

TEKNİK GÖSTERGELERİ MAKİNE ÖĞRENİMİ MODELLERİNE ENTEGRE EDEREK ŞEHİR ENDEKSİ HAREKETİNİN YÖNÜNÜ TAHMİN ETMEYE YÖNELİK BİR YAKLAŞIM

Öyküm Esra YİĞİT¹
Merve KARAKÖSE²

Atıf/©: Yiğit, Ö., E. ve Karaköse, M. (2021). Teknik göstergeleri makine öğrenimi modellerine entegre ederek şehir endeksi hareketinin yönünü tahmin etmeye yönelik bir yaklaşım. *Hitit Sosyal Bilimler Dergisi*, 14(2), 556-575. doi: 10.17218/hititsbd.979391

Özet: Borsa İstanbul (BIST) tarafından hesaplanan şehir endeksleri, belirli bir bölgeye yatırım yapmak isteyen yatırımcılara önemli bir yol göstericidir. Türkiye'de 13 ilin finansal performansları 2009 yılı başından itibaren bu endeksler tarafından yansıtılmaktadır. Bu çalışmada, şehir endeksi serilerinin çoğunlukla oynaklığı ile ilgilenen önceki çalışmalardan farklı olarak, teknik göstergeleri farklı makine öğrenmesi modellerine entegre ederek şehir endeksi serilerine ait hareketin yönünü tahmin etmede kullanılabilen bileşik bir prosedür önerilmiştir. Önerilen prosedür, BIST'de işlem gören en yüksek hisse senedi sayısına sahip İstanbul şehir endeksi (XSIST) serisine uygulanmıştır. Hacim, oynaklık, trend ve momentuma dayalı 38 farklı teknik gösterge hesaplanmış ve XSIST serisinin günlük değişiminde en etkili göstergeler, 6 farklı makine öğrenmesi modellerine girdi olarak seçilmiştir. Öğrenme modellerin performansı, karmaşıklık matrislerine dayalı metrikler yardımıyla karşılaştırılmıştır.

Anahtar Kelimeler: Şehir Endeksi, Makine Öğrenmesi, Öznitelik Mühendisliği, Teknik Göstergeler

An Approach to Predicting Direction of City Index Movement by Integrating Technical Indicators into Machine Learning Models

Citation/©: Yiğit, O., E. and Karaköse, M. (2021). An approach to predicting direction of city index movement by integrating technical indicators into machine learning models. *14(2)*, 556-575. doi: 10.17218/hititsbd.979391

Abstract: City indexes that have been calculated by Borsa Istanbul provides a deeper understanding to investors who are interested on making investments to a specific region. Financial performances of 13 cities in Turkey are reflected by these indexes from the beginning of the year 2009. This study differs from the earlier ones mostly focused on the volatility by integrating the technical indicators into different machine learning models for the purpose of forecasting the direction of movement in the daily city indexes. The proposed procedure was applied to the Istanbul city index (XSIST) series, which has the highest number of stocks traded in BIST. 38 different technical indicators based on volume, volatility, trend and momentum were calculated and the most effective indicators in the daily change of XSIST series were selected as inputs to 6 different machine learning models. The performance of the learning models was compared with the help of metrics based on confusion matrices.

Keywords: City Index, Machine Learning, Feature Engineering, Technical Indicators

Araştırma Makalesi / Research Article

Makale Geliş Tarihi / Submitted: 5.8.2021 Makale Kabul Tarihi / Accepted: 30.12.2021

¹ Sorumlu Yazar, Doç. Dr., Yıldız Teknik Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, İstanbul/Türkiye, oeyigit@yildiz.edu.tr, <http://orcid.org/0000-0001-7805-3979>

² Yüksek Lisans Öğrencisi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Programı, İstanbul/Türkiye, merve.karakose34@gmail.com, <http://orcid.org/0000-0002-9693-8997>

Finansal Destek: Bu çalışma Yıldız Teknik Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından FYL-2021-4436 nolu proje kapsamında desteklenmiştir.

1. GİRİŞ

Borsa, yatırımcıları bir araya getiren, varlıkların alım/satımının yapıldığı ülke ekonomisinin önemli bileşenlerinden olan bir piyasadır. Borsadaki yatırımcılar piyasa bilgilerini derinlemesine analiz ederek en düşük riskle en yüksek karı elde etmeyi amaçlamaktadırlar. Finansal piyasalarda oldukça uzun bir süre, Fama (1970) tarafından “Etkin Piyasalar Hipotezi” olarak tanımlanan ve finans biliminin temel taşı olarak kabul edilen kuralların tamamen geçerli olduğuna inanılmaktaydı. Etkin bir piyasada varlıkların davranışının bir rasgele yürüyüş süreci olduğu ve dolayısıyla yatırımcının geçmiş hareketlerden yararlanarak geleceği tahmin etmesinin mümkün olmadığı belirtilmekte, bütün hareketler olağan akışta olduğu için spekülasyon davranışlarından arbitraj kazanımının olamayacağına altı çizilmektedir. Günümüzde finansal varlıklara ait hareketlerin rastlantısallığı genel olarak kabul edilmekle birlikte bu teorinin geçerliliğini sorgulayan ve varlıkların gelecekteki davranışını tahmin etmeye odaklanan önemli çalışmalar mevcuttur (Khaidem ve diğerleri, 2016, s.1). Bu çalışmalar iki ana başlık halinde incelenebilir. İlki temel analiz olarak adlandırılır ve bir yatırımcının ilgili varlığa ait yatırım kararı verme sürecinde varlıkların içsel değerleri, ekonomi ve ilgili sektörün performansı ve politik iklim gibi bazı temel göstergeler kullanılır. İkincisi ise teknik analizdir ve temel analizden farklı olarak varlıklara ait değerlerin geçmiş hareketlerine (fiyat, hacim, vb.) dayalı bir tahminleme yapma amacı güdüldür (Patel ve diğerleri, 2015, s.259). Teknik analizin temel amacının bir varlığın içsel değerini ölçmekten ziyade gelecekteki davranışını öngörmek için modeller kurmak olması, etkin piyasalar hipotezinin temel kuralı ile çelişmekte ve uzun bir süredir tartışılan bir konu olma niteliği taşımaktadır.

Varlıklara ait hareketlerin öngörülmesi amacıyla kurulan modeller yatırımcılara potansiyel piyasa risklerine karşı etkin bir korunma sağlarken, spekülasyonun ilgili tahmine dayanan reaksiyonlarında yüksek kar elde etme fırsatı sunmaktadır (Kumar ve diğerleri, 2006, s.2). Her ne kadar varlıkların hareketine ait kurulan modeller temelde klasik istatistiksel yöntemlere dayansa da zaman serisinin doğasından kaynaklanan gürültülü, kompleks ve lineer olmayan yapısı, bu yöntemlerin bazı varsayımlarına ters düşmekte ve uygulamada geçerliliğini/genelleştirilebilirliğini bozmaktadır. Son yıllarda artan bilgi teknolojilerinin büyük miktardaki karmaşık veriyi işleme ve saklama olanağı sağlaması ve bu gelişmeye paralel olarak makine öğrenme sistemlerinin lineer olmayan yapıdaki serileri modellemede oldukça yüksek başarılar gösteren algoritmaların geliştirilmesi, “Finansal Serilerden Bilgi Keşfi” kavramının güçlenmesine olanak sağlamıştır. Finansal verilerin gelecekteki davranışlarına yönelik öngörü çabalarında makine öğrenmesi teknikleri karar vericilere oldukça önemli faydalar sunmaktadır. Bu sebeple araştırmacılar, yapay zekâ tekniklerinin bir dalı olan makine öğrenmesi yöntemlerini, finansal verilerdeki karmaşık ilişkileri modellemede, saklı örüntüleri keşfetmede ve dahası geçmiş verilerden öğrenmeye dayalı bir öngörü modeli elde etmede kullanmakta ve oldukça başarılı sonuçlar elde etmektedirler.

Öte yandan, finansal varlıkların gelecekteki davranışını öngörmeye kullanılan tekniklerin başarısı çoğunlukla modele alınan özneliklerin seçimine bağlıdır. Öznelikler (bağımsız değişken) ile çıktı (bağımlı değişken) arasındaki ilişkinin doğru bir şekilde belirlenmesi sürecinde araştırmacı açısından en önemli ve en meşakkatli adım özneliklerin etkin bir şekilde belirlenmesidir. İstatistiksel açıdan modele anlamlı katkı sağlayan ve kuramsal teori ile çelişmeyen özneliklerin eksiksiz bir şekilde ölçülerek analize dâhil edilmesi arzu edilen bir stratejidir. Ayrıca sade/basit/anlaşılır ve hesaplama yükü az (cost effective) bir model kurmak da bu stratejiye dâhildir. Kısacası, çıktıyı etkileyen tüm öznelikleri kullanmak ve tutumlu (parsimonist) davranmak arasındaki dengenin etkin bir şekilde kurulması gerekmektedir. Bu sebeple makine

öğrenme yöntemlerinin uygulanışında etkili özniteliklerin çıkartılması/seçilmesi, modele dâhil etme aşamasında gerekli dönüşümlerin yapılması ve yüksek boyuttaki öznitelik uzayının boyutunun azaltılarak model karmaşıklığının düşürülmesi gibi işlemlerin bir bütünü olan “Öznitelik Mühendisliği” (feature engineering) kavramı son yıllarda önemini arttırmıştır.

BIST’te işlem gören payların gruplar halinde getiri performanslarının ölçüldüğü Pay Endeksleri kapsamında makine öğrenmesi teknikleri kullanarak geleceğe yönelik tahmin çalışmaları incelendiğinde çoğunluğunda özniteliklerin mikro ve/veya makroekonomik göstergelerden seçildiği görülmektedir. Bu göstergelerden bazıları döviz kuru, petrol fiyatı, altın fiyatı, faiz oranı, para arzı, TÜFE, cari açık ve ödemeler dengesi olarak tanımlanabilir (Akcan ve Kartal, 2011; Altınbaş ve Bişkin, 2015; Boyacıoğlu ve Avcı, 2010; Filiz ve diğerleri, 2017; Kantar, 2020; Karaatlı ve diğerleri, 2005; Özdemir ve diğerleri, 2011; Telli ve Coşkun, 2016; Yakut ve diğerleri, 2014). Öte yandan, özellikle borsa tüccarları tarafından kullanılan ve finansal varlığa ait meta-bilgiden (açılış değeri, kapanış değeri, en yüksek/en düşük değer, hacim) yararlanarak üretilen sezgisel veya model tabanlı göstergelerin öznitelik olarak seçilerek makine öğrenmesi algoritmalarına girdi olarak alınan çalışmalar da mevcuttur (Hasan ve diğerleri, 2020; Kara ve diğerleri, 2011; Pabuçcu, 2019; Tekin ve Çanakoglu, 2018). BIST pay endeksleri kapsamında yukarıda bahsi geçen çalışmaların hemen hemen hepsi Türkiye borsasının uluslararası alanda güçlü bir yansıması olarak nitelendirilen BIST 100, BIST 30 veya BIST 50 endeksleri için gerçekleştirilmiş, bölgesel kalkınmanın önemli bir göstergesi olan şehir endeksleri için makine öğrenmesine dayalı bir öngörü modeli oluşturulan çalışmaya rastlanmamıştır.

Bu çalışma, BIST şehir endeksleri için ulusal dizinde bulunan çalışmalardan iki açıdan farklılık göstermektedir. İlki, şehir endekslerinin gelecekteki yönünü tahmin etmede klasik yaklaşımların getirdiği varsayımsal kısıtların üstesinden gelebilen makine öğrenmesi yöntemlerinin kullanılabilirliğinin gösterilmesidir. İkincisi, sadece mikro ve makro düzeyde ilişkilerin incelendiği diğer çalışmalara kıyasla bu çalışmada serilerin meta bilgisinden çıkartılan ve teknik göstergeler olarak tanımlanan özniteliklerin tahmin sürecine nasıl dâhil edildiğinin gösterilmesidir. Tanımlanan amaçlar doğrultusunda, çalışmada teknik göstergelerin makine öğrenmesi modellerine entegre edilerek serinin değişim yönünü yakalamada yüksek doğruluk ile tahminler üreten bir yaklaşım önerilmiş ve bu yaklaşım en fazla sayıda şirketi bünyesinde bulunduran ve 31.12.2008 tarihinden itibaren hesaplanmaya başlanan İstanbul şehir endeksi (XSIST) serilerine uygulanmıştır.

Çalışmanın izleyen bölümleri şu şekilde tasarlanmıştır. Çalışmanın ikinci bölümünde bölgesel endeks kavramından kısaca bahsedilmiş ve sonrasında bu endeksler ile yapılan hemen hemen tüm çalışmalar ve bu çalışmalarda kullanılan yöntemler irdelenmiştir. Çalışmanın dördüncü bölümünde çalışmada kullanılan veri seti, çalışmanın yöntem ve metodolojisi ele alınmış, beşinci bölümde ise uygulama sonuçları ele alınmıştır. Çalışmanın altıncı ve son bölümünde sonuçlara ve bazı önerilere yer verilmiştir.

2. BÖLGESEL ENDEKS

Bölgesel endeks, ilgilenilen bir coğrafi bölgenin spesifik özelliklerini yansıtan bir portföy olma özelliği taşımakta ve uzun yıllardan bu yana farklı amaçlar doğrultusunda hesaplanmaktadır. Standard & Poors tarafından hesaplanan ve ABD’de bulunan farklı eyaletlerdeki şehirlerin gayrimenkul fiyatlarının dinamiklerini yansıtan Case-Shiller Home Price Indeks, bölgesel bazda hesaplanan endekslere örnek olarak gösterilebilir. 2008 yılında hesaplanmaya başlanan ve ticari faaliyetler, beşeri sermaye, bilgi değişimleri ve kültürel tecrübe olmak üzere 4-boyutlu göstergesi bulunan Global Cities Index, küresel çapta bir gündem oluşumuna olanak tanımaktadır.

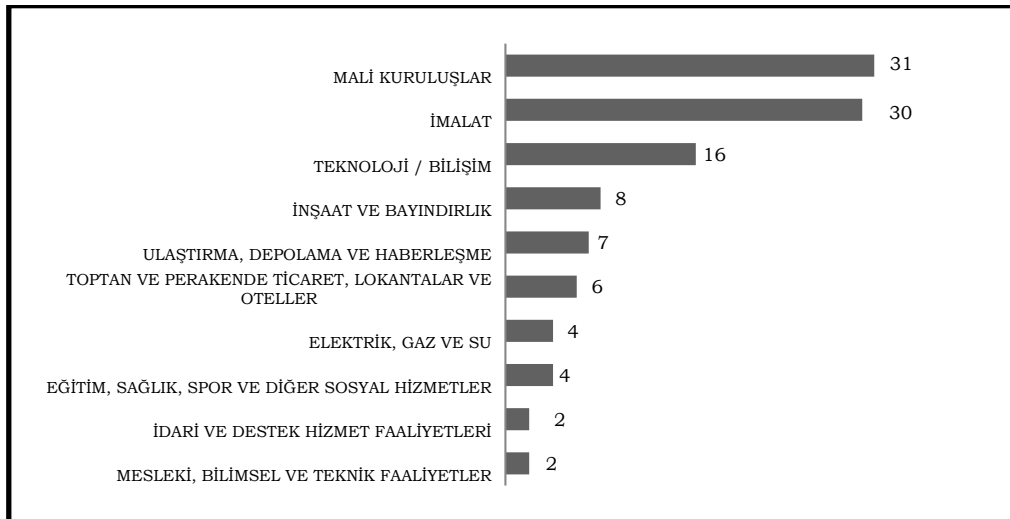
Dünyanın her yerinden ticari işletmeleri ve yaratıcı bireyleri çekmek amacı ile 2008 yılında hesaplanmaya başlanan Global City Index Power ise temelde 6-boyuttan (ekonomi, araştırma-geliştirme, kültürel etkileşim, yaşanabilirlik, çevre, erişilebilirlik) oluşan ve her yıl hesaplanarak sıralamanın revize edildiği bir küresel şehir endeksidir (Güneş, 2019, s.118). Ekonomik güç, beşeri sermaye, kurumsal etkinlik, finansal olgunluk, küresel çekicilik, fiziki sermaye, sosyal-kültürel ve çevresel karakteristikler gibi farklı boyutların ağırlıklı ortalamaları ile hesaplanan bir diğer küresel endeks Global City Competitiveness Index, EIU tarafından hesaplanmakta ve 120 şehir için bir rekabet sıralaması sunmaktadır. Bölgesel endeksler, şehirlerin sadece finansal, beşeri ve fiziki kaynaklarını yansıtmakla değil aynı zamanda çevresel durumunu da karşılaştırmak amacı ile hesaplanmaktadır. Örneğin European Green City Index, farklı göstergelerle 30 şehrin çevresel sürdürülebilirliğini ölçmeyi amaçlayarak ilgili şehirlerin çevre politikalarının değerlendirilmesi ve geliştirilebilmesine katkıda bulunur (Meijering ve diğerleri, 2014, s.132).

2.1. BİST Şehir Endeksleri

BİST, 2009 yılında 9 şehrin finansal performanslarını yansıtmayı amaçlayarak şehir endeksi hesaplamaya başladığını duyurmuştur. Günümüzde Adana (XSADA), Ankara (XSANK), Antalya (XSANT), Aydın (XSAYD), Balıkesir (XSBAL), Bursa (XSBUR), Denizli (XSDEN), İstanbul (XSIST), İzmir (XSIZM), Kayseri (XSKAY), Kocaeli (XSKOC), Konya (XSKON) ve Tekirdağ (XSTKR) olmak üzere 13 adet şehir endeksi hesaplanmaktadır.³

2.2. İstanbul Şehir Endeksi

İstanbul, Türkiye'nin en aktif ticari ve finansal ilişkilerin bulunduğu şehirlerin başında gelmektedir. 2009 yılında hesaplanmaya başlanan XSIST, günümüzde 110 şirketin bünyesinde bulunduran bir portföydür. Bu şirketlerin büyük bir kısmını %28,1 ile mali kuruluşlar oluştururken onu %27,2 ile imalat sektörü takip etmektedir (bkz. Şekil 1). En az payı bulunan sektörler ise; mesleki, bilimsel ve teknik faaliyetler ile idari ve destek hizmet faaliyetleri sektörleridir.



Şekil 1. XSIST Kapsamında Yer Alan Şirketlere Ait Frekanslar

³ Bir şehrin şehir endeksi kapsamına girebilmesi için bazı kriterler bulunmakta olup ilgili kriterlere BİST portalından ulaşılabilir.

3. LİTERATÜR ARAŞTIRMASI

Şehir endeksleri ile ilgili yapılan kısıtlı sayıdaki çalışmalar incelendiğinde araştırmacıların çoğunlukla oynaklık üzerine yoğunlaştığı görülmektedir. Öte yandan bahsedilen çalışmaların çoğunluğunun modelleme sürecinde makroekonomik göstergelerden faydalandığı görülmektedir.

Bayramoğlu ve Pekkaya (2010) şehir endeksleri ile yaptıkları çalışmada özetleyici istatistiksel analizler kullanmışlardır. Riskin ve oynaklığın en yüksek olduğu şehir endekslerinin XSANT, XSKOC, XSANK ve XSBUR iken oynaklığın en düşük olduğu endeksin XSADA olduğunu raporlamışlardır. Aksoy (2013) çalışmasında şehir endekslerini kukla değişken olarak ele almıştır. Yabancı yatırımcıların yatırım tercihlerinde ilgili firmanın şehir endeksleri içinde yer almasının belirleyici bir faktör olup olmadığını araştırıldığı regresyon analizinde kukla değişkenler istatistiksel açıdan anlamlı bulunmamış olup katsayılarının pozitif çıkması, bu firmaların büyük şehirlere yatırım yapma eğiliminde olduğu sonucuna ulaşmıştır. Çakır (2016), finansal performans ölçüm yöntemleri ile XSTKR, XSBUR ve XSKOC endekslerine ait performansın yüksek olduğu ve bu endekslere yatırım yapılabileceği sonucuna ulaşmışlardır. Kula ve Baykut (2018), çalışmalarında simetrik ve asimetrik ARCH modellerini (ARCH, GARCH, TGARCH, EGARCH ve PARCH) kullanmış ve oynaklığın en yüksek bulunduğu endeks XSANT iken en stabil XSKOC olduğunu belirtmişlerdir. Ayrıca şehir endekslerinin düşük rejimi hızlıca terk etme eğiliminde olduklarını belirtmişlerdir. Davaslıgil Atmaca (2018) ise çalışmasında ham petrol, Türk lirası ve döviz kuru değişkenlerini kullanarak Dinamik Koşullu Korelasyon GARCH Modelini (DCC-GARCH) ele almıştır. Ham petrol, Türk lirası ve döviz kuru değişkenleri ile ham petrol ve şehir endeksleri hareketinin (XSANT dışında) pozitif korelasyonlu olduğunu belirtmiştir. Yapraklı ve diğerleri (2018) çalışmalarında simetrik ve asimetrik ARCH Modellerinden GARCH, EGARCH ve TGARCH kullanarak XSANT için GARCH(1 1), XSBAL, XSBUR, XSIST, XSKAY, XSKOC ve XSTKR için EGARCH(1 1), XSADA, XSANK ve XSIZM için TGARCH(1 1) modellerinin uygun olduklarını belirtmişlerdir. Bayrakdaroğlu ve Tepeli (2018) çalışmalarında Risk-Getiri Analizi uygulamış ve dışsal/içsel değişken olarak beklenen getiri oranı, beta değerleri ve risksiz faiz oranını ele almışlardır. Çalışmanın sonucunda XSKAY endeksinin beklenen getirisinin en yüksek olduğu, XSKON ise en düşük olduğu bulunmuştur. Şehir endekslerinin çoğunun BIST 100 endeksi ile kıyaslandığında ortalama getiri açısından daha fazla getiri sağladığı bulunmuştur. Kayral ve Tandoğan (2019) çalışmalarında kukla değişken (ay içi ve ay dönümü anomalisi) kullanmış ve oynaklığı simetrik ARCH ile modellemişlerdir. XSANK, XSIZM, XSBUR ve XSANT endekslerinde ay içi, XSIST, XSANK ve XSIZM endekslerinde ay dönümü anomalisi tespit edilmiştir. Aşkın (2019) çalışmasında saklı Markov modelini kullanarak XSIST'in artan getiri davranışının, döviz kuru, faiz oranı, para arzı ve tüketici fiyat endeksinin artan davranışları ile açıklanabildiğini göstermiştir. Gürsoy ve diğerlerinin (2020) çalışmalarında çok değişkenli VAR- EGARCH kullanmış ve XSANT endeksinin dolar ve euro fiyatlarındaki volatiliteden etkilendiğini göstermişlerdir. Aşkın (2020) ise farklı koşullu olasılık dağılım türleri ile simetrik ve asimetrik ARCH Modellerini kullanarak XSANT dışında tüm şehir endeksleri için asimetrik modellerin uygun olduğunu belirtmiştir. Çalışmada endekslerin kalın kuyruklu dağılım yapısı sergiledikleri ve bu sebeple de Student-t dağılımının Normal dağılıma kıyasla daha başarılı olduğu gösterilmiştir. Kayral (2020) çalışmasında ARDL Sınır Testini kullanmış ve XSIST ve XSIZM endeksleri ile döviz kuru arasında uzun dönemli ilişkinin olduğunu göstermiştir. Özkan ve Ulaş (2021) ise yine aynı yönteme dayalı ve bölgesel COVID-19 vakaları, gram altın fiyatı ve döviz kuru değişkenleri ile bir çalışma yürütmüşlerdir. Çalışmada vaka sayılarındaki artış ile şehir endeksleri fiyatları arasındaki pozitif ilişki gösterilmiştir. XSIST ve XSIZM endekslerinin altın fiyatı ile negatif, döviz kuru ile pozitif uzun dönemli ilişkisi olduğu tespit edilmiştir.

4. YÖNTEM VE METODOLOJİ

Bu çalışma kapsamında bölgesel kalkınmanın önemli göstergelerinden biri olan ve BIST tarafından hesaplanan İstanbul şehir endeksinin gelecekteki davranışının tahminine yönelik bir sınıflandırma tabanlı modelleme süreci önerilmiştir. Önerilen süreç, öznitelik mühendisliği yöntemlerine kapsamlı bir bakış açısı sunmaktadır. Modelleme sürecinin ilk aşaması, XSIST zaman serisinin meta-bilgisinden yararlanarak teknik göstergelere dayalı öznitelikler çıkartmadır. İkinci aşaması, modele katkı sağlayacak, en az bilgi kaybı ile en sade modeli kurmada kullanılacak özniteliklerin seçilmesidir. Üçüncü aşama, makine öğrenmesi teknikleri ile modelleri tahmin etmektir. Dördüncü ve son aşama, modellerin istatistiksel metriklere göre performanslarını karşılaştırmak ve genelleştirilebilirliğinin incelendiği bazı tanısal bazı testleri yapmaktır.

4.1. Öznitelik Çıkartma

Finansal tahmin amacı ile kullanılan hesaplamalı yaklaşımlara dayalı çalışmaların yaklaşık %20'si modelleme sürecinde girdi olarak kullanılan öznitelik vektörünü Teknik Analizden (TA) faydalanarak oluşturmaktadır (Shynkevich ve diğerleri, 2017, s.71). TA ile hesaplanan her bir gösterge, bir varlığın fiyatından elde edilen ek bilgileri kullanır. Bu göstergeler varlığa ait açılış değeri, kapanış değeri, en yüksek/en düşük değer ve işlem hacmi bilgilerini kullanarak her işlem günü için hesaplanır. Her işlem günü için hesaplama yapılabilmesi için tek değişkenli serinin çok değişkenli hale dönüştürülmesi gerekmektedir. Bu amaç doğrultusunda seriye pencere kaydırma (rolling) işlemi yapılmaktadır. Literatürde kullanılan TA tabanlı göstergeler incelendiğinde hesaplamaların temelde 4 ana başlık altında toplandığı görülmektedir (hacme, oynaklığa, trende ve momentuma dayalı). Çalışmada 38 farklı gösterge ele alınmış olup Tablo 1'de gösterilmiştir.

Tablo 1. Çalışmada Kullanılan Teknik Göstergeler

Hacim Tabanlı	Oynaklık Tabanlı	Trend Tabanlı	Momentum Tabanlı
Accumulation/ Distribution Index (ADI)	Average True Range (ATR)	Moving Average Convergence Divergence (MACD)	Relative Strength Index (RSI)
Chaikin Money Flow (CMF)	Bollinger Bands (BB)	Moving Average Signal Line (MACD)	Stochastic RSI
Force Index (FI)	Keltner Channel (KC)	MACD Histogram	Stochastic RSI %k
Money Flow Index (MFI)	Donchian Channel (DC)	Simple Moving Average (SMA)	Stochastic RSI %d
Ease of movement (EoM, EMV)	Ulcer Index (UI)	Exponential Moving Average (EMA)	True strength index (TSI)
Volume-price trend (VPT)	-	Average Directional Movement Index (ADX)	Ultimate Oscillator (UO)
Negative Volume Index (NVI)	-	Mass Index (MI)	Stochastic Oscillator (SR)
Volume Weighted Average Price (VWAP)	-	Commodity Channel Index (CCI)	Signal Stochastic Oscillator
-	-	Detrended Price Oscillator (DPO)	Awesome Oscillator (AO)
-	-	KST Oscillator (KST Signal)	Kaufman's Adaptive Moving Average
-	-	Ichimoku Kinkō Hyō (Ichimoku)	Rate of Change (ROC)
-	-	-	The Percentage Price Oscillator (PPO)
-	-	-	Percentage Price Oscillator Signal Line
-	-	-	Percentage Price Oscillator Histogram

4.2. Öznitelik Seçme

Öznitelik seçme işleminde temel amaç modellere girdi olarak seçilecek özniteliklerin belirlenmesidir. Öznitelik çıkartma sürecinde hesaplanan özniteliklerin tamamı modele katkı sağlamayabilir. Bu durumda modele katkı sağlayacak, çıktıyı anlamlı şekilde etkileyecek ve modelde yer almadığında bilgi kaybına yol açacak öznitelikleri seçmek ve bu özellikleri sağlamayan öznitelikleri kapsam dışı bırakmak gerekmektedir. Öznitelik seçme uygulandığında hem modelin karmaşıklığı azaltılabilir hem de sınıflandırma sürecinde yaşanabilecek muhtemel bir aşırı uyum (overfitting) probleminden kaçınılabilir.

Öznitelik seçme teknikleri genel olarak filtreleyici (filtering), sarmal (wrapper) ve gömülü (embedded) olmak üzere 3 kategoride incelenebilir. Tüm bu tekniklerin nihai hedefi en yüksek doğrulukla en az sayıda özneliğe bağlı sade bir model elde etme çabasında istatistiki açıdan anlamlı öznitelikleri belirlemektir. Her ne kadar sarmal tekniklerin hesaplama süresi diğer tekniklere kıyasla daha yüksek olsa da, tek başına bilgi sağlamayan ancak diğer öznitelikler ile beraber modele dâhil edildiğinde yüksek performans gösteren özniteliklerin kapsam dışı edilmesinin önüne geçer. Bu teknikte model performansının en yükseklenmesi hedeflenerek yinelemeli işlem mantığı ile en uygun öznitelik alt kümesi aranır. Sarmal teknikler arasında ileriye doğru seçim (forward selection), geriye doğru eleme (backward elimination) ve yinelemeli öznelik eleme (recursive feature elimination-RFE) en sık kullanılan yöntemler arasındadır. İleriye doğru seçimde başlangıçta öznelik uzayı boş bir kümedir. Tüm öznitelikler önem seviyesine göre sıralanır. Çıktı ile en yüksek ilişkiye sahip ilk öznelikten başlanarak öznelik uzayı diğerlerinin teker teker eklenmesi ile belirlenir. Geriye doğru elemelerde ise süreç tam tersi işler. Başlangıçta öznelik uzayında tüm öznitelikler bulunmakta ve modele katkı sağlamayan öznitelikler bu uzaydan teker teker dışlanmaktadır. Yinelemeli öznelik eleme yöntemi ise en iyi performans gösteren alt kümeyi bulmayı amaçlayan bir açgözlü arama (greedy search) optimizasyon algoritmasıdır. Bu yöntemde optimal öznelik sayısına ulaşılan kadar en zayıf öznelik (veya öznitelikler) ortadan kaldırılarak model kurulur. Farklı öznelik altkümeleri için modelin performansını bulmak ve en yüksek performans gösteren özniteliklerin koleksiyonu seçmek için çapraz doğrulama yöntemi kullanılır ve böylece optimal öznelik sayısı belirlenmiş olur.

Bu çalışmada, RFE yöntemi ele alınmıştır. Yöntemin uygulanması, sonuçların elde edilmesi ve görselleştirme işlemi için Python Yellowbrick kütüphanesi kullanılmıştır. Yöntemin performansı, büyük ölçüde özniteliklerin önem sıralanmasında kullanılan sınıflandırıcıya bağlı olmakla birlikte çalışmada Rasgele Orman (RO) sınıflandırıcısı kullanılmıştır.

4.3. Sınıflandırma

Öznitelik seçme işleminden sonra XSIST şehir endeksinin gelecekteki hareketinin tahmin edilmesinde farklı makine öğrenmesi algoritmaları ele alınmıştır. Bu algoritmalar izleyen bölümde kısaca özetlenmiştir. Çalışmada kullanılan ve aynı zamanda hemen hemen tüm makine öğrenme modellerinin kendine has parametreleri olmakta ve modelin eğitilmesi esnasında farklı optimizasyon teknikleri kullanılarak modellere ait parametrelerin optimal değerlerinin bulunması amaçlanmaktadır. Modele ait hiperparametrelerin optimizasyonu, aşırı uyum ve az uyum (underfitting) dengesini sağlamak amacıyla yapılır. Bu çalışmada kullanılan ve yukarıda da bahsi geçen makine öğrenmesi algoritmaları için hiperparametre optimizasyonu ızgara arama (grid search) tekniği ile yapılmıştır. ızgara aramada önceden belirlenen hiperparametre değerlerinin tüm kombinasyonları için farklı modeller denenir ve spesifik bir performans metriğine en iyi sonuçlar üreten hiperparametreler bulunur. Araştırmacı, denenmesini istediği hiperparametre sayısını arttırdıkça denenilen her model çapraz doğrulama ile test edildiği için işlem maliyeti

oldukça yükselecek ancak buna paralel olarak yüksek sayıda kombinasyonlar denendiğinden en iyi performans veren hiperparametreleri bulma şansı artacaktır.

4.3.1. Lojistik Regresyon (LR)

LR, klasik doğrusal regresyondan farklı olarak ikili sonuç veren çıktının modellenmesinde kullanılır. İkili durum söz konusu olan modelde genellikle 0 veya 1 olarak tanımlanmaktadır. Lojistik fonksiyon $-\infty/+\infty$ aralığındaki tüm değerleri girdi olarak kabul edebilmektedir. LR'de eğitim verisinde yer alan örneklerle beraber bir sınıflandırma modeli oluşturulmakta ve elde edilen örnekler en yüksek olasılık değerine sahip sınıfa atanmaktadır.

4.3.2. Destek Vektör Makineleri (DVM)

DVM, 1990'lı yıllarda Vapnik ve ekibi tarafından geliştirilmiş dağılımdan bağımsız bir denetimli öğrenme yöntemidir. Doğrusal ve doğrusal olmayan verileri sınıflandırmada kullanılan denetimli bir makine öğrenmesi algoritmasıdır. DVM'lerde, eşleme yöntemi ile veriler üst bir boyuta dönüştürülür. Bu yeni oluşturulan boyutta, veriyi en uygun halde ayırması beklenen karar sınırı adında bir üst düzlem bulunması amaçlanır. DVM, genelleştirme yeteneklerinin yanı sıra uç değerler içeren verilere karşı dayanıklı algoritmalarıdır (Onan, 2017, s.7). Doğrusal olmayan vektör makinalarında modelin performansı büyük ölçüde çekirdek fonksiyonunun seçimine ve model parametre değerlerine bağlıdır. Doğrusal dönüşüm için kullanılan farklı çekirdek fonksiyonları vardır. RBF (Radial Basis Function), lineer ve poly (polynomial) sıkça kullanılan çekirdek fonksiyonları olup "No Free Lunch" (NFL) prensibine göre birbirlerine kıyasla üstünlükleri araştırmacının üzerinde çalıştığı veri setine bağlıdır. Çalışmada bahsi geçen 3 çekirdek fonksiyonu da ele alınmıştır.

4.3.3. Rasgele Orman (RO)

RO, Leo Breiman tarafından geliştirilen bir kolektif öğrenme yöntemidir. Birden çok karar ağacının birleşiminden oluşan bir modeldir. Veriler N adet karar ağacı üzerinde işlenir ve ortaya çıkan tahminlerin ortalaması alınarak doğru tahminler üretilmeye çalışılır. Geleneksel karar ağaçları yöntemlerinde sıklıkla karşılaşılan aşırı uyum sorununu verileri parçalara ayırıp birden fazla ağaçta işleyerek ortadan kaldırmaya çalışır (Sevli, 2019, s.179). RO'da örneklem, Bootstrap tekniğiyle oluşturulur. Orijinal veri setinden elde edilen Bootstrap örnekleri eğitim veri seti ve test veri seti olarak ikiye ayrılır. Eğitim veri setine tahminler uygulanarak bir topluluk oluşturulur. Tahminlerden bilgi kazancı en yüksek olan ile dallara ayrılma gerçekleşir. RO algoritmaları, diğer makine öğrenmesi tekniklerine kıyasla yanlılığı oldukça düşük sonuçlar üretir.

4.3.4. K- En Yakın Komşu (k-NN)

k-NN, benzerlik öğrenilmesi tabanlı bir yöntemdir. Eğitim örnekleri n-boyutlu sayısal niteliklerle tanımlanır. Her örnek, n-boyutlu uzayda bir noktayı gösterir. Böylece eğitim örnekleri n-boyutlu uzayda depolanır. Uygulaması kolay olan bu denetimli öğrenme algoritması türü, hem sınıflandırma hem de regresyon problemleri için kullanılır. Algoritmanın çalışmasında araştırmacı tarafından bir k-değeri atanır. En yakın k-kadar eleman alınarak uzaklık hesaplanır. Uzaklıkların hesaplanması işleminde çoğunlukla Öklid fonksiyonu kullanılırken ek olarak Manhattan, Minkowski ve Hamming fonksiyonları da tercih edilebilir. Uzaklık hesabı yapıldıktan sonra sıralanır ve en uygun olan sınıfa atama gerçekleştirilir.

4.3.5. AdaBoost

Boosting kavramı, tahminlerin paralel olarak birbirinden bağımsız değil de sırayla gerçekleştiği bir yaklaşımdır. Bu yaklaşımda tahminler önceki tahminlerde meydana gelen hataları

kullanır. AdaBoost Freund ve Schapire (1996) tarafından geliştirilen ilk başarılı boosting algoritması olarak bilinir. Genellikle, karar ağaçları modellemesi için kullanılır. Her biri son modeldeki hataları düzelteren çoklu sıralı modeller oluşturulur. AdaBoost algoritmasında yanlış tahmin edilen gözlemlere ağırlık atanır ve ardından oluşturulan model bu değerleri doğru şekilde tahmin etmek için çalışır.

4.3.6. Gradyan Boosting (GB)

Breiman tarafından 1997 yılında geliştirilen sınıflandırma ve regresyon problemlerinde kullanılan makine öğrenmesine dayalı bir topluluk algoritmasıdır. GB algoritmasında amaç zayıf tahminleri iterasyonlar ile aşamalı olarak güçlü tahminlere dönüştürmektir ve tahminler sıralı olarak gerçekleşir (Keleş ve diğerleri, 2020, s.74).

4.4. Model Performanslarının Karşılaştırılması

Makine öğrenmesi modelleri kurulduktan sonraki aşama sınıflandırma performanslarının karşılaştırılmasıdır. Sınıflandırmalara ait model başarıları, karmaşıklık (confusion) matrisi kullanılarak hesaplanan metriklere dayanır. Bu metrikler arasında en sık kullanılanları doğruluk (DO-accuracy), özgüllük (specificity), duyarlılık (sensitivity) ve F1-skordur. Tablo 2, tipik bir karmaşıklık matrisinin yapısını göstermektedir. Karmaşıklık matrisinde yer alan TP, TN, FP ve FN değerlerine göre adı geçen performans metrikler aşağıda verilen formüller yardımı ile hesaplanmaktadır.

Tablo 2. Karmaşıklık Matrisi

Gerçek Durum	Tahmin Edilen Durum	
	+1	-1
+1	Gerçek Pozitif (True Positive-TP)	Yanlış Negatif (False Negative-FN)
-1	Yanlış Pozitif (False Positive-FP)	Gerçek Negatif (True Negative-TN)

$$\text{Doğruluk} = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

$$\text{Özgüllük} = \frac{TN}{FP+TN} \quad (2)$$

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1-skor} = 2 \times \frac{\left(\frac{TP}{TP+FP}\right) \left(\frac{TP}{TP+FN}\right)}{\frac{TP}{TP+FP} + \frac{TP}{TP+FN}} \quad (4)$$

Yukarıda verilen performans metrikleri haricinde çalışmada ROC (Receiver Operating Characteristic) eğrisi altında kalan alan (AUC) ve hata değerleri hesaplanmıştır. ROC eğrisinde bulunan her nokta bir karar eşiğine göre x-ekseninde (1-özgüllük), y-ekseninde ise duyarlılık değerlerini göstermektedir. Ayrıca eğitim ve test setlerinin tahmin doğrulukları kullanılarak her bir modelin hatası hesaplanmıştır. Eğitim ve test setlerine ait hata değerleri sırasıyla $\text{hata}_{\text{eğitim}} = 1 - \text{DO}_{\text{eğitim}}$ ve $\text{hata}_{\text{test}} = 1 - \text{DO}_{\text{test}}$ olmak üzere $\text{hata} = \text{hata}_{\text{test}} - \text{hata}_{\text{eğitim}}$ olarak hesaplanmıştır.

5. UYGULAMA

Bu başlık altında uygulamaya dair bilgilere yer verilmiştir.

5.1. Veri Seti

Çalışmada kullanılan veri seti <http://tr.investing.com> adresinden ve BIST elektronik dağıtım sisteminden çekilmiştir. İstanbul şehir endeksine ait hacim bilgisine 13.11.2009 tarihinden itibaren ulaşılabilmesi sebebiyle veri setinde 13.11.2009 ve 22.04.2021 tarihleri arasında

toplamda 2878 güne ait açılış, kapanış, en yüksek, en düşük değer ve hacim değerleri bulunmaktadır. Tablo 3, ilgili seriye ait bazı özetleyici istatistikleri göstermektedir. Basıklık testine göre serinin leptokurtik dağılım sergilediği görülmektedir. Normallik varsayımı XSIST serisi için geçerli değildir.

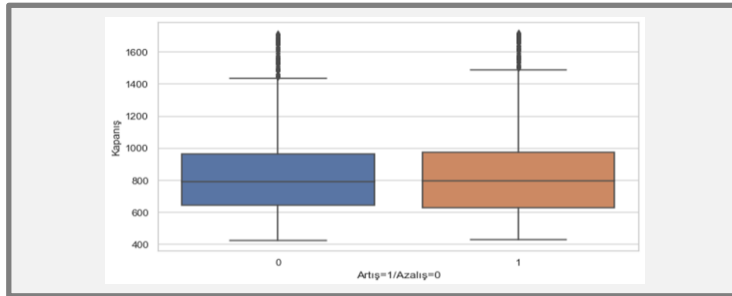
Tablo 3. Özetleyici İstatistikler

XSIST	Açılış	Kapanış	En Düşük	En Yüksek	Hacim
Ortalama	828.88	828.13	821.44	834.70	2.68+08
Standart Sapma	248.28	247.64	245.71	249.97	1.78+08
Medyan	792.36	791.20	785.33	797.46	2.34+08
Basıklık	1.3754***	1.3739***	1.3533***	1.3768***	3.1192***

5.2. Adımsal Süreç

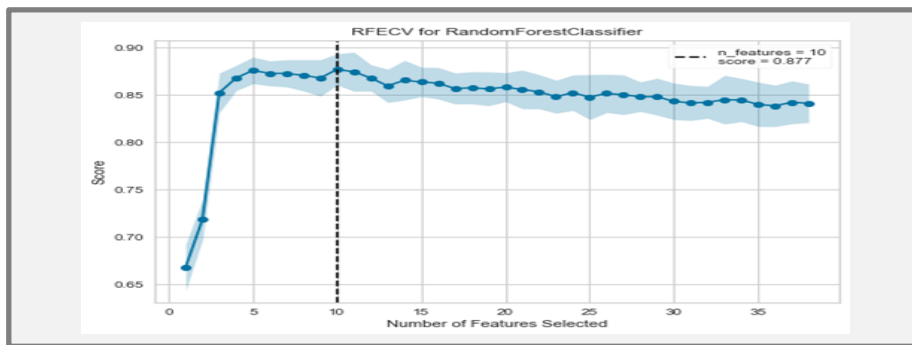
Adım 1: Zaman serilerinden matematiksel fonksiyonlar ile istatistiksel veya teknik öznitelikler çıkartmada Python programlama dilinde yer alan farklı kütüphaneler bulunmaktadır. Teknik göstergelerin hesaplanmasında, TA 3 kütüphanesi kullanılmıştır.

Özniteliklerin hesaplanmasından sonra makine öğrenimi algoritmalarının tahmin edeceği ve çıktı değişkeni betimleyen yeni bir sütun vektörü oluşturulmuştur. Bu çalışmada ikili sınıflandırma problemi ele alınmış olup XSIST endeksine ait ilgili günün kapanış fiyatı bir önceki güne kıyasla artış halinde olması durumunda +1, tersi durumda ise -1 ile etiketlenmiştir. Şekil 2 ile 1564 adet artış ve 1314 adet azalış olarak etiketlenen ham veri setine ait kutu grafikleri verilmiştir.

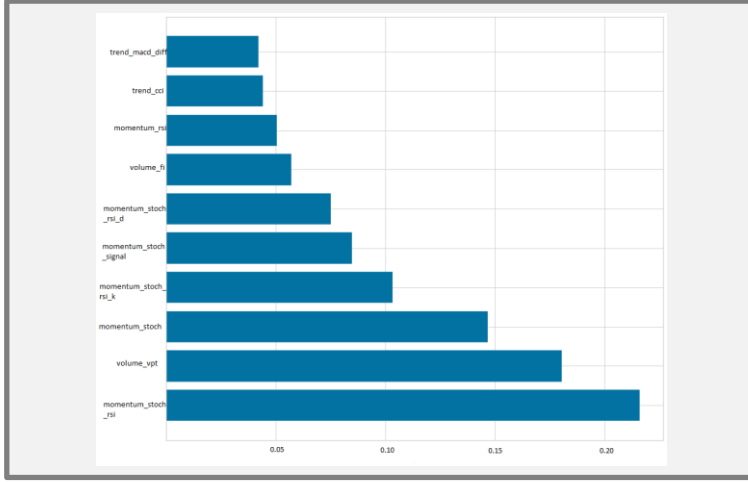


Şekil 1. XSIST Kapanış Değerleri

Adım 2: Çıkarılan özniteliklerin ölçeklemesinde min-max dönüşümü uygulanmış ve veri seti 0,3 oranında test setine ayrılmıştır. RFE tabanlı RO öznitelik seçim işlemine göre, bir sonraki aşamada öğrenme modellerine girdi olarak alınması önerilen öznitelik sayısı ve İstanbul şehir endeksinin bir önceki güne göre artış/azalışını etkileyen özniteliklerin önem sırası sırasıyla Şekil 3 ve Şekil 4'te verilmiştir.



Şekil 3. Önerilen Öznitelik Sayısı



Şekil 4. Özniteliklerin Önem Sırası

Sınıflandırma modellerine girdi olarak alınacak öznitelik sayısına ait optimal değer 10 olduğu görülmektedir. Önem seviyesi en yüksek gösterge, hız göstergesi olarak da bilinen stokastik RSI osilatörüdür (Stochastic Relative Strength Index). Momentum tabanlı bir teknik gösterge olup karar vericilere mevcut RSI değerinin aşırı alım veya aşırı satım olup olmadığı konusunda fikir vermektedir. Ayrıca %k (hızlı stokastik) ve %d (yavaşlatılmış stokastik) çizgisel değerleri de seçilen öznitelik uzayında bulunmaktadır. Hacim fiyat trendi (Volume Price Trend), ikinci en önemli gösterge olmakla beraber hacim temelli göstergeler arasında yer alır. Bu gösterge bir talep ve arz arasındaki denge hakkında yatırımcıya fikir verir. Gösterge, menkulün yukarı veya aşağı hareketine bağlı olarak, bir hisse fiyatının trendindeki ve mevcut hacmindeki yüzde değişiminin katlarını ekleyen veya çıkaran kümülatif bir hacim çizgisinden oluşur (www.investopedia.com). Öznitelik uzayına seçilen bir diğer momentum tabanlı gösterge stokastik osilatördür (Stochastic). Bu gösterge, fiyat momentumunu ölçer ve geçmiş dönem tarafından tanımlanan kapanış fiyatına dayanır. Stoch doğrudan fiyata dayalı iken StochRSI, RSI'nın kapanış fiyatına dayalıdır ve dolayısıyla RSI'nın momentumunu ölçer. Diğer bir ifade ile bir göstergenin göstergesidir. Stokastik Sinyal osilatörü de (Stochastic Signal) öznitelik vektöründe bulunan bir diğer göstergedir. Bir diğer hacim tabanlı gösterge güç indeksidir (Force Index). Gerçek alış veya satış baskısının ne kadar güçlü olduğunu gösterir. Yüksek pozitif değerler güçlü bir yükseliş eğilimini, düşük değerler ise güçlü bir düşüş eğilimi olduğunu işaretidir. Bağlı güç endeksi (Relative Strength Index) bir menkul kıymetin belirli bir zaman aralığındaki fiyat hareketlerinin hızını ve değişimini ölçmek için kullanılan bir diğer momentum tabanlı teknik göstergedir. Emtia kanalı endeksi (Commodity Channel Index) ve hareketli ortalama yakınsama (Moving Average Convergence Divergence) öznitelik vektöründe bulunan trend tabanlı teknik göstergelerdir. CCI, belirli bir süre boyunca ortalama fiyat seviyesine göre mevcut fiyat seviyesini ölçerken MACD fiyatların iki hareketli ortalamaları arasındaki ilişkiyi göstermektedir.

Yukarıda ayrıntısıyla bahsedilen teknik göstergelerle oluşturulan öznitelik vektöründe çoğunlukla momentum tabanlı göstergelerin bulunduğu görülmektedir. Hacim tabanlı göstergeler VPT ve FI olmak üzere XSIST endeksine ait hareketi anlamlı bir şekilde etkileyen ve trend tabanlı olan iki gösterge CCI ve MACD olarak bulunmuştur. Oynaklık tabanlı göstergelerin hiçbirinin öznitelik seçme yöntemiyle oluşturulan öznitelik vektöründe bulunmaması dikkat çeken bir bulgudur.

Adım 3: Çalışmada kullanılan sınıflandırma performanslarının değerlendirilmesi ve kıyaslaması sürecine geçmeden önce modellere ait hiperparametrelerin optimal setinin belirlenmesi

gerekmektedir. Her bir öğrenme modeli için en iyi sınıflandırmayı sağlayan hiperparametrelerin belirlenmesinde 10-katlı çapraz doğrulama ve ızgara aramasına dayalı bileşik prosedür uygulanmıştır. Eğitim seti üzerinde çalışılarak belirlenen optimal hiperparametrelerle test seti üzerinde modeller yeniden çalıştırılmış ve sonuçlar elde edilmiştir. Her bir öğrenme modeli için ızgara aramasında kullanılan, diğer bir deyişle hiperparametre konfigürasyon arama uzayı Tablo 4'te verilmiştir.

Tablo 4. Hiperparametre Konfigürasyon Arama Uzayı

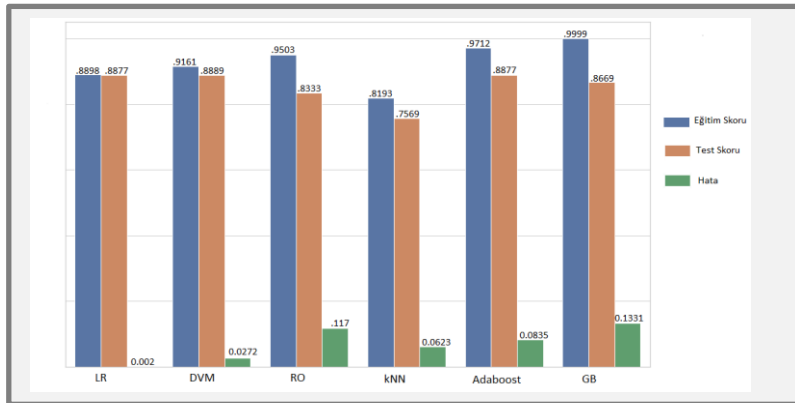
Model	Arama Uzayı
LR	'penalty' : ['l1','l2']; 'tol' : [0.0001,0.0002,0.0003]; 'solver':['liblinear']; 'max_iter':[100,200,300]; 'C' : [0.01, 0.1, 1, 10, 100]; 'intercept_scaling' : [1, 2, 3, 4].
DVM	'kernel' : ['rbf','linear','poly']; 'gamma' : [0.0001, 0.001, 0.01, 0.1, 1,10,50] 'C' : [1, 10, 50, 100, 200, 300]
RO	'max_depth' : np.linspace(1, 10, 10, endpoint=True); 'min_samples_split' : [2, 6, 20]; 'min_samples_leaf' : [1, 4, 16]; 'n_estimators':[100,200,300,400]; 'criterion' : ['gini']
k-NN	'n_neighbors' : [1, 7, 9]; 'weights' : ['uniform', 'distance']; 'algorithm' : ['auto', 'ball_tree', 'kd_tree', 'brute']; 'leaf_size' : [10,20, 30, 50].
Adaboost	'n_estimators' : [10, 50, 100, 300,500]; 'learning_rate' : [0.01, 0.1, 0.5,1]
GB	'loss' : ['deviance']; 'n_estimators' : [100,200,300]; 'max_depth' : [4, 8,16]; 'min_samples_leaf' : [100,150,250]; 'max_features' : [0.3, 0.1]; 'learning_rate' : [0.1, 0.05, 0.01, 0.001]

Adım 4: 10-boyutlu öznitelik vektörü ile öğrenme modellerinin sınıflandırma performansları Tablo 5'de verilmiştir. Çalışmada kullanılan tüm performans metrikleri baz alındığında sınıflandırma performansı en yüksek olan ilk 3 öğrenme modelinin LR, DVM ve Adaboost olduğu görülmektedir. En yüksek ve ikinci en yüksek DO (LR ve Adaboost DO değerleri yaklaşık olarak eşittir) arasındaki fark sadece %0.12 olmaktadır. Bununla birlikte tüm sınıflandırıcılar arasında en düşük hataya sahip sınıflandırıcı LR'dir (%0.204). DVM ve Adaboost modellerinin hataları sırasıyla %2.719 ve %8.347 olarak bulunmuştur. Sınıflandırma sonuçlarında en yüksek ilk 3 performansı gösteren öğrenme modellerinde optimal parametreler LR için C=100, DVM için C=300, gamma=0.1, kernel=poly ve Adaboost için learning rate=0.5, n-estimators=500 olarak bulunmuştur.

Tablo 5. Öğrenme Modellerinin Sınıflandırma Performansları

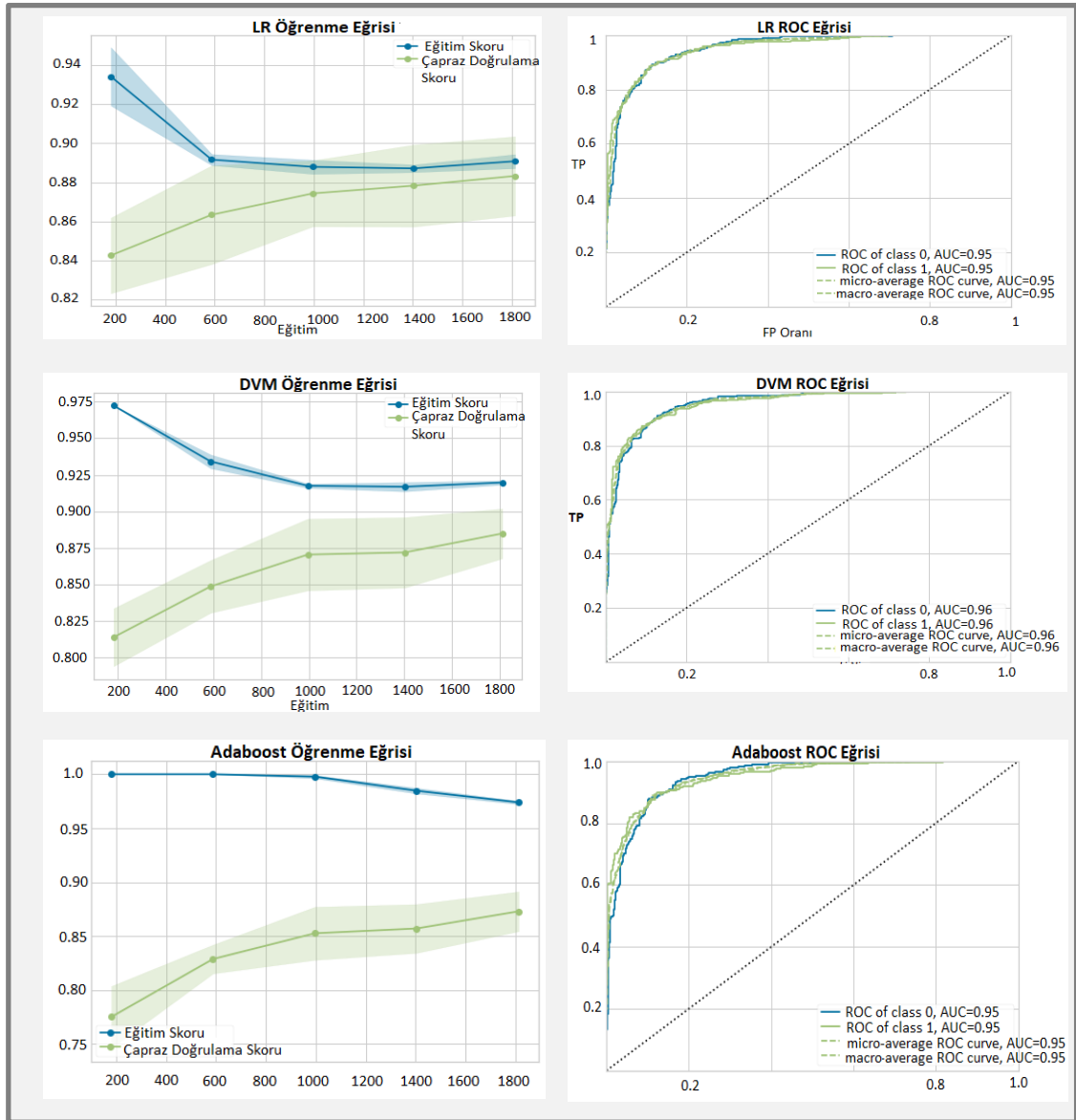
Model	Performans Metrikleri					
	DO	Duyarlılık	Özgüllük	F1 skor	AUC	Hata
LR	0.8877	0.902542	0.869898	0.897787	0.952655	0.002040
DVM	0.8889	0.896186	0.880102	0.898089	0.957470	0.027198
RO	0.8333	0.855932	0.806122	0.848739	0.918983	0.117014
k-NN	0.7569	0.832627	0.665816	0.789157	0.824563	0.062321
Adaboost	0.8877	0.898305	0.875000	0.897354	0.951985	0.083470
GB	0.8669	0.875000	0.857143	0.877790	0.954233	0.133102

Şekil 5'de her model için eğitim ve test setlerine ait DO skorları görsel halde verilmiş olup her bir model için ayrıca hata değerleri de verilmiştir. RO, k-NN, Adaboost ve GB modellerinin hata değerlerinin oldukça yüksek olduğu göze çarpmaktadır.



Şekil 5. Eğitim-Test Seti Skorları ve Hata Değerleri

Şekil 6'da en yüksek DO değerlerine sahip LR, DVM ve Adaboost modellerine ait öğrenme ve ROC eğrileri verilmiştir. Her üç model için de ROC eğrisi altında kalan alan %95'in üzerinde bulunmuştur. Ayrıca karmaşıklık matrislerine göre LR, DVM ve Adaboost modellerinin yanlış sınıflandırdıkları gözlem sayısı birbiriyle hemen hemen eşit olmakla birlikte sırasıyla 97, 96 ve 97 olduğu bulunmuştur. Ancak, makine öğrenme modellerinde olası bir aşırı uyum/az uyum probleminin araştırılması amacıyla kullanılan ve modelin varyansı ve sapması hakkında önemli fikirler veren öğrenme eğrileri incelendiğinde, yüksek DO sahip Adaboost modelinin yüksek varyans ve yüksek sapmalı olduğu görülmektedir. Doğrulama ve eğitim skorları, eğitim sayısı artmasına rağmen birbirine yakınsamamış ve doğrulama skoru eğitim eğrisinin oldukça altında kalmıştır. Yüksek varyanslı bir model, eğitim setinde bulunan gözlemlerin değişimine karşı duyarlıdır ve hatanın büyük ölçüde kullanılan eğitim kümesine bağlı olduğu anlamına gelir. Bu durumda modelde bir aşırı uyum söz konusudur. Yüksek sapmalı bir model ise gözlemlerdeki yapıyı yakalama yeteneğinin düşük olduğunu göstermektedir.



Şekil 6. Modellerin Öğrenme ve ROC Eğrileri

6. TARTIŞMA VE SONUÇ

Şehir endeksleri, bir şehrin çevresel, finansal ve eğitsel gibi birçok farklı kavramlarda gelişmelerin takip edilmesi ve gerekli aksiyonların alınması için özellikle gelişmiş ve gelişmekte olan ülkeler tarafından hesaplanan endekslerdir. BIST tarafından 2009 yılı itibari ile şehirlerin finansal performanslarını incelemek amacı ile hesaplanmaya başlanan şehir endeksleri ilgili şehre yatırım yapmayı planlayan ulusal/uluslararası iştirakler tarafından özenle takip edilmektedir. Günümüzde 13 adet şehir endeksi BIST pay endeksleri altında hesaplanmakta ve ana üretim ve/veya faaliyet merkezinin ilgili şehirde bulunan şirketlerin finansal performanslarını yansıtmaya olanak sağlamaktadırlar.

BIST şehir endeksleri ile ilgili ulusal dizin incelendiğinde hemen hemen tüm çalışmaların bu endekslerin oynaklığının modellenmesi ile ilgili olduğu görülmektedir. Oynaklık, zaman serileri analizi kapsamında otoregresif koşullu değişen varyans (ARCH) tipi modellerle incelenmektedir. Bu çalışmada, ulusal dizindeki çalışmalardan farklı olarak, şehir endekslerine ait gelecekteki davranışın (artış/azalış) tahminlenmesine odaklanılmış ve bu amaca yönelik şehir endeksleri arasında BIST’de işlem gören en yüksek hisse senedine sahip olan İstanbul şehir endeksi verileri kullanılarak bahsedilen amaca yönelik bileşik bir prosedür kullanılmıştır. Çalışmada önerilen yöntem, teknik analiz ve makine öğrenmesi modellerini aynı anda kullanma olanağı sağlamaktadır. Serilerin meta bilgisinden yola çıkılarak hacme, oynaklığa, trende ve momentumu dayalı 38 farklı teknik gösterge hesaplanmış ve öğrenme modellerine girdi olarak girecek öznitelikler RFE tabanlı öznitelik seçim yöntemiyle önem seviyelerine göre sıralanmıştır. Endeks yönünün değişiminde en etkili 10 gösterge ile öğrenme modelleri kurulmuş ve performansları değerlendirilmiştir.

Çalışmanın ana bulgularından birisi XSIST endeks değişim yönünü istatistiksel olarak etkileyen teknik göstergeler arasında oynaklık tabanlı bir göstergenin bulunmamasıdır. Önem seviyesi en yüksek göstergeler çoğunlukla momentumu dayalıdır. Teknik analizde sıkça kullanılan ve hacim tabanlı göstergelerden olan VPT ve FI’nın İstanbul şehir endeksinin değişiminde anlamlı bir etkiye sahip olduğu görülmüştür. Ayrıca trend tabanlı iki gösterge, CCI ve MACD, yürütülen öznitelik seçim prosedürüne göre XSIST endeks yönüne ait artış veya azalışta etkili göstergeler olarak bulunmuştur.

Eğitim verileri kullanılarak ızgara araması ve 10-katlı çapraz doğrulama prosedürlerinin bileşik olarak yürütülmesi ile LR, DVM, RO, k-NN, Adaboost ve GB öğrenme modellerinin optimal parametreleri belirlenmiş ve bu parametrelerle test verileri kullanılarak tahmin sonuçları elde edilmiştir. Çalışmada ele alınan tüm performans metrikleri için k-NN en düşük performansı göstermiştir. Yüksek varyans ve sapmadan muzdarip olan bu öğrenme modelinde daha fazla eğitim verisiyle veya farklı bir k değeriyle çalışılırsa modelin performansında iyileşme olabilir. RO ve GB öğrenme modelleri de benzer şekilde performans metrikleri açısından yüksek sınıflandırma değerlerine sahip olsalar da, modellerin geçerliliğinden bahsedebilmek için daha çok eğitim verisine ihtiyaç duymaktadırlar. En yüksek DO değerlerine sahip LR, DVM ve Adaboost modellerinin öğrenme eğrileri incelendiğinde %88.77 doğrulukla sınıflandırma yapan Adaboost modelinin aşırı uyum problemine sahip olduğu görülmüştür. LR ve DVM öğrenme modellerine ait tüm performans metrikleri karşılaştırıldığında hemen hemen benzer sonuçların bulunduğu görülmüştür. LR öğrenme modelinin hatası %0.204 ile en düşük hataya sahip olmakla birlikte bu oran DVM’de %2.719 olarak bulunmuştur.

Bu çalışma şehir endekslerinin değişim yönü ile ilgilenen gelecek araştırmalar için karşılaştırma yapılabilecek çıktılar sunmaktadır. Şehir endeksleri ilgili şehrin finansal performansına dair bir

görünüm sunmaktadır. Riskten kaçınan yatırımcının geçmiş bilgilerden yola çıkarak geleceği etkin bir şekilde öngörme amacı, beraberinde bu amaca yönelik “karar verme” sürecinde varlığa ait açılış değeri, kapanış değeri, en yüksek/en düşük değer ve işlem hacmi gibi eldeki tüm bilgileri kullanma yoluna götürür. Böylelikle yatırımcı eldeki tüm bilgileri ihmal etmeden kullanarak doğru bir portföy yönetimi sürecini yürütmüş olabilecektir. Bu sebeple de yatırımcı açısından ilgili endeksin değişiminde etkili olan faktörlerin ortaya konulması ve geçerli bir tahmin modeli kurması oldukça önem arz etmektedir. Teknik analizde kullanılan göstergeler ile makine öğrenmesi yöntemlerinin entegre edildiği bu çalışmada bazı kısıtlar bulunmaktadır. Her ne kadar tahmin performansı oldukça yüksek oranda bulunmuş olsa da (~%88), istatistiki açıdan anlamlı olarak şehir endeksleri hareketini etkileyen makro-ekonomik göstergelerin de ele alınması ile tahmin başarısını yükseltmek mümkün olabilir. Ayrıca daha karmaşık yapıdaki farklı öğrenme modelleri kullanılarak da sınıflandırma başarısının arttırılması sağlanabilir. Bunlara ek olarak bu çalışmada sadece İstanbul şehir endeksine dayalı bir uygulama yapılmıştır. Gelecek çalışmalarda tüm şehir endeksleri için hangi göstergelerin ortak olarak anlamlı bulunduğu ve hangi öğrenme modellerinin çoğunlukla yüksek başarı gösterdiği araştırılacaktır.

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemiştir.

Finansal Destek: Bu çalışma Yıldız Teknik Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından FYL-2021-4436 nolu proje kapsamında desteklenmiştir.

Etik Onay: Bu makale, insan veya hayvanlar ile ilgili etik onay gerektiren herhangi bir araştırma içermemektedir.

Yazar Katkısı: Öyküm Esra Yiğit(%50), Merve Karaköse(%50)

Peer-review: Externally peer-reviewed.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: This research was funded by the Yıldız Technical University Scientific Research Projects Coordination Unit as part of the FYL-2021-4436 project.

Ethical Approval: This article does not contain any studies with human participants or animals performed by the authors.

Author Contributions: Öyküm Esra Yiğit(50%), Merve Karaköse(50%)

KAYNAKÇA

- Akcan, A., ve Kartal, C. (2011). İMKB sigorta endeksini oluşturan şirketlerin hisse senedi fiyatlarının yapay sinir ağları ile tahmini. *Muhasebe ve Finansman Dergisi*, 51, 27-40. Erişim adresi: <https://dergipark.org.tr/tr/pub/mufad/issue/35630/396054>
- Aksoy, M. (2013). İstanbul Menkul Kıymetler Borsası'nda finansal kriz döneminde yabancı yatırımcıların hisse senedi tercihlerinin analizi. *İstanbul Üniversitesi Siyasal Bilgiler Fakültesi Dergisi*, 0(48). Erişim Adresi: <https://dergipark.org.tr/tr/pub/iusiyasal/issue/630/6413>
- Altınbaş, H., ve Biskin, O.,T. (2015). Selecting macroeconomic influencers on stock markets by using feature selection algorithms. *Procedia Economics and Finance*, 30, 22-29. doi: 10.1016/S2212-5671(15)01251-4
- Aşkın, Ö.E. (2019). Analysis of changes in İstanbul city index values with hidden Markov model. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 41(2), 319-337. doi: 10.14780/muiibd.665045
- Aşkın, Ö.E. (2020). BIST şehir endekslerine ait volatilitenin modellenmesi. *Muhasebe ve Finansman Dergisi*, (85), 223-242. Erişim Adresi: <https://dergipark.org.tr/en/pub/mufad/article/673733>

-
- Bayrakdaroğlu, A., ve Tepeli, Y. (2018). BİST şehir endekslerinin risk-getiri analizi üzerine bir inceleme. *Muhasebe ve Finansman Dergisi*, (80), 147-160. Erişim Adresi: <https://dergipark.org.tr/tr/pub/mufad/issue/39476/465922>
- Bayramoğlu, M.,F., ve Pekkaya, M. (2010). İMKB tarafından hesaplanan endekslerde yeni gelişmeler ve İMKB şehir endeksleri. *Muhasebe ve Finansman Dergisi*, (45), 200-215. Erişim Adresi: <https://dergipark.org.tr/tr/pub/mufad/issue/35624/395912>
- Boyacioglu, M.A., ve Avcı, D. (2010). An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the İstanbul Stock Exchange. *Expert Systems with Applications*, 37(12), 7908-7912. doi: 10.1016/j.eswa.2010.04.045
- Çakır, Z. (2016). Şehir endekslerinin finansal performanslarının ölçülmesi ve değerlendirilmesi (Yayımlanmamış Yüksek Lisans Tezi). Hitit Üniversitesi, Çorum.
- Davaslıgil Atmaca, V. (2018). BİST şehir endeksleri oynaklığının DCCGARCH model ile analizi. *Yönetim Bilimleri Dergisi*, 16(31), 287-308. Erişim Adresi: <https://dergipark.org.tr/en/pub/comuybd/issue/44733/556047>
- Fama, E., F. (1970). Efficient capital markets: a review of empirical work. *Journal of Finance*, 25(2),383-417. doi: 10.2307/2325486
- Filiz, E., Karaboğa, H.A., ve Akoğul, S. (2017). BİST-50 endeksi değişim değerlerinin sınıflandırılmasında makine öğrenmesi yöntemleri ve yapay sinir ağları kullanımı. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26(1), 231-241. Erişim Adresi: <https://dergipark.org.tr/tr/pub/cusosbil/issue/31939/350973>
- Freund, Y., ve Schapire, R., E. (1996). Experiments with a new boosting algorithm. *Proceedings of International Conference on Machine Learning*, (96), 148-156. Erişim Adresi: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.6252&rep=rep1&type=pdf>
- Güneş, İ. (2019). Şehirlere küresel şehir endekslerinden bakış. *Strategic Public Management Journal*, 5(10), 111-130. doi:10.25069/spmj.585811
- Gürsoy, S., Alptürk, Y., ve Tunçel, M., B. (2020). Dolar ve euro kurundaki değişimlerin Antalya şehir endeksi üzerindeki etkisinin incelenmesi: çok değişkenli VAR-EGARCH uygulaması. *Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 10(2), 43-56. doi:10.47147/ksuüibf.828592
- Hasan, A., Kalıpsız, O., ve Akyokuş, S. (2020). Modeling traders' behavior with deep learning and machine learning methods: evidence from BIST 100 index. *Complexity*, 2020, 1-16. doi: 10.1155/2020/8285149
- Kantar, L. (2020). BİST 100 endeksinin yapay sinir ağları ve ARMA modeli ile tahmini. *Muhasebe ve Finans İncelemeleri Dergisi*, 3(2) , 121-131. doi:10.32951/mufider.725722
- Kara, Y., Boyacioglu, M.A., ve Baykan Ö., K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the İstanbul stock exchange. *Expert Systems with Applications*, 38(5), 5311-5319. doi: 10.1016/j.eswa.2010.10.027
- Karaatlı, M., Güngör, İ., Demir, Y., ve Kalaycı, Ş. (2005). Hisse senedi fiyat hareketlerinin yapay sinir ağları yöntemi ile tahmin edilmesi. *Yönetim ve Ekonomi Araştırmaları Dergisi*, 3(3), 38-48. Erişim Adresi: <https://dergipark.org.tr/tr/pub/yead/issue/21828/234628>
-

-
- Kayral, İ.E. (2020). BİST şehir endeksleri ile döviz kurları arasındaki ilişkinin incelenmesi: bir ARDL sınır testi uygulaması. *IBAD Sosyal Bilimler Dergisi*, (6), 272-284. doi: 10.21733/ibad.668915
- Kayral, İ.E., ve Tandoğan, N.Ş. (2019). BİST şehir endekslerinde ay içi ve ay dönümü anomalilerinin incelenmesi. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, 8(4), 3114-3133. doi: 10.15869/itobiad.633844
- Keles, M., Keles, A., ve Keles, A. (2020). Makine öğrenmesi yöntemleri ile uçuş fiyatlarının tahmini. *Euroasia Journal of Mathematics, Engineering, Natural and Medical Sciences*, 7(11), 72-78.
- Khaidem L., Saha, S., ve Dey, S., R. (2016). Predicting the direction of stock market prices using random forest. Retrieved from: <https://arxiv.org/pdf/1605.00003.pdf>
- Kula, V., ve Baykut, E. (2018). BİST şehir endekslerinin volatilité yapıları ve rejim değişimlerinin analizi. *Muhasebe ve Finans İncelemeleri Dergisi*, 1(1), 38-59. doi:10.32951/mufider.382687
- Kumar, M., ve Thenmozhi, M. (2006). Forecasting stock index movement: a comparison of support vector machines and random forest. *Indian Institute of Capital Markets 9th Capital Markets Conference Paper*, Retrieved from: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=876544
- Meijering, J.V., Kern, K., ve Tobi, H. (2014). Identifying the methodological characteristics of European green city rankings. *Ecological Indicators*, 43, 132-142. doi: 10.1016/j.ecolind.2014.02.026
- Onan, A. (2017). Twitter mesajları üzerinde makine öğrenmesi yöntemlerine dayalı duygu analizi. *Yönetim Bilişim Sistemleri Dergisi*, 3(2), 1-14. Erişim Adresi: https://atif.sobiad.com/index.jsp?modul=makaledetay&Alan=sosyal&Id=AWHRGnkWoDuH9Br_eOCd
- Özdemir A.K., Tolun, S., ve Demirci, E. (2011). Endeks getirisi yönünün ikili sınıflandırma yöntemiyle tahmin edilmesi: İMKB-100 endeksi örneği. *Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 4(2), 45-59. Erişim Adresi: <https://dergipark.org.tr/tr/pub/niguiibfd/issue/19749/211376>
- Özkan, N., ve Ulaş, Ü. (2021). Bölgesel Covid-19 vaka sayıları, altın fiyatları, Euro ve BIST şehir endeksleri arasındaki ilişki: bir ARDL sınır testi yaklaşımı. *Ekonomi Politika ve Finans Araştırmaları Dergisi*, 6(1), 240-253. doi: 10.30784/epfad.880244
- Pabuçcu, H. (2019). Borsa endeksi hareketlerinin tahmini: trend belirleyici veri. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 22(1), 246-256. doi: 10.29249/selcuksbmyd.487862
- Patel, J., Shah, S., Thakkar, P., ve Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, 42(1), 259-268. doi: 10.1016/j.eswa.2014.07.040
- Sevli, O. (2019). Göğüs kanseri teşhisinde farklı makine öğrenmesi tekniklerinin performans karşılaştırması. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 176-185. doi:10.31590/ejosat.553549
- Shynkevich, Y., McGinnity, T.M., Coleman, S.A., Belatreche, A., ve Li, Y. (2017). Forecasting price movements using technical indicators: investigating the impact of varying input window length. *Neurocomputing*, 264(15), 71-88. doi.org/10.1016/j.neucom.2016.11.095
-

- Tekin, S., ve Çanakoğlu, E. (2018). Prediction of stock returns in İstanbul Stock Exchange using machine learning methods. *Signal Processing and Communications Applications Conference (SIU)*, 1-4. doi: 10.1109/SIU.2018.8404607.
- Telli, Ş., ve Coşkun, M. (2016). Forecasting the BİST 100 index using artificial neural networks with consideration of the economic calendar. *International Review of Economics and Management*, 4(3), 26-46. doi: 10.18825/irem.67309
- Yakut, E., Elmas, B., ve Yavuz, S. (2014). Yapay sinir ağları ve destek vektör makineleri yöntemleriyle borsa endeksi tahmini. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(1), 139-157. Erişim Adresi: <https://dergipark.org.tr/tr/pub/sduiibfd/issue/20816/222712>
- Yapraklı, S., Bozma, G., ve Akdağ, M. (2018). BİST şehir endekslerinde oynaklığın ölçülmesi: alternatif ekonometrik modellerin karşılaştırmalı olarak incelenmesi. *Finans Politik ve Ekonomik Yorumlar*, (639), 67-86. Erişim Adresi: <https://dergipark.org.tr/tr/pub/fpeyd/issue/47984/607050>

SUMMARY

Introduction

The regional index is a portfolio that reflects the specific characteristics of a geographical region of interest and has been calculated for many years for different purposes. In 2009, Borsa Istanbul (BIST) announced that it started to calculate city indexes in order to reflect the financial performances of 9 cities. Nowadays, 13 city indices namely Adana (XSADA), Ankara (XSANK), Antalya (XSANT), Aydın (XSAYD), Balıkesir (XSBAL), Bursa (XSBUR), Denizli (XSDEN), Istanbul (XSIST), Izmir (XSIZM), Kayseri (XSKAY), Kocaeli (XSKOC), Konya (XSKON) and Tekirdağ (XSTKR) are calculated. Istanbul is one of the cities with the most active commercial and financial relations in Turkey. XSIST, which was started to be calculated in 2009, is a portfolio that includes 110 companies today. Financial institutions constitute the majority of these companies with 28.1%, followed by the manufacturing sector with 27.2%.

Background

When the limited numbers of studies on city indices are examined, it is seen that researchers mostly focus on the volatility. Volatility is generally examined with autoregressive conditional heteroscedasticity type models (ARCH-type) within the scope of time series analysis. On the other hand, it is seen that most of the mentioned studies benefit from macroeconomic indicators in the modelling process. In this study, a classification-based modelling process has been proposed for the prediction of the future behaviour of the Istanbul city index, which is one of the important indicators of regional development and calculated by BIST. The proposed process provides a comprehensive perspective on feature engineering methods.

Method

The first step of the modelling process is to extract features based on technical indicators by utilizing the meta-information of the XSIST time series. When the studies for financial forecasting are examined, it is seen that approximately 20% of them use indicators obtained from Technical Analysis. Each indicator calculated with Technical Analysis uses additional information from an asset's price. These indicators are calculated for each trading day using the asset's opening value, closing value, high/low value and trading volume information. In this study, 38 different indicators based on volume, volatility, trend and momentum are extracted.

The second step of the study is to select the features that will contribute to the model and will be used to construct the simplest model with the least information loss. In this study, recursive feature elimination-RFE method is used for the feature selection step.

The third step of the study is to construct learning models with obtained features in the previous step. Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), k-nearest neighbourhood (k-NN), Adaboost and Gradient Boosting (GB) classifiers are used and for each learning model, a composite procedure based on 10-fold cross-validation and grid search is applied to identify the hyper parameters that provide the best classification. With the optimal hyper parameters determined by working on the training set, the models are re-run on the test set and the results are obtained.

The fourth and the final step of this study is to compare the performance of learning models based some statistical metrics such as accuracy, sensitivity, specificity, F1-Score and area under ROC curve.

Results and Conclusions

Among 38 technical indicators, it is found that the optimal value of the number of features to be taken as input to the classification models is 10. The most important indicator is the stochastic RSI oscillator (Stochastic Relative Strength Index), also known as the speed indicator. Volume Price Trend is the second most important indicator and is among the volume-based indicators. It is seen that there are mostly momentum-based indicators in the feature vector created with the technical indicators.

For all performance metrics considered in the study, k-NN showed the lowest performance. This learning model, which suffers from high variance and bias, may improve the performance of the model if it is studied with more training data or with a different k value. Although RF and GB learning models have high classification values in terms of performance metrics, they need more training data to talk about the validity of the models. When the learning curves of the LR, SVM and Adaboost models with the highest accuracy values are examined, it is seen that the Adaboost model, which classifies with an accuracy of 88.77%, has an overfitting problem. When all performance metrics of LR and SVM models are compared, it is seen that almost similar results are found. The error of the LR learning model has the lowest with 0.204%, but this rate is found to be 2.719% in the SVM.