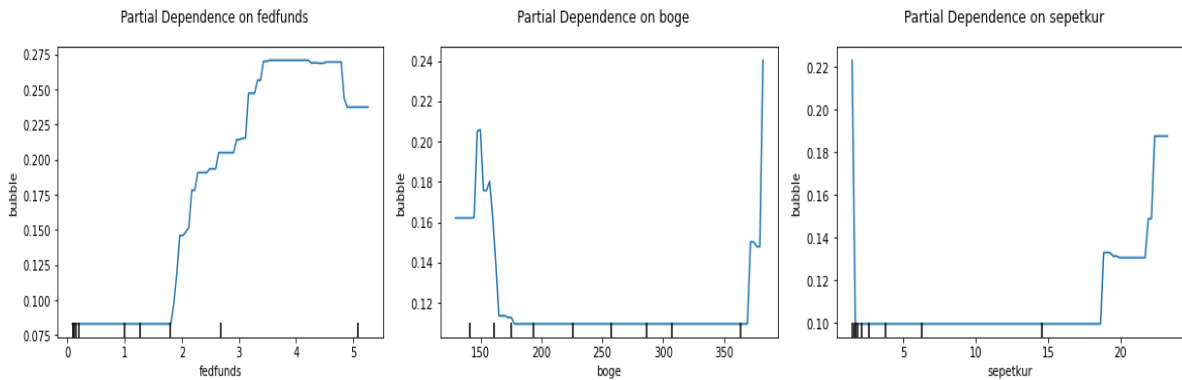
Para arzı ve TÜFE grafikleri incelendiğinde sepet kur ile benzer bağımlılık gösterdikleri görülebilecektir. Yüksek sepet kur enflasyonist baskıları artırarak şirketlerin maliyet yapısını olumsuz etkileyebileceğinden TÜFE ve para arzının belli bir seviyeden sonra artması balon oluşum olasılığına önemli derecede etki etmektedir. Şekil 7'de yer alan XUSIN ve para arzı için balon olasılığı grafiğinde açıkça görülmektedir ki para arzındaki artışlar sanayi sektöründeki hisse senedi fiyatlarını pozitif yönde etkileyebileceğinden bu durum yatırımcılar için önemli bir gösterge oluşturmaktadır.

Şekil 6'da yer alan XBANK endeks fiyatındaki balon oluşum olasılığı ve sepet kur arasındaki ilişki de XU100 endeksine benzer şekilde ifade edilebilmektedir. Sepet kur, enflasyon, faiz oranları ve diğer makroekonomik göstergelerle etkileşim halindedir. Özellikle döviz kurundaki volatilité, bankacılık sektörünün finansal sağlığı üzerinde doğrudan ve dolaylı etkiler yaratabilmektedir. Yüksek döviz kuru, bankaların döviz pozisyonlarını bozarak, kredi riskini artırabilir ve karlılıklarını olumsuz etkileyebilir. Ayrıca, kur dalgalanmaları, dış ticaret dengesi üzerindeki etkisiyle de bankacılık sektörünün faaliyetlerini dolaylı olarak şekillendirmektedir (Aka, 2020).

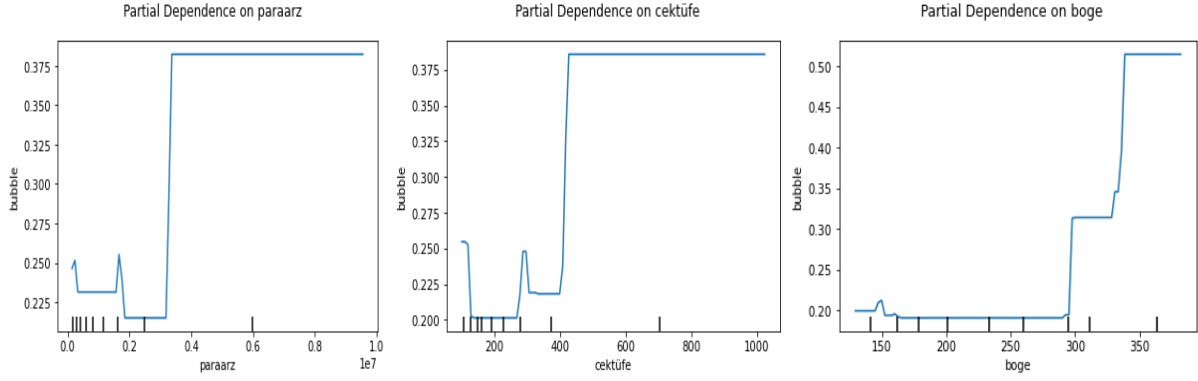
Şekil 5: XU100 için Kısmi Bağımlılık Grafikleri



Şekil 6: XBANK için Kısmi Bağımlılık Grafikleri



Şekil 7: XUSIN için Kısmi Bağımlılık Grafikleri



Açıklama: Dikey eksen, olasılık oranının logaritmasını göstermektedir. Olasılık oranı, " $p / (1-p)$ " olarak ölçülür ve burada p , balon olasılığını temsil etmektedir. Yatay eksen, tahmin edici değerlerin seviyelerindeki değerlerini göstermektedir.

Şekil 6'da yer alan grafiğin başlangıcında, FED faiz oranı 2'nin altındayken balon olasılığı çok düşük seviyede kalmaktadır. Bu durum, faiz oranlarının düşük olduğu dönemlerde XBANK endeksinde balon oluşma olasılığının çok düşük olduğunu göstermektedir. FED faiz oranı 2'nin üzerine çıktığında balon olasılığı hızla artmakta ve 4 civarına geldiğinde yaklaşık 0.27 seviyesine ulaşmaktadır. Bu, faiz oranlarının belirli bir seviyeye kadar arttıkça balon oluşma ihtimalinin yükseldiğini göstermektedir. FED'in politika faiz oranı yüzde 4,5'u aştığında XBANK endeks fiyatında balon oluşma olasılığının azaldığı ve bu nedenle parasal sıkılaşma ve yükselen faiz oranlarının banka hisselerine etkisinin dikkate değer olduğu görülmektedir.

Şekil 6 ve 7'de BOGE'de 300-350 seviyelerinden sonra meydana gelen artışın XBANK ve XUSIN endekslerindeki balon olasılığını arttırdığı görülmektedir. Araştırmalar, bileşik öncü göstergelerin XBANK üzerindeki etkisinin istatistiksel olarak anlamlı olduğunu ortaya koymuştur. BOGE sanayi sektöründeki şirketlerin geleceğe yönelik beklentilerini yansıttığı için, yatırımcıların bu göstergeleri dikkate alarak stratejilerini belirlemeleri önerilmektedir. Özellikle sanayi üretimindeki değişimlerin ve piyasa beklentilerinin, XUSIN üzerindeki etkileri yatırım kararlarını etkileyebilmektedir (Ünal, 2021).

Genel olarak, makroekonomik faktörler endeks fiyat balonlarının öngörülmesinde kritik bir rol oynuyor gibi görünmektedir. Ancak, mevduat ve kredi faizlerinin gücü benzerlerine göre nispeten daha zayıftır. Tüm endekslerde balon oluşumuna etki eden dışsal değişkenlere bakıldığında, incelenen makroekonomik ve finansal göstergeler arasında fiyat balonu üzerinde diğer göstergelere nazaran en az etkili olan göstergeler kredi ve mevduat faizi olarak belirlenmiştir.

5. SONUÇ

Borsa endeksleri ve bu endekslerin fiyat değişimleri, sermaye piyasalarında kullanılan önemli finansal araçlar olup, bir ülkenin finansal gelişmişlik düzeyini gösteren önemli belirleyiciler arasında yer almaktadır. Bir ekonominin gerçek kaynak tahsisi varlık fiyatlarından etkilendiği için (Brunnermeier, 2016; Stiglitz, 1990), bu fiyatların ne zaman ve neden temel değerlerinden saptığını anlamak kritik öneme sahiptir. Bu anlayış, ilgili paydaşların bilinçli kararlar almalarına rehberlik eden erken uyarı sistemleri (EWS) geliştirmelerine de yardımcı olabilir. Bu bağlamda, finansal gelişmişlik düzeyi yüksek ekonomilerde, ekonomik istikrarın sağlanması durumunda volatilitenin düşeceği ve bunun da riski azaltacağı düşünülmektedir.

Finansal balonlar, finansal varlık fiyatlarındaki sürekli ve sistematik artışların, temel değerlerden sapmalara neden olarak ortaya çıkmakta ve bu durum önemli finansal krizlere yol açabilmektedir. Spekülatif hareketler, varlık fiyatlarında büyük dalgalanmalara sebep olabilmektedir. Bu bağlamda,

çalışmada Borsa İstanbul'da yer alan BIST100, Banka ve Sınai endekslerini analiz ederek fiyat dalgalanmalarının balon etkisi oluşturup oluşturmadığı, Phillips, Shi ve Yu (2015) tarafından geliştirilen özyinelemeli Sağ Kuyruklu Genelleştirilmiş Supremum Artırılmış Dickey-Fuller Testi (GSADF) kullanılarak 2002-2024 dönemindeki aylık fiyatlar için incelenmiştir. Araştırma kapsamında sınanan temel hipotezler reddedilerek BIST 100, BIST Banka ve BIST Sınai endekslerinde finansal balonların varlığı gözlemlenmiştir. Ayrıca, Endekslerde varlığı ortaya çıkarılan finansal balonların genel itibarıyla yabancı yatırımlar, Türkiye ve ABD Merkez Bankası'ndaki gelişmeler, gibi faktörlerden kaynaklandığı söylenebilir. Buna göre, pay piyasalarının aslında etkin olmadığı ve yatırımcıların rasyonel olmayan spekülasyon davranışları ile sürü psikolojisi sergiledikleri sonucuna varılmaktadır. İlk aşama analizinin sonuçları, her bir endeksin farklı sürelerde balonlar yaşadığını göstermektedir.

Makine öğrenme algoritmalarının etkinliği, kullanılan özniteliklerin niteliği ve aralarındaki ilişkinin yanı sıra bu özniteliklerin temsil gücüne bağlıdır. Yüksek temsil gücüne sahip öznitelikler bile, uygun bir kombinasyon oluşturulmadığı takdirde, makine öğrenme algoritmalarının başarısını olumsuz etkileyebilir. Çalışmanın ikinci aşamasında, balon oluşumunun gerçek zamanlı tahmininde ise rastgele orman (Random Forest) makine öğrenmesi algoritmaları kullanılmıştır. Makine öğrenmesi analizlerinin bulguları, fiyat balonlarının ortaya çıkışının endekslere özgü özellikler gösterebileceğini ortaya koymaktadır. Ayrıca, her bir endeks fiyatı için balon oluşumunu etkileyen değişkenlerin eşik değerleri, balonları tespit etmek için erken uyarı göstergeleri geliştirmede kullanılabilir.

Çeşitli finansal ve makroekonomik faktörler ile balonlar arasındaki ilişki, literatürde bir ilk olarak makine öğrenmesi modelleri kullanılarak incelenmiştir. Makine öğrenmesi algoritmalarının en önemli avantajı, değişkenler arasında herhangi bir önceden belirlenmiş fonksiyonel forma ihtiyaç duymamaları ve kesin dağılımsal varsayımlara bağlı kalmamalarıdır. Ayrıca bu algoritmalar, geliştirilen modellerin örneklem dışı performanslarını da başarıyla test etme imkânı sunmaktadır. Finansal piyasalar ve ülkenin önemli borsa endeks göstergeleri söz konusu olduğunda, EWS, yatırımcıların portföylerini daha verimli bir şekilde yeniden tahsis etmelerine yardımcı olabilir.

Bu çalışma doğrultusunda gelecekteki çalışmalarda söz konusu endeksleri veya XU100 ve XBANK gibi diğer endekslerdeki balon oluşumlarının da benzer bir şekilde incelenmesi, sonuçların genel geçerliliğini artıracaktır. Bu bağlamda makine öğrenmesi alanındaki gelişmeler doğrultusunda, farklı algoritmaların kullanılması ile sonuçların doğrulanması önemlidir. Balon oluşumları ve makroekonomik değişkenler arasındaki non-lineer ilişkilerin incelenmesi, balon oluşumunun daha iyi anlaşılmasına katkı sağlayabilir. Sonuç olarak, bu çalışma borsa endeksindeki balon oluşumunu anlamak için önemli bir adım teşkil etmektedir. Ancak, balon riskini tamamen ortadan kaldırmak mümkün olmamakla birlikte, uygun politikalar ve düzenlemeler ile bu risk azaltılabilir.

Bu çalışmada elde edilen bulgular, potansiyel balonlar konusunda farkındalık oluşturmak üzere erken uyarı sistemleri (EWS) geliştirmek amacıyla kullanılabilir. Gelecekte bu konuda yapılacak çalışmalar için, EWS'lerin farklı veri kaynakları ve algoritmalarla entegrasyonu, farklı piyasa koşullarına ve endekslere uyarlanabilirliği üzerine odaklanılabilir. Çalışmada kullanılan Random Forest algoritması, balon oluşumunu tahmin etmede umut vaatsetse de derin öğrenme modelleri gibi diğer makine öğrenmesi tekniklerinin de araştırılması ve karşılaştırılması önemlidir. Balon dinamikleri ve etkileşimlerini daha ayrıntılı analiz etmek için zaman serisi analizi ve diğer ekonometrik teknikler kullanılabilir.

KAYNAKÇA

- Aka, K. (2020). Seçilmiş makroekonomik göstergelerin döviz kuru üzerinde etkisi: Türkiye ekonomisi üzerine bir uygulama. *BDDK Bankacılık ve Finansal Piyasalar Dergisi*, 14(1), 99-117.
- Akkaya, M. (2018). Borsa İstanbul hisse senedi getirilerinde balon oluşumu üzerine bir uygulama. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*.
- Astill, S., Harvey, D. I., Leybourne, S. J., ve Taylor, A. R. (2017). Tests for an end-of-sample bubble in financial time series. *Econometric Reviews*, 36(6-9), 651-666.
- Başoğlu Kabran, F., ve Ünlü, K. D. (2021). A two-step machine learning approach to predict S&P 500 bubbles. *Journal of Applied Statistics*, 48(13-15), 2776-2794.
- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Breiman, L., Friedman, J. H., Olshen, R. A., ve Stone, C. J. (1984). *Classification and regression trees*. Wadsworth and Brooks, Monterey, CA.
- Brunnermeier, M. K. (2016). Bubbles. In *Banking Crises: Perspectives from The New Palgrave Dictionary* (pp. 28-36). London: Palgrave Macmillan UK. https://doi.org/10.1057/9781137553799_5.
- Case, K. E., ve Shiller, R. J. (2003). Is there a bubble in the housing market?. *Brookings papers on economic activity*, 2003(2), 299-362.
- Caspi, I. (2017). Rtaf: Testing for bubbles with EViews. *Journal of Statistical Software*, 81, 1-16. <https://doi.org/10.18637/jss.v081.c01>
- Chang, T., Aye, G. C., ve Gupta, R. (2014). Testing for multiple bubbles in the BRICS stock markets. *University of Pretoria Department of Economics Working Paper Series*, 7.
- Çağlayan Akay, E. (2020). Ekonometride büyük veri ve makine öğrenmesi: Temel kavramlar. AVESİS. <https://avesis.marmara.edu.tr/yayin/609c108f-c897-4877-a16c-cddb401d46f5/ekonometride-buyuk-veri-ve-makine-ogrenmesi-temel-kavramlar>
- Deperlioğlu, Ö., ve Köse, U. (2024). Python ile makine öğrenmesi. Ankara: Seçkin Yayıncılık,
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189-1232. <https://www.jstor.org/stable/2699986>
- Gilles, C., & LeRoy, S. F. (1992). Asset price bubbles. *The New Palgrave Dictionary of Money and Finance*. Macmillan Reference, London.
- Hennequin, M., ve Hommes, C. (2023). Managing bubbles in experimental asset markets with monetary policy. *Journal of Money Credit and Banking*, 56(2-3), 429-454. <https://doi.org/10.1111/jmcb.13050>
- Hoque, K., ve Aljamaan, H. (2021). Impact of hyperparameter tuning on machine learning models in stock price forecasting. *IEEE Access*, 9, 163815-163830. <https://doi.org/10.1109/access.2021.3134138>
- James, G., Witten, D., Hastie, T., ve Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: Springer.
- Ji, M., Dou, Z., Li, H., ve Wang, M. (2021). Multi-characteristic product price research based on gsadf-bp model. *IEEE Access*, 9, 166870-166885. <https://doi.org/10.1109/access.2021.3132726>

- Koç Ustalı, N., Tosun, N., ve Tosun, Ö. (2021). Makine öğrenmesi teknikleri ile hisse senedi fiyat tahmini. *Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 16(1), 1-16. <https://doi.org/10.17153/oguiibf.636017>
- Korkmaz, Ö., Erer, D., ve Erer, E. (2016). Alternatif yatırım araçlarında ortaya çıkan balonlar Türkiye hisse senedi piyasasını etkiliyor mu? BİST 100 üzerine bir uygulama. *BDDK Bankacılık ve Finansal Piyasalar*, 10(2), 29-61.
- Lancaster University, Vasilopoulos, K., Pavlidis, E., Lancaster University, Martínez-García, E., ve Federal Reserve Bank of Dallas. (2021). exuber: Recursive Right-Tailed Unit Root Testing with R. Federal Reserve Bank of Dallas, *Globalization Institute Working Papers*, 2020(383). <https://doi.org/10.24149/gwp383r1>
- Liu, C.-F., Yeh, C.-Y., ve Lee, S.-J. (2012). Application of type-2 neuro-fuzzy modeling in stock price prediction. *Applied Soft Computing*, 12(4), 1348-1358. <https://doi.org/10.1016/j.asoc.2011.11.028>
- Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, 8(4), 283-298. [https://doi.org/10.1016/S0001-2998\(78\)80014-2](https://doi.org/10.1016/S0001-2998(78)80014-2)
- Nikou, M., Mansourfar, G., ve Bagherzadeh, J. (2019). Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting Finance & Management*, 26(4), 164-174. <https://doi.org/10.1002/isaf.1459>
- O'Brien, M., ve Wosser, M. (2018). An Early Warning System for Systemic Banking Crises: A Robust Model Specification. *Central Bank of Ireland Research Technical Paper*, 9.
- Phillips, P. C. B., ve Shi, S. (2020). Real time monitoring of asset markets: Bubbles and crises. In *Handbook of statistics* (Vol. 42, pp. 61-80). Elsevier.
- Phillips, P. C. B., Shi, S., ve Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, 56(4), 1043-1078. <https://doi.org/10.1111/iere.12132>
- Rouf, N., Malik, M., Arif, T., Sharma, S., Singh, S., Aich, S., ve Kim, H. (2021). Stock market prediction using machine learning techniques: A decade survey on methodologies, recent developments, and future directions. *Electronics*, 10(21), 2717. <https://doi.org/10.3390/electronics10212717>
- Sağlam Bezgin, M., ve Başar, M. (2020). The research of asset price bubble at Borsa Istanbul and financial crisis relationship. *Anadolu University Journal of Social Sciences*, 20(2), 143-156.
- Santoni, G. J. (1987). The great bull markets 1924-29 and 1982-87: Speculative bubbles or economic fundamentals. *Federal Reserve Bank of St. Louis Review*, 69(9), 16-29.
- Shi, S. P., Phillips, P. C., ve Yu, J. (2011). Specification sensitivities in right-tailed unit root testing for financial bubbles. *Hong Kong Institute for Monetary Research Working Paper*, 17.
- Stiglitz, J. E. (1990). Symposium on bubbles. *Journal of Economic Perspectives*, 4(2), 13-18.
- Tran, K. L., Le, H. A., Lieu, C. P., ve Nguyen, D. T. (2023). Machine learning to forecast financial bubbles in stock markets: Evidence from Vietnam. *International Journal of Financial Studies*, 11(4), 133. <https://doi.org/10.3390/ijfs11040133>
- Uğuz, S. (2019). *Makine öğrenmesi teorik yönleri ve Python uygulamaları ile bir yapay zekâ ekolü*. Ankara: Nobel Yayıncılık.

- Ural, M. (2022). Varlık fiyatlarında spekülâtif fiyat baloncuklarının sađ kuyruklu ADF yöntemiyle analizi. *İzmir İktisat Dergisi*, 37(1), 189-205. <https://doi.org/10.24988/ije.910996>
- Ünal, S. (2021). Makro ekonomik faktörler ile borsa istanbul hisse senedi endekslerinin getirileri arasındaki ilişki. *Pamukkale University Journal of Social Sciences Institute/Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (47).
- Zeren, F., ve Yılançı, V. (2019). Are there multiple bubbles in the stock markets? Further evidence from selected countries. *Ekonomika*, 98(1), 81-95. <https://doi.org/10.15388/ekon.2019.1.5>



© Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY NC) license.
(<https://creativecommons.org/licenses/by-nc/4.0/>).

EXTENDED ABSTRACT

Identifying and Forecasting of Financial Bubbles Based on Right-Tailed ADF and Machine Learning Methods

1. Introduction

A bubble in financial markets is defined as a sustained and systematic deviation of the market value of an economic asset from its fundamental value (Santoni, 1987). Bubbles, which occur when asset prices exceed their fundamental value, present significant risks for policymakers and financial regulators, who have a role in the financial system (Case and Shiller, 2003). Consequently, the detection of anomalies in stock market data is of great importance in identifying the onset of financial crises and monitoring asset price bubbles.

The objective of this study is to identify and predict financial bubbles in the BIST 100, BIST Bank, and BIST Industrials indices over the period 2002M01-2024M06 by combining econometric and machine learning methods. Accordingly, the Generalised Supremum Augmented Dickey-Fuller (GSADF) test is first used to identify bubble periods. Subsequently, the Random Forest Algorithm, a powerful machine learning method, is used to establish a predictive model for bubble occurrences by establishing correlations with a set of macroeconomic indicators. Moreover, the Random Forest algorithm determines the relative importance of financial and macroeconomic factors that influence the probability of bubble formation.

2. Data Set and Method

The existence of speculative bubbles and the dates of their formation are determined by GSADF, the right-tailed unit root test developed in Phillips, Shi and Yu (2015), using monthly data since 2002 for the Borsa Istanbul 100 index (BIST100), Banks (XBANK) and Industrials (XUSIN) prices. Subsequently, random forest trees are employed. A random forest is a supervised machine learning algorithm that employs ensemble learning, a technique that combines multiple classifiers. In essence, the objective of the random forest approach is to integrate the outputs of multiple decision trees, each trained on a distinct dataset, rather than generating a single decision tree (Çağlayan Akay, 2020).

The classification tree, in which the dependent variable represents the presence or absence of a bubble in stock market indices, is constructed in accordance with the methodological framework proposed by Breiman et al. (1984). This framework defines two categories: "There is a Bubble" and "No Bubble." The following macroeconomic and financial indicators are used as explanatory variables in the model: consumer price index (TÜFE), core inflation, producer price index, international reserves, lending rate, interest rate on Turkish lira deposits, composite leading indicators index, exchange rate, M2 broad money, geopolitical risk index for Turkey, and Federal Reserve interest rates. These variables are posited to affect the formation of bubbles in the stock market. As with the BIST 100, BIST Industrials, and BIST Bank Price Index, all exogenous variables are employed on a monthly basis. The relative importance of all indicators affecting bubble formation is determined through the application of a random forest algorithm.

3. Empirical Findings

Upon evaluation of the GSADF test results in general, the existence of speculative financial bubbles in the XU100, XBANK, and XUSIN indices is identified. Concurrently, the GSADF test was employed to ascertain the commencement and conclusion of the price bubble in the indices. Upon examination of the aggregate representation of the bubble periods in the indices, it becomes evident that the majority of these periods occur concurrently.

The Random Forest algorithm is employed to ascertain the relative importance of financial and macroeconomic factors influencing the probability of bubble formation. In the field of machine learning, the evaluation criteria are of paramount importance for the understanding and comparison of the performance of the models in question. The performance of the Random Forest algorithm was evaluated using a variety of metrics. The performance metrics employed were accuracy rate, precision, sensitivity, specificity, F1 score, and AUC value. The results of the analysis demonstrate that the money supply, exchange rate, consumer price index, Federal Reserve interest rate, and Composite Leading Indicators Index are significant determinants of the formation of price bubbles in the BIST 100, BIST Bank, and BIST Industrials indices in Turkey. Concurrently, these metrics illustrate the predictive efficacy of the random forest model in the analysis. Consequently, it underscores the necessity for the prudent implementation of monetary policy and macroprudential measures to mitigate bubble risks and safeguard financial stability. Additionally, it demonstrates that machine learning methodologies possess considerable potential for the early detection of financial bubbles and the management of market risks. The results of the model evaluation indicate that the proposed approach exhibits high predictive power and can serve as an effective alternative for predicting financial bubbles.

4. Discussion and Conclusion

Given the critical role of bubble detection and analysis in ensuring financial stability, this study proposes a two-stage machine learning model based on a real-time bubble detection test and random forest trees (RFA) to predict bubbles in the Istanbul stock market. This study makes a significant contribution to the existing literature in three key ways. Firstly, it develops a methodology for the detection of bubbles in stock market index prices. Secondly, it elucidates the relationship between macroeconomic factors and bubble formation. Thirdly, it employs machine learning methods to predict bubble periods within the aforementioned framework. In this context, the objective is to gain a deeper comprehension of market behaviour by examining the dynamic interactions of macroeconomic variables.

This research makes a notable contribution to the existing literature by employing the methodological advantages of machine learning techniques, which have significantly impacted the detection and prediction of financial bubbles. In contrast to conventional statistical techniques, machine learning methodologies offer a number of benefits. Furthermore, as existing research has focused on linear relationships between macroeconomic indicators and stock returns, this study makes a significant contribution to the literature by addressing the nonlinear and complex patterns of relevant macroeconomic variables on BIST index returns in a nonparametric manner. This study builds upon previous research in a number of ways. Firstly, the research employs random forest algorithms in conjunction with the GSADF test. Secondly, in contrast to the previous study, which employed daily data, our analysis utilises a more comprehensive range of monthly data spanning from 2002:01 to 2024:06. In conclusion, our study adopts a macroeconomic perspective, incorporating a range of variables, including the basket exchange rate, the consumer price index, the producer price index, the core inflation, the money supply, the Federal Reserve interest rate, international reserves, the composite leading indicators index, the geopolitical risk index for Turkey, and the deposit and loan rates. In this way, our study offers policymakers and investors valuable insights into the formation mechanisms of financial bubbles, providing them with the information and knowledge they require to make informed decisions.