





## A feature-based hybrid ARIMA-ANN model for univariate time series forecasting

Ümit Çavuş Büyüksahin\* , Şeyda Ertekin 

Department of Computer Engineering, Middle East Technique University, Ankara, 06800, Turkey

### Highlights:

- A feature-based hybrid ARIMA-ANN method is proposed for time series forecasting.
- Our new hybrid method calculates features of a given time series and selects most important ones.
- The method achieved better forecasting accuracy than the examined individual and hybrid methods.

### Keywords:

- Time series forecasting
- Artificial neural network
- Autoregressive integrated moving average
- Gradient boosting trees

### Graphical/Tabular Abstract

The architecture of the proposed feature-based hybrid method is indicated in Figure 1.

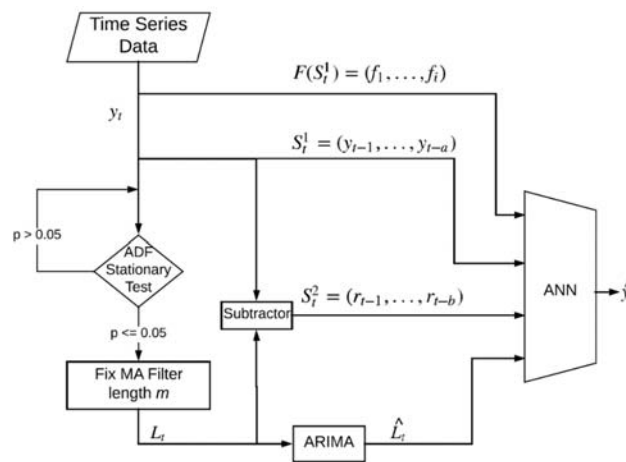


Figure 1 Proposed feature-based hybrid method

### Article Info:

Research Article  
Received: 04.01.2019  
Accepted: 17.08.2019

### DOI:

10.17341/gazimmfd.508394

### Correspondence:

Author: Ümit Çavuş Büyüksahin  
e-mail:  
umit.buyuksahin@ceng.metu.edu.tr  
phone: +90 506 705 6885

**Purpose:** In this study, it is aimed to improve the high-accuracy hybrid ARIMA-ANN model for univariate time series forecasting by converting it to a feature-based model.

### Theory and Methods:

The proposed method firstly obtains a linear component from a given univariate time series by using moving average (MA) filter. Then, nonlinear residual series, which is the difference between given time series and linear component, is achieved. After that, a set of classical and advanced features is extracted from the observed data and their importance scores are computed by using gradient boosting trees. The most informative features are selected and used in the hybrid method as explanatory variables which lead to improve time series forecasting accuracy results.

### Results:

The performance of the proposed feature-based hybrid method is evaluated on four different time series each which come from different fields and have different statistical characteristics. The proposed method improves the absolute accuracy of the hybrid method that we built on it, between 5.5% and 11.8% in the evaluated datasets.

### Conclusion:

The univariate time series forecasting methods can use only values in the prior time steps to predict next value. They do not take account of underlying structural characteristics of past values which might provide a more meaningful summarized information that lead to yield better approximations in the forecasting. The developed method, extracts features which basically measure the characteristics of the given series. After that, the most explanatory features are selected and used in the hybrid method for improving forecasting performance results. Our experimental results indicate that our feature-based hybrid method gives superior accuracy as compared to well-known existing methods.



## Tek değişkenli zaman serileri tahmini için öznitelik tabanlı hibrit ARIMA-YSA modeli

Ümit Çavuş Büyüksahin\*<sup>ID</sup>, Şeyda Ertekin<sup>ID</sup>

Orta Doğu Teknik Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 06800, Ankara, Türkiye

### Ö N E Ç I K A N L A R

- Zaman serileri tahmini için öznitelik tabanlı bir hibrit ARIMA-YSA yöntemi önerilmiştir
- Yeni hibrit yöntemimiz verilen zaman serisinin özelliklerini hesaplar ve en önemlilerini seçer
- Yöntem, incelenen tekil metotlardan ve hibrit modelinden daha iyi tahmin doğruluğu sağlamaktadır

### Makale Bilgileri

Araştırma Makalesi

Geliş: 04.01.2019

Kabul: 17.08.2019

### DOI:

10.17341/gazimmfd.508394

### Anahtar Kelimeler:

Zaman serisi tahmini,  
yapay sinir ağı,  
otoregresif bütünlük  
hareketli ortalama,  
gradyan artırma ağaçları

### ÖZET

Zaman serilerinde yüksek performanslı tahminleme yapabilmek birçok uygulama alanı için temel öneme sahiptir. Literatürde zaman serisi tahmin doğruluğunu artırmak için birçok metod önerilmiştir. Bu metotlardan tek değişkenli zaman serilerine odaklanmış olanlar, serinin sadece geçmiş tarihinde yer alan değerleri kullanarak, gelecekteki değerlerin tahminini yapmaktadır. Bu çalışmada, tek değişkenli zaman serilerinin geçmiş değerlerinin yanında, serinin karakteristiğini özetleyen yapısal özniteliklerinin de kullanılarak, tahmin performansının artırılması hedeflenmiştir. Zaman serilerinin istatistiksel özetini sunan özniteliklerin önem puanları gradyan artırma ağaçları (GBT) ile belirlenmektedir. En yüksek önem puanına sahip olan öznitelikler, hibrit ARIMA-YSA modeline açıklayıcı ilave değişkenler olarak verilmektedir. Geliştirilen yöntemin değerlendirilmesi dört farklı veri seti üzerinde yapılmış olup, mevcut kabul görmüş yöntemlere kıyasla daha başarılı sonuçlar elde edildiği görülmüştür.

## A feature-based hybrid ARIMA-ANN model for univariate time series forecasting

### H I G H L I G H T S

- A feature-based hybrid ARIMA-ANN method is proposed for time series forecasting
- Our new hybrid method calculates features of a given time series and selects most important ones
- The method achieved better forecasting accuracy than the examined individual and hybrid methods

### Article Info

Research Article

Received: 04.01.2019

Accepted: 17.08.2019

### DOI:

10.17341/gazimmfd.508394

### Keywords:

Time series forecasting,  
artificial neural network,  
autoregressive integrated  
moving average,  
gradient boosting trees

### ABSTRACT

High prediction accuracies at time series modeling and forecasting is of the utmost importance for a variety of application domains. Many methods have been proposed in the literature to improve time series forecasting accuracy. Those which focus on univariate time series forecasting methods use only the values in the prior time steps to predict the next value. In this study in addition to the historical values, it is aimed to increase the forecasting performance by using extra statistical and structural features which summarize characteristics of the time series. Feature importance scores are determined by gradient boosting trees (GBT). Features with the highest importance score are given as explanatory additional variable to the hybrid ARIMA-ANN model. The evaluation of the developed method is performed on four different publicly available datasets. Our experimental results show higher accuracy performance for the proposed method as compared to the currently well-accepted methods.

\*Sorumlu Yazar/Corresponding Author: umit.buyuksahin@ceng.metu.edu.tr, seyda@ceng.metu.edu.tr / Tel: +90 506 705 6885

## 1.GİRİŞ (INTRODUCTION)

Gelişmekte olan teknolojiler sayesinde veri toplamadaki artış ve kolaylık, büyük sayıda zaman serisi verisini de beraberinde getirmektedir. Zaman serisi verilerinden bilgi çıkararak, gelecekteki olası değerleri anlama işlemine zaman serisi tahmini denir. Zaman serisi tahminlemesi, finans alanında döviz kurları, elektrik fiyatları, borsa endeksi değeri; enerji alanında santrallerde enerji üretimi, elektrik tüketimi; meteoroloji alanında hava durumu; medikal alanda tahminlemelerinde kritik önem taşımaktadır [1]. Zaman serileri tahminlemelerinin karar destek sistemlerinde vazgeçilmez önemi nedeniyle, verimli ve doğru çalışan modellerin geliştirilmesi gittikçe önem kazanmaktadır. Zaman serileri, kullanılan değişken sayısına göre iki kategoriye ayrılır: Tek değişkenli ve çok değişkenli zaman serileri. Tek değişkenli zaman serileri bir zaman dizisinin tek bir değişkenin ardışık kayıtlarını içerdiği anlamına gelir. Çok değişkenli seriler ise, zaman içinde ardışık olarak kaydedilen birden fazla değişken verisinden oluşmaktadır. Zaman serilerinin birçoğu sadece zamana dayalı gözlemlerden oluşur; bu gözlemlerle ilişkilendirilecek, çok değişkenli serilerde olduğu gibi başka açıklayıcı değişkenlere sahip değildir. Bu nedenle, zaman serileri tahmin yöntemlerinin çoğu tek değişkenli zaman serilerine odaklanmaktadır. Tek değişkenli serilerde, sadece serinin geçmiş tarihinde yer alan değerler kullanılarak, gelecek değerlerin tahmini yapılır. Tek değişkenli seriler için tasarlanan modeller, serideki mevcut değer, geçmiş gözlemlerin doğrusal veya doğrusal olmayan bir fonksiyonu olarak kabul edilip edilmediğine bağlı olarak ikiye ayrılır. Bu modellere sırasıyla doğrusal ve doğrusal olmayan modeller denir.

Otoregresif bütünleşik hareketli ortalama (ARIMA) algoritması, doğrusal zaman serisi tahmin yöntemleri arasında yaygın olarak kullanılan bir metottur. Literatürde enerji [2], hidroloji [3], mühendislik [4] ve finans [5] alanları dahil olmak üzere, çok sayıda alanda ARIMA uygulanmıştır. ARIMA yöntemi, otoregresif (AR), hareketli ortalama (MA), AR ve MA modellerinin kombinasyonu olan ARMA modelinden yararlanmaktadır. Diğer doğrusal yöntemlere benzer şekilde, ARIMA, gelecekteki veri değerlerinin mevcut ve geçmiş veri değerlerine doğrusal olarak bağlı olduğu konusunda güçlü bir varsayım yapmaktadır [6]. Bu sayede ARIMA, durağan zaman serileri tahmininde yüksek doğruluk sonuçları verir. Fakat gerçek dünyada karşılaşılan zaman serileri, doğrusal bir ARIMA yöntemi ile modellenemeyen, lineer olmayan hareketler içerir. Pratikte karşılaşılan zaman serilerindeki doğrusal olmayan, karmaşık hareketleri yakalamak için literatürde birçok doğrusal olmayan yöntem önerilmiştir. Literatürde en bilinen algoritmalarından biri olan Yapay Sinir Ağları (YSA) biyoinformatik [7], sosyal bilimler [8], enerji [9] ve finans [10] gibi pek çok alanda uygulanmıştır. YSA doğrusal olmayan karmaşık bir fonksiyonu modelleyebilir ve yüksek doğrulukta tahminler üretebilir [11, 12]. Ayrıca, YSA'lar doğası gereği veri güdümlü algoritmalar; dolayısıyla istenen YSA modelleri zaman serisi verilerinin tek

değişkenli veya çok değişkenli özelliklerine göre oluşturulabilirler [13]. YSA'lar bu yetenekleri sayesinde, çok değişkenli zaman serileri modellemesinde yaygın olarak tercih edilmektedir. Fakat YSA modellemesi zorlu ve problemleri bir süreçtir. Çünkü YSA'lar bir eğitim verisine aşırı uyma (overfitting) veya zayıf genelleme (poor generalization) problemleriyle karşılaşabilir. Literatürde lineer modellerin, belirli koşullar altında, YSA'dan daha iyi doğruluk sonuçları verdiğini gösteren çeşitli çalışmalar bulunmaktadır [14, 15]. Son zamanlarda, literatürde doğrusal ve doğrusal olmayan tekil yöntemleri birleştiren çeşitli hibrit yöntemler önerilmiştir. Bu hibrit yöntemler, her bir tekil yöntemin benzersiz mukavemetinin avantajlarını kullanabilmeleri için sunulmuştur. Bu sayede hibrit yöntemlerin, kullandığı tekil yöntemlerden daha iyi performans elde ettikleri gösterilmiştir [13, 16-18]. Hibrit yöntemlerdeki yaygın uygulama şu şekildedir: Önce, belirli bir zaman serisi verisi doğrusal ve doğrusal olmayan bileşenlerine ayrıştırılır; daha sonra uygun tekil yöntemler bu bileşenler üzerinde ayrı ayrı uygulanır. Hibrit yöntemler, zaman serisi verisini bileşenlerine ayrıştırma ve model oluşturma yaklaşımlarına göre birbirlerinden ayrılırlar. Örneğin, [18] tarafından önerilen hibrit yöntem, doğrusal ve doğrusal olmayan bileşenleri düzgün bir şekilde üretebilmekte ve farklı yaklaşımların avantajlarını etkili bir şekilde kullanabilmektedir. Bu sayede sadece tekil yöntemlerden değil, diğer hibrit yaklaşımlardan da daha iyi tahmin performansı elde etmektedir.

Bu tek değişkenli doğrusal, doğrusal olmayan ve hibrit yöntemler, bir sonraki değeri tahmin etmek için sadece önceki zaman adımlarındaki değerleri kullanırlar. Diğer taraftan, geçmiş değerlerin altında yatan yapısal özniteliklerinin yakalanması, tahminlerin performansını artıran daha anlamlı özetlenmiş bilgiler sağlayabilir. Zaman serilerinin öznitelikleri, belirli zaman aralıklarındaki değerlerin istatistiksel hesaplamalarından oluşur. Zaman serilerinde öznitelik kullanarak modelleme yapma fikri üzerine çeşitli çalışmalar bulunmaktadır. Örneğin, zaman serilerinde kümelenmeyi hedefleyen çalışmalar [19-21] ve anormallikleri tespiti için yapılan çalışmalar [22, 23] bulunmaktadır. Zaman serilerinin kümelenmesini hedefleyen bu çalışma [19], zaman serilerinden istatistiksel yöntemlerle elde edilen özniteliklerle zaman serilerini tanımlamayı hedeflemiştir. Benzer bir şekilde [20]'deki çalışma, zaman serilerinde kümeleme işlemi için geçmiş verilere bakmak yerine, zaman serilerinin küresel özniteliklerini çıkarmıştır. Çıkarılan öznitelikler için boyut küçültme yöntemi uygulayarak, zaman serilerinin az sayıda öznitelikle daha etkili ve daha başarılı bir şekilde kümelenebileceğini göstermiştir. [21]'deki çalışma çok büyük miktarda zaman serisi veri tabanından, zaman serilerinin korelasyon yapıları, dağılımları, entropi, durağanlık, ölçekleme gibi birçok öznitelikliğini çıkarmıştır. Daha sonra, bu öznitelikler arasında en fazla bilgi verenleri seçerek, özellik tabanlı sınıflandırıcılar ile zaman serilerinde sınıflar arasındaki farkları otomatik olarak öğrenmiştir. [22]'deki çalışma, Yahoo e-posta sunucularında zaman

içindeki davranışlarında alışılmadık durumları tespit etmek için öznitelik tabanlı yöntem geliştirmiştir. Bu yöntem, her bir sunucudan gelen zaman serisi verisinden, serinin özelliklerini ölçen (korelasyon, mevsimsellik gücü, entropi, vb.) öznitelik vektörünü hesaplamıştır. Daha sonra bu vektörden, aykırı saptama yöntemi ile olağandışı hareketleri tespit etmiştir. [23]'deki çalışmada Uber'deki zaman serisi veri setlerinde bayram ya da başka özel günlerde ortaya çıkan alışılmadık durumları tespit etmek için öznitelik tabanlı bir yöntem önermiştir. Bunun için zaman serisindeki her bir noktayı bir öznitelik vektörü ile tanımlamıştır. Daha sonra öznitelik tabanlı veri setleri Uzun/Kısa Süreli Bellek (LSTM Long-Short Term Memory) sinir ağı metodu ile modellemiştir. Bu sayede, alışılmadık durumların tespitinde %18'a kadar performans artışı elde edebilmişlerdir. Bu çalışmada, yüksek doğruluk veren hibrit ARIMA-YSA modelini [18], öznitelik tabanlı bir modele çevirerek, geliştirmeyi hedefledik. İstatistiksel yöntemlerle hesaplanan öznitelikler, verilen zaman serisini açıklayıcı değişkenler olarak hibrit yöntemde sunulmuştur. En önemli öznitelikleri içeren küme, verilen zaman serilerine göre değişebileceğinden, önerdiğimiz yöntem bir öznitelik seçimi ön-işlem adımını uygulamaktadır. Bu adımda, öznitelik önem puanlarını akıllıca elde etmek için özniteliklerden gradyan artırma ağaçları (GBT) oluşturulmaktadır. GBT'ler aracılığı ile özniteliklerin önem puanları hesaplanır ve en bilgilendirici olanları seçilmektedir. Daha sonra, bu seçilmiş öznitelikler hibrit yöntemde zaman serileri tahmin doğruluğu sonuçlarının geliştirilmesini sağlayan, açıklayıcı değişkenler olarak kullanılmaktadır.

## 2. ÖNERİLEN METOT (PROPOSED METHOD)

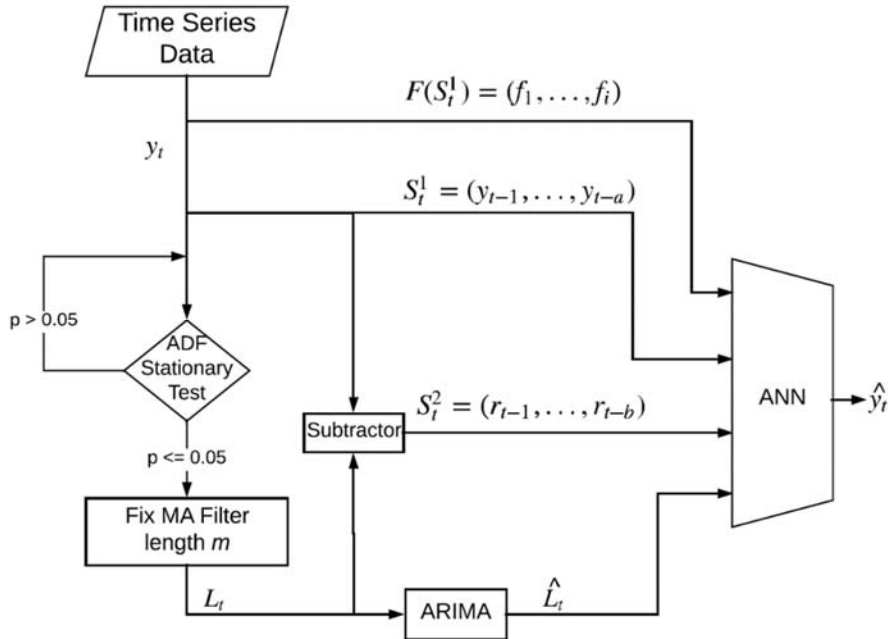
Tahmin doğruluğunu artırmak için literatürde birçok tekil veya hibrit yaklaşımda kullanılan zaman serisi tahmin

metodu geliştirilmiştir. Fakat bu metotlar, iki önemli soruna çözüm üretmekte zorluk çekmiştir. İlk sorun, zaman serileri doğrusal veya doğrusal olmayan veya her iki temel sürecin birleşiminden oluşmuş olabilirler [24]. Bu nedenle, her durum için doğru tekil tahminleme yönteminin seçilmesi zordur. Metodumuzun hibrit modelleme kısmı, her bir tekil yöntemin avantajını birleştirerek bu problemi hafifletmekte, uygun olmayan bir metot kullanma riskini azaltmakta ve daha doğru sonuçlar elde etmektedir. İkinci problem, zaman serilerinin açıklayıcı değişkenler tarafından tatmin edici bir şekilde ilişkilendirilememesidir. Bu nedenle, literatürdeki birçok yöntem, zaman serisinin ayırt edici özelliklerini belirlemek için genellikle yetersizdir [13]. Önerdiğimiz yöntem, öznitelikleri çıkararak zaman serilerinin karakterini daha iyi yakalamayı amaçlamaktadır. Bu sorunları çözmeyi hedefleyen metodumuz, zaman serilerinde daha tutarlı ve yüksek doğrulukta tahminler yapabilmektedir. Önerilen öznitelik tabanlı hibrit yöntemin mimarisi, Şekil 1'de gösterilmiştir.

### 2.1. Hibrit Metot (Hybrid Method)

Öznitelik tabanlı hibrit yöntemimiz, tek değişkenli zaman serileri için tasarlanmış ve yüksek doğruluk veren hibrit bir metodun [18] üzerine inşa edilmiştir. Önerdiğimiz metot, tek değişkenli zaman serisini çıkarılan özniteliklerle çok değişkenli formata çevirmektedir. Tek değişkenli zaman verisi  $y_t$ , sıralı  $n$  tane değerden oluşmaktadır:  $y_t = y_1, y_2, \dots, y_n$ . Bu yöntemde zaman serileri, Eş. 1'de verilen doğrusal bileşen  $L_t$ , doğrusal olmayan bileşen  $N_t$  ve seçilen öznitelik kümesi  $F_t$  'nin bir fonksiyonu olarak kabul edilmektedir.

$$y_t = f(L_t, N_t, F_t) \quad (1)$$



Şekil 1. Önerilen öznitelik tabanlı hibrit metot (Proposed feature-based hybrid method)

Kullandığımız hibrit metot Eş. 2 'de gösterilen hareketli ortalama (MA) filtresi kullanarak, verilen zaman serisinden, doğrusal bileşen  $l_t$  çıkarılmaktadır.

$$l_t = \frac{1}{m} \sum_{i=t-m+1}^t y_i \quad (2)$$

Eş. 2'de gösterilen  $m$ , verilen seriden doğrusal bileşen elde etmek için uygun şekilde ayarlanması gereken filtre uzunluğunu ifade etmektedir. Filtre uzunluğu, filtreden doğrusal bir bileşen elde edilene kadar ayarlanmaktadır. Bu aşamada, filtre çıktısının doğrusallığı, bir birim kök testi Augmented Dickey Fuller (ADF) testi ile ölçülmektedir. Birim kök, belirli bir veri seti üzerinde öngörülemez bir hareketin varlığını gösterir. Eğer verilen bir veri setinde birim kök varsa, ADF test kabul edilen eşik değer 0.05'ten daha büyük sonuç vermektedir. Filtre çıkışı ADF testini geçene ve doğrusal sonuç elde edilene kadar filtrenin uzunluğu ayarlanır.

Elde edilen doğrusal bileşen  $l_t$ , geçmiş veriler  $l_{t-1}, l_{t-2}, \dots, l_{t-p}$  ve rastgele hata serileri  $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$  ile doğrusal ARIMA metodu  $g$  ile Eş. 3'te görüldüğü gibi modellenmektedir.

$$\hat{L}_t = g(l_{t-1}, l_{t-2}, \dots, l_{t-p}, \epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}) \quad (3)$$

Kalan seri, Eş. 4'te gösterildiği gibi orijinal zaman serisi  $y_t$  ile elde edilen doğrusal seri  $l_t$  arasındaki, doğrusal olmayan farktır.

$$r_t = y_t - l_t \quad (4)$$

Doğrusal ve doğrusal olmayan bileşenler elde edildikten sonra, ortalama (mean), varyans (variance), çarpıklık (skewness), basıklık (kurtosis), vb. gibi verilen bir zaman serisinin karakterini açıklayan klasik ve ileri istatistiksel özellikler hesaplanmaktadır. Bu özellikler, geçmiş veriler  $y_{t-1}, y_{t-2}, \dots, y_{t-a}$  üzerinden hesaplanır. Bu özelliklerin gradyan artırma yöntemi (GBT) ile önem puanları hesaplanmaktadır. Bu önem puanları, her bir özneliğin verilen zaman serisini ne kadar açıkladığını göstermektedir.

En yüksek önem puanına sahip olanlar, yani en açıklayıcı  $k$  tane öznelikler  $F_t = f_1, f_2, \dots, f_k$  hibrit metotta kullanılmak üzere seçilmektedirler. Öznelik çıkarma ve seçim adımının detayları bir sonraki bölümde verilmektedir.

Önerilen yöntemin son aşamasında ( Şekil 1), doğrusal olmayan modelleme YSA, Eş. 5'te belirtildiği gibi, elde edilen bileşenler ile seçilen öznelik kümesi arasındaki fonksiyonel ilişkiyi tanımlamak için kullanılmaktadır. YSA modeli burada, geçmiş verileri  $y_{t-1}, y_{t-2}, \dots, y_{t-a}$ , mevcut ARIMA tahmin sonucunu  $\hat{L}_t$ , doğrusal olmayan kalan verileri  $r_{t-1}, r_{t-2}, \dots, r_{t-b}$ , ve seçilen öznelikleri  $f_1, f_2, \dots, f_k$  girdi olarak almaktadır.

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-a}, \hat{L}_t, r_{t-1}, r_{t-2}, \dots, r_{t-b}, f_1, f_2, \dots, f_k) \quad (5)$$

Eş. 5'te,  $f$  doğrusal olmayan YSA fonksiyonunu;  $a$  ve  $b$  parametreleri ise YSA'da kullanmak için zamanda ne kadar geriye gideceğimizi göstermektedir. Bu parametreler deneysel olarak belirlenmektedir. Deneylerden elde edilen gözlemlerimize göre, durağan olmayan veri setlerinde, durağan olanlara göre daha eski verileri görmek, tahmin sonuçlarının performansını artırmaktadır.

## 2.2. Öznelik Çıkarma (Feature Extraction)

Tek değişkenli zaman serileri tahminlemesi, yalnızca serinin kendisinin geçmiş tarihinde yer alan gözlemleri kullanılmaktadır. Fakat bu yöntemle verilen zaman serisinin karakteri yeterince iyi kavranamaz. Bu nedenle, yöntemimiz zaman serisi veri setini türetilmiş özellikler kümesi olarak temsil ederek, serideki dinamik değişkenliği, daha statik hale dönüştürmektedir. Diğer bir deyişle, tek değişkenli zaman serileri, öznelik bazlı temsil oluşturularak çok değişkenli forma çevrilmektedir.

Öznelikler, gözlemlenen geçmiş veriler üzerinde, belirli bir uzunlukta kaydırılan bir çerçeveden hesaplanmaktadır. Bu öznelikler, belirli bir zaman serisinin özelliklerini tanımlayan klasik ve ileri istatistiksel hesaplamalardan oluşmaktadır. Hesaplanan öznelikler ve açıklamaları Tablo 1'de listelenmiştir.

**Tablo 1.** Zaman serilerinden çıkarılan özelliklerin özeti (Summary of extracted features from time series)

Öznelik	Tanım
Ortalama (Mean)	Ortalama
Varyans (Variance)	Varyans
MDeğişim (Mchange)	Ortalama değişimi
Yayılm İndeksi, (Index of Dispersion /IoD)	Göreceli varyans
Çarpıklık (Skewness)	Dağılım asimetrisi
Basıklık (Kurtosis)	Dağılımın dağınıklığı
Öziliniti (Autocorrelation)	Seri korelasyon
Entropi (Entropy)	Spektral entropi
Eğilim (Trend)	Eğilimin gücü
Doğrusallık (Linearity)	Doğrusallık gücü
Kesişme Noktası (Crossing points/Cpoints)	Geçiş noktası sayısı

Bunlara ek olarak, öznitelik kümesi her bir özneliğin değişim oranını hesaplayarak zenginleştirilmiştir. Belirli bir zaman  $t$ 'de, belirli bir öznitelik  $f_t$  'in, değişim oranı hesaplaması Eş. 6'daki gibidir.

$$ch\_rate(f_t) = \frac{f_t - f_{t-1}}{f_{t-1}} \quad (6)$$

Hesaplanmış ve zenginleştirilmiş öznitelik kümesi, veri setinin boyutunu artırır. Yüksek boyutluluk içeriği zenginleştirmesine rağmen, modelin eğitim setindeki verileri ezberleyerek, eğitim verisine aşırı uyum (overfitting) problemine yol açabilir. Bu nedenle, veri setindeki en bilgi verici öznitelikleri bularak ve seçerek boyutu azaltmak, tahmin etme işlemi için önemlidir. Öznitelik seçimi (feature selection) sadece veri setinin boyutunu azaltmakla kalmaz, aynı zamanda tahmin performansına katkıda bulunmayan veya doğruluk sonuçlarının düşmesine sebep olan, gereksiz öznitelikleri de ortadan kaldırır. Ayrıca, daha az özniteliklere sahip olmak, veri setinin karmaşıklığını azaltarak daha hızlı ve daha uygun maliyetli bir model oluşmasını sağlamaktadır. Literatürdeki birçok veri madenciliği ve makine öğrenmesi çalışmasında, öznitelik kullanımının ve seçiminin önemi vurgulanmıştır [21-23, 25, 26].

Literatürdeki öznitelik seçim metodları üç grupta toplanmaktadır. 1) Filtre bazlı (filter-based) metodlar, istatistiksel testler yardımıyla çıktı değer ile en güçlü korelasyona sahip öznitelikleri ortaya çıkarır. Varyans Eşiği (Variance threshold) [27] gibi yöntemler bu kategori altında toplanır. Bu yaklaşımın dezavantajı, öznitelikler ile çıktı değer arasında sadece doğrusal ilişkileri yakalamasıdır. 2) Sarıca bazlı (wrapped-based) metodlar yinelemeli arama yöntemi ile buluşsal bir şekilde en açıklayıcı öznitelikleri bulmayı hedefler. Bu kategorideki en bilinen yöntem, Genetik algoritmadır (Genetic algorithm) [28]. Bu yaklaşım, yinelemeli arama sırasında makine öğrenmesi algoritması kullanmasından dolayı zaman alıcıdır ve aşırı uyma problemine yatkındır. 3) Gömülü metodlar (embedded methods) öznitelik seçimini belirli bir öğrenme modelinin oluşturma sürecine dahil ederler. Bu sayede, seçilen öznitelikler, öğrenme modelinin optimize olmasıyla elde edilir. Bu grupta gradyan artırma ağacı (GBT) [29] gibi ağaç tabanlı yöntemler bulunmaktadır. Çıktı değer ile öznitelikler arasında doğrusal olmayan ilişkileri de yakalaması ve aşırı uyma riskinin az olması, gömülü metodları öznitelik seçimi için diğer metodların önüne çıkarmaktadır.

Öznitelik seçimi için gradyan artırma ağacı (GBT) yöntemi [29] kullanılmıştır. Bu yöntem temel olarak zayıf öğrenen karar ağacı modelleri oluşturarak, bunlardan birleşik bir model yaratır. Gradyan artırım ile karar ağaçları oluşturulurken, her bir özneliğin önem puanı hesaplanır. En açıklayıcı ve bilgilendirici öznitelikler daha fazla önem puanına sahiptir. Önem puanını hesaplamak için kullanılan üç ana metod vardır [30]. Birincisi olan *kazanım (gain)* her ağaç için verilen özneliğin modele olan göreceli katkısını hesaplar. İkinci metod *frekans (frequency)*, tüm ağaçlardaki göreceli özellik sayısını hesaplar ve yüzdesel olarak temsil

edilir. Üçüncü metod *kapak (cover)*, verilen özneliğe göre gözlemlerin nispi sayısıdır. Bu çalışmada, her bir öznitelik için önem puan değerlerini tayin etmek için *kazanım (gain)* metodu kullanılmıştır.

Breiman vd. [31], Eş. 7'de gösterildiği gibi tek bir karar ağacı  $T$  için verilen bir  $X_i$  özneliğinin önem puanını hesaplamıştır.

$$w_t^2(T) = \sum_{t=1}^{j-1} \widehat{\tau}_t^2 \quad (7)$$

Karar ağacı,  $X_i$  özneliği ile bir düğümü alt düğümlere böler, böylece  $j-1$  tane iç düğümler elde eder. Ayrık noktadaki  $\tau_t^2$  ile gösterilen öznitelik maksimum tahmini iyileşim değerini sağlar. Sonuç olarak, belirli bir  $T$  karar ağacı için  $X_i$  özneliğinin karesi alınmış önem puanı ( $w_t^2(T)$ ),  $j-1$  tane düğüm üzerindeki karesi alınmış tahmini iyileşim hesaplamalarının toplamıdır.

### 3. DENEYLER (EXPERIMENTS)

Önerilen öznitelik tabanlı hibrit yönteminin performansı, her biri farklı alanlardan gelen ve farklı istatistiksel özelliklere sahip olan dört farklı zaman serisinde değerlendirilmiştir. Bunlardan üçü, zaman serileri tahmini modelleme literatüründe yaygın olarak kullanılan referans veri setleridir: Wolf'un güneş lekeleri verileri, Kanada'da yakalanan vaşak verileri ve İngiliz sterlini/ABD doları döviz kuru verileridir [13, 16-18]. Dördüncü veri seti ise, halka açık olan Türkiye gün içi piyasasının elektrik fiyatıdır [32]. Veri setlerinin ayrıntıları ve deney düzenekleri Tablo 2'de gösterilmiştir.

Deneylerde, önerilen yöntemin performans sonuçları, tüm veri setleri için, üzerine inşa edilen hibrit yöntem [18] ve ARIMA, YSA gibi tekil yöntemlerin performans sonuçlarıyla karşılaştırılmıştır. Deneylerin doğruluk sonuçlarını karşılaştırmak için iki hata metriği kullanılmıştır: Ortalama Mutlak Hata (MAE) ve Ortalama Kareler Hatası (MSE). Eş. 8'de formüle edilen bu metrikler gerçekleştirilen tahminin üzerindeki mutlak hataların ortalamasını ve bu hatalarının karelerinin ortalamasını ölçer.

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \quad MSE = \frac{1}{n} \sum_{t=1}^n e_t^2 \quad (8)$$

Eş. 8'de, tahminlerde yapılan hata  $e_t$  ile gösterilmiştir. Bu hata gerçek zaman serisi  $y_t$  ile tahmin edilen serinin  $\widehat{y}_t$  farkı olarak hesaplanmıştır. Tablo 3, bu dört veri setinin üzerinde incelenen tüm yöntemlerin, bu iki hata metriği ile olan sonuçlarını göstermektedir. Tablonun son sütununda öne sürülen öznitelik tabanlı yöntemin, referans aldığı hibrit yöntemi ne kadar iyileştirdiği verilmektedir.

Şekil 2'de gösterilen *Sunspot* veri seti, 0,08 ADF test sonucuyla (eşik 0,05'ten daha büyük), durağan olmayan bir zaman serisi olarak kabul edilir. Bu veri setinde doğrusal bileşenin çıkarılması için, önerilen yöntemin hibrit kısmında kullanılan MA filtresinin uzunluğu 15 bulunmuştur. Filtre

**Tablo 2.** Veri setleri açıklaması ve deney düzeni (Dataset description and experimental setup)

Veri Seti	İsim	Tanım	Eğitim seti boyutu	Test seti boyutu
Wolf'un güneş lekesi	Sunspot	Güneşin yüzeyinde 1700-1987 yılları arasında görülen güneş lekelerinin yıllık değişimi	221	67
Kanada'da yakalanan vaşak	Lynx	Kuzey Kanada'nın Mackenzie Nehri bölgesinde 1821-1934 yılları arasında yakalanan yıllık vaşak sayısı	100	14
İngiliz sterlini/ABD doları döviz kuru	GbpUsd	İngiliz sterlini ile ABD doları arasındaki döviz kurunun, 1980- 1993 yılları arası haftalık değişimi	679	52
Türkiye Gün İçi Piyasasının elektrik fiyatı	Intraday	Türkiye günüçi elektrik piyasası fiyatının Temmuz 2015- Aralık 2017 arasında günlük ortalama değerleri	521	60

**Tablo 3.** Tüm veri setleri için performans karşılaştırması (Performance comparison for all datasets)

Veri Setleri	Metotlar / Metrikler	YSA	ARIMA	Hibrit Metot	Önerilen Metot	İyileşme (%)
Sunspot	MAE	14,23	13,37	10,48	9,82	5,7
	MSE	353,12	306,97	194,29	16,40	16,5
Lynx	MAE	0,1249	0,1198	0,1013	0,0891	11,8
	MSE	0,0241	0,0231	0,0162	0,0143	10,6
Gbp/Usd	MAE	428,55	435,72	404,90	382,57	5,5
	MSE	3,4681	3,5272	2,9538	2,8426	3,7
Intraday	MAE	20,10	20,22	18,81	17,20	8,1
	MSE	617,46	652,72	581,38	487,21	16,1

\*Gbp/Usd veri seti için MAE ve MSE sonuçları  $10^{-5}$  ile çarpılmıştır

\*\*En iyi sonuçlar kalın tipte gösterilmiştir.

sonucunda elde edilen doğrusal bileşenin ADF test sonucu, bileşenin durağanlığını gösteren 0,006 değerini göstermektedir. Böylece bileşen doğrusal bir yöntem ARIMA ile kolayca modellenebilir hale gelmiştir.

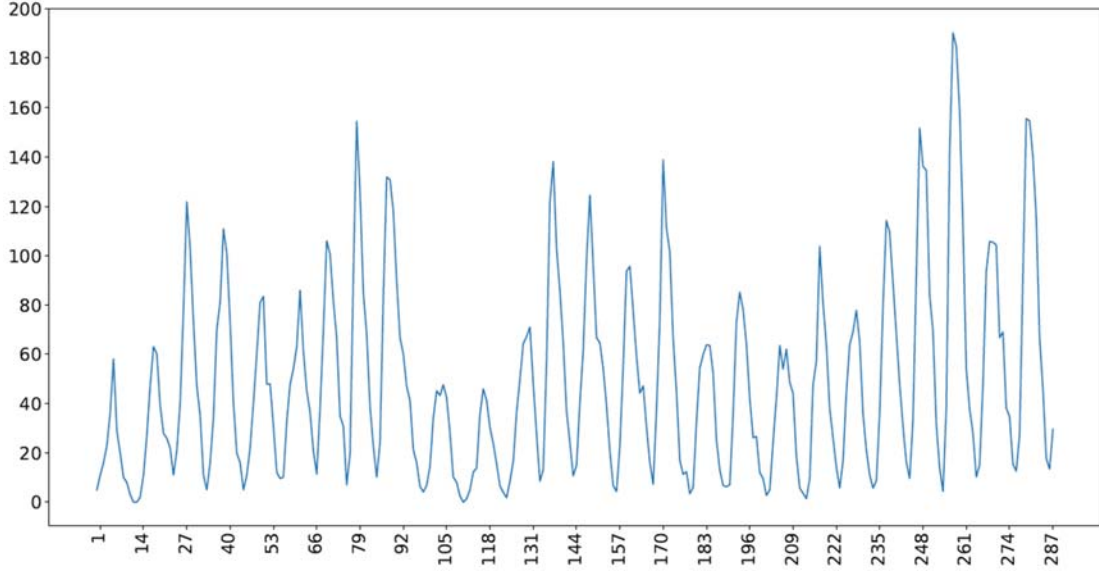
Şekil 3, Sunspot veri seti için hesaplanan özniteliklerin önem puanlarına göre sıralanmış halini göstermektedir. *Mean* özniteliği tahmin performansına en yüksek katkıyı sağlarken, *Cpoints* en düşük katkıyı yapmaktadır. Deneysel gözlemlerimiz göstermiştir ki, nispi önem puanı 0,04'ten az olan öznitelikler zaman serisini yeteri kadar açıklayamamaktadır. Düşük önem puanına sahip öznitelikler tahmin performansına etkili bir şekilde katkıda bulunmadığından, sadece en önemli ilk 12 öznitelik ( Şekil 3) final YSA modellemesine dahil edilmiştir. Final YSA modeli ayrıca, verilen zaman serisinin son 4 değerini, doğrusal olmayan kalan serinin son 2 değerini ve doğrusal bileşen tahminini, girdi katmanından almaktadır ( Şekil 1). Burada kullanılan YSA modeli 39 tane düğüm içeren 1 gizli katmana (hidden layer) sahiptir. Bu modelin eğitiminde 0.01 öğrenme oranı (learning rate), 0.001 sapma değeri (bias), ReLU aktivasyon fonksiyonu ve Nadam optimizasyon algoritması kullanılmıştır.

Sunspot veri setinin sayısal sonuçları analiz edildiğinde ( Tablo 3), hibrit yöntemlerin, ARIMA ve ANN tekil yöntemlerinden belirgin bir şekilde daha iyi performans sağladığı görülmektedir. Bu, ARIMA ve ANN'nin tek tek kullanıldığında, veri serisindeki genel hareketleri

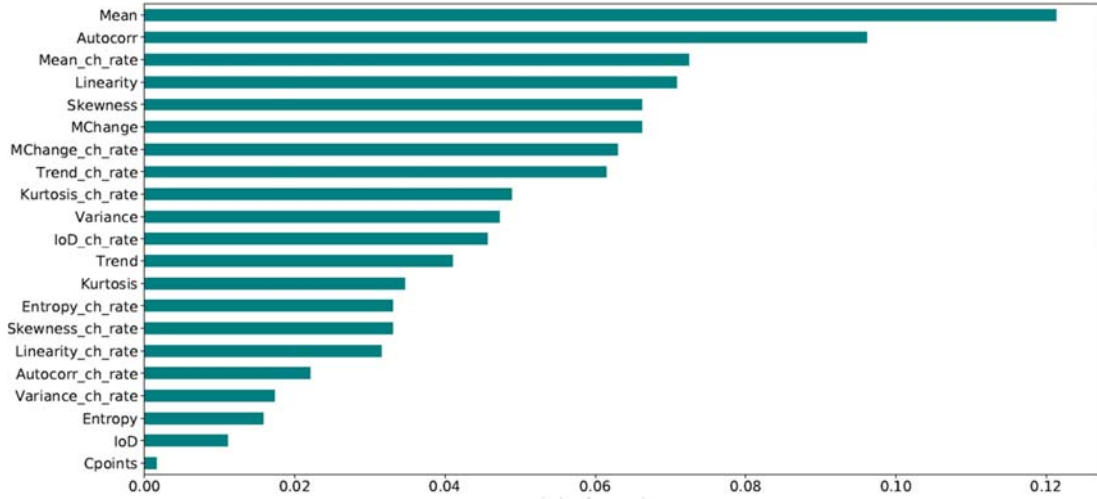
yakalamada yetersiz olmasından kaynaklanmaktadır. Fakat iki yöntemin de etkili yönlerini birleştiren hibrit yöntemler, bu eksikliğin üstesinden gelmek için etkili bir yoldur. Tablo 3'teki hibrit metotların sonuçları da bunu doğrulamaktadır. Diğer taraftan bu çalışmada önerilen yöntem, hibrit yöntemin avantajlarını kullanmakla beraber, veri setinin öznitelik tabanlı temsilini de sağlayarak daha yüksek tahmin performansı vermektedir. Sonuç olarak, önerilen yöntem, üzerine inşa edilen hibrit metodun MAE ve MSE hata ölçümlerini sırasıyla %5,7 ve %16,5 iyileştirmiştir.

Şekil 4'te gösterilen *Lynx* veri seti 0,056 ADF test sonucu ile neredeyse durağan bir zaman serisidir. Bu nedenle, bu veri setinden lineer bileşeni çıkarmak için 5 uzunluğunda kısa bir filtre kullanılması yeterli olmaktadır. Filtreyi uyguladıktan sonra elde edilen doğrusal bileşen 0,006 ADF test sonucu ile ARIMA tarafından düzgün bir şekilde modellenebilmektedir.

Hesaplanan öznitelik önem puanları karşılaştırıldığında ( Şekil 5), *Lynx* veri setinin *Mean*, *Variance*, *Skewness*, *Linearity* gibi istatistiksel özniteliklere duyarlı olduğu görülmektedir. Deneysel gözlemlerimiz sonucu bu veri setinde, en önemli ilk 11 öznitelik ( Şekil 5) final YSA modellemesinde kullanılmak üzere seçilmiştir. Ayrıca, final YSA modelinin girdi katmanında, verilen zaman serisinin son 5 değeri, doğrusal olmayan kalan serinin son 3 değeri ve doğrusal bileşen tahmini olmak üzere 9 tane daha düğüm bulunmaktadır. Burada kullanılan YSA modeli 41 tane



Şekil 2. Sunspot serisi (1700-1987) (Sunspot series (1700-1987))



Şekil 3. Sunspot veri seti için öznelik önemi (Feature importance for Sunspot dataset)

düğüm içeren 1 gizli katmana sahiptir. Bu modelin eğitiminde 0.01 öğrenme oranı, 0.001 sapma değeri, ReLU aktivasyon fonksiyonu ve Nadam optimizasyon algoritması kullanılmıştır.

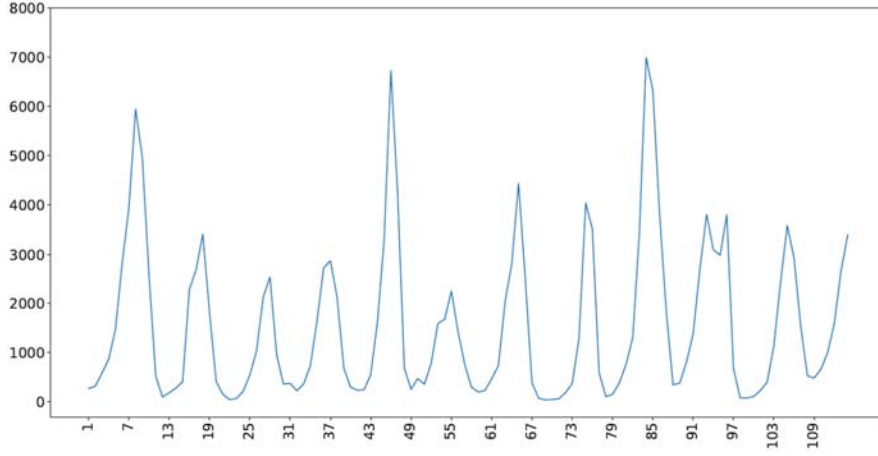
Lynx veri setinde tekil kullanılan yöntemler arasında, ARIMA, Sunspot veri setinin tersine YSA'ya kıyasla daha iyi bir doğruluk sonucu sağlamaktadır (Tablo 3). Bu durum, Lynx veri setinin Sunspot veri setine kıyasla daha durağan olmasından kaynaklanmaktadır. Bu durağanlık nedeniyle, çıkarılan öznelikler kendi içlerinde daha fazla düzenlilik göstermekte ve zaman serileri karakteristiğini daha doğru bir şekilde temsil etmektedir. Bu sayede, yöntemimiz, her iki hata metriğinde de hibrit yöntem doğruluk oranını %10'dan fazla artırmıştır.

Gbpsd veri seti 0,58 ADF test sonucuyla, oldukça yüksek volatilitelere sahip ve durağan olmayan bir veri setidir. Bu

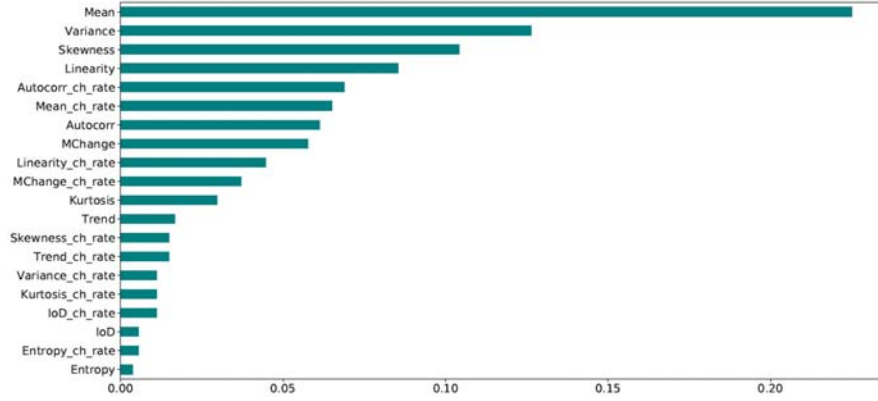
yüksek volatilitelere, zaman serisindeki sayısız değişim noktalarını gösteren Şekil 6'daki görselinde de görülmektedir. Doğrusal bileşenin elde edilmesi için kullanılan MA filtresinin uzunluğu 40 olarak çıkmaktadır. Elde edilen doğrusal bileşenin 0,007 çıkan ADF test sonucu, bu bileşenin doğrusal bir ARIMA yöntemi ile modellenmeye uygun olduğunu göstermektedir.

Gbpsd veri seti için hesaplanan öznelik önem puanları incelendiğinde, diğer veri setlerinde de görüldüğü gibi *Mean* özneliğinin tahmin performansına en önemli katkıyı yaptığı görülmektedir (Şekil 7). Bu veri setinde yaptığımız deneyler sonucu, en yüksek puanlı ilk 7 öznelik final YSA modellenmesinde kullanılmak üzere seçilmiştir. Ayrıca, final YSA modeline, verilen zaman serisinin son 5 değeri, doğrusal olmayan kalan serinin son 3 değeri ve doğrusal bileşen tahmininin sonucu da girdi olarak verilmektedir. Burada kullanılan YSA modeli 33 tane düğüm içeren 1 gizli

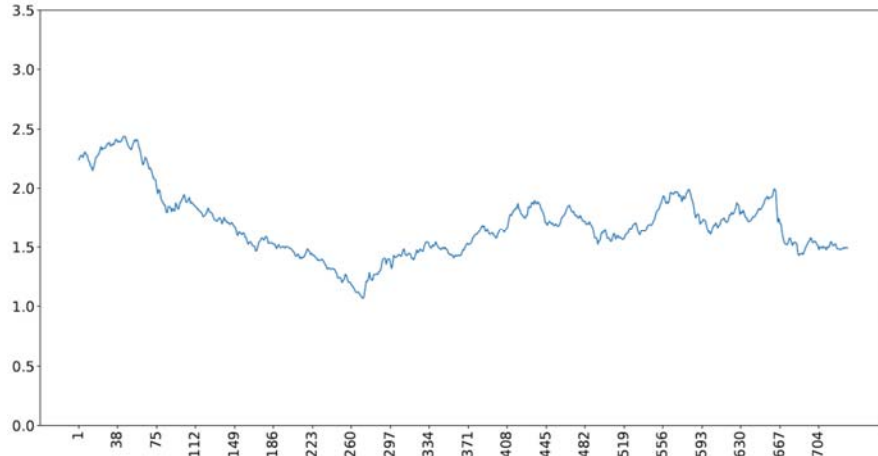




Şekil 4. Lynx serisi (1821-1934) (Lynx series (1821-1934))



Şekil 5. Lynx veri seti için öznelik önemi (Feature importance for Lynx dataset)



Şekil 6. GbpUsd Serisi (1980-1993) (Gbp/Usd seies (1980-1993))

katmana sahiptir. Bu modelin eğitiminde 0.01 öğrenme oranı, 0.001 sapma değeri, ReLU aktivasyon fonksiyonu ve Nadam optimizasyon algoritması kullanılmıştır.

GbpUSD veri seti sonuçları incelendiğinde ( Tablo 3), YSA metodu yüksek düzeyde volatilitesi olan bu veri seti için, doğrusal bir metod olan ARIMA'dan daha iyi bir performans vermektedir. Hibrit yöntem, tekil metotlara göre daha iyi bir

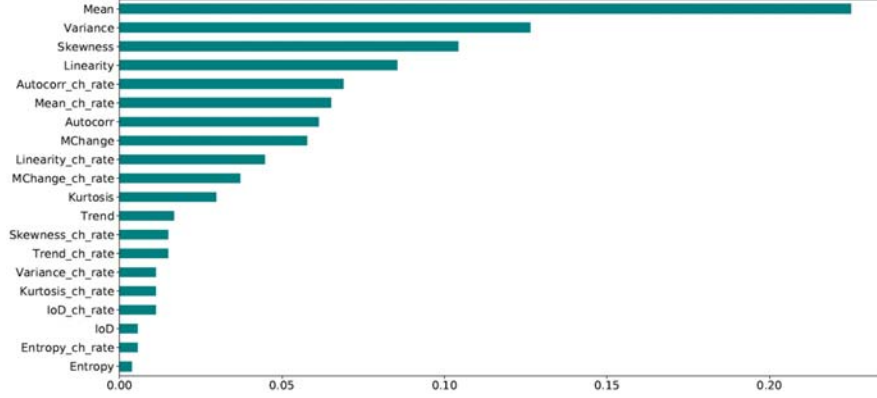
doğruluk performansı sağlasa da önerilen yöntem, seçilen öznelikleri kullanarak yüksek volatilitayı diğer metotlardan daha iyi bir şekilde kavramaktadır. Önerilen yöntem MAE ve MSE hata ölçümlerinde hibrit metod sonuçlarını sırasıyla %5,5 ve %3,7 iyileştirmeyi başarmaktadır.

Intraday veri seti 0,27 ADF test sonucuyla, oldukça yüksek volatiliteye sahip ve durağan olmayan bir zaman serisidir (

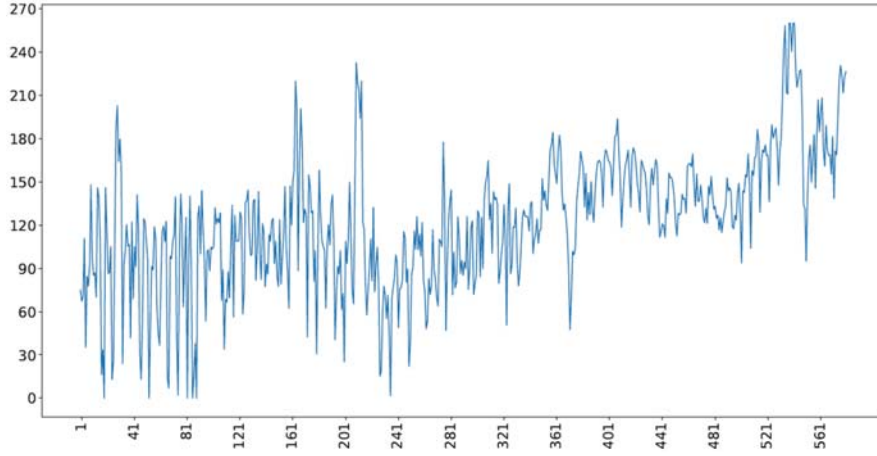
Şekil 8). Önerilen yöntemin hibrit kısmında, uzunluğu 6 olan bir MA filtresi, ADF test sonucunu 0,004 veren, istenen durağanlıkta bir bileşeni çıkarabilmektedir.

Şekil 9 bu veri seti için hesaplanan özneliklerin önem puanlarına göre sıralanmış halini göstermektedir. *Mean* özneliği tahmin performansına en yüksek katkıyı sağlarken, *Cpoints* özneliği en düşük katkıyı yapmaktadır. Bu veri setinde yapılan deneyler sonucunda, en yüksek önem

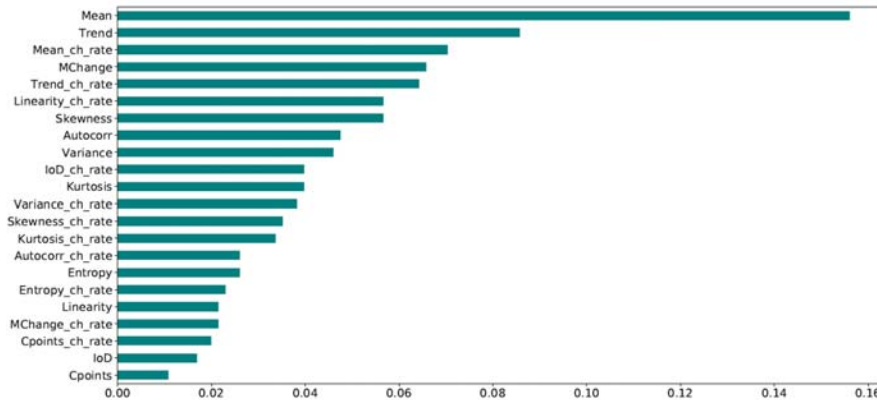
puanına sahip olan ilk 9 özellik final YSA modellemesinde kullanılmak üzere seçilmiştir. Ayrıca, final YSA modeli, verilen zaman serisinin son 8 değerini, doğrusal olmayan kalan serinin son 8 değerini ve doğrusal bileşenin tahminini girdi olarak almaktadır. Burada kullanılan YSA modeli 53 tane düğüm içeren 1 gizli katmana sahiptir. Bu modelin eğitiminde 0.01 öğrenme oranı, 0.001 sapma değeri, ReLU aktivasyon fonksiyonu ve Nadam optimizasyon algoritması kullanılmıştır.



Şekil 7. GbpUsd veri seti için öznelik önemi (Feature importance for GbpUsd dataset)



Şekil 8. Intraday serisi (Temmuz 2015 - Aralık 2017) (Intraday series (Jul. 2015 - Dec. 2017))



Şekil 9. Intraday veri seti için öznelik önemi (Feature importance for Intraday dataset)

Tablo 3'teki Intraday veri seti için olan sonuçlar incelendiğinde, doğrusal bir metot olan ARIMA'nın, bu yüksek volatiliteli veri setinin doğrusal olmayan hareketlerini yakalamak için yeterli olmadığı görülmektedir. Önerilen yöntem ise, veri setinin karakteristiğini hesaplayan, seçilmiş özniteliklerin avantajlarını kullanır. Bu sayede hibrit yöntemin MAE ve MSE hata ölçümleri sırasıyla %8,1 ve %16,1 oranında gelişmektedir.

#### 4. SONUÇLAR (CONCLUSIONS)

Pratikteki çeşitli uygulamalardan türetilen zaman serisi verileri, genel olarak doğrusal ve doğrusal olmayan varyasyonlar içerir. Bu farklı varyasyonları yakalamak için, doğrusal ve doğrusal olmayan yöntemlerin kuvvetli taraflarını birleştiren çeşitli hibrit yöntemler literatürde önerilmiştir. Hibrit yöntemler, her iki yöntem tipinin avantajlarını aynı anda kullanabildikleri için tekil yöntemlerden daha iyidir. Fakat hibrit modeller, serideki bir sonraki değeri tahmin etmek için önceki zaman adımlarındaki değerleri kullanarak, tek değişkenli yöntemler olarak tasarlanmıştır. Serilerin altta yatan yapısal özelliklerini dikkate almazlar. Bu çalışmada, zaman serilerinde tahmin performansını artırmak için bir metot önerilmektedir. Bu metot, zaman serilerinin anlamlı özetini sunan özniteliklerini hesaplayarak, en açıklayıcı olanları hibrit bir yöntemle girdi olarak sunmaktadır. Her bir serinin en açıklayıcı özniteliklerini bulmak için, özniteliklerin önem puanı gradyan artırma ağaçları kullanılarak ayrı ayrı hesaplanmaktadır. Daha sonra, en bilgilendirici öznitelikler önem puanı sırasına göre seçilmekte ve hibrit yöntemle diğer girdilerle beraber sunulmaktadır. Deneyler herkese açık referans veri setlerinde gerçekleştirilmiştir. Deneysel sonuçlarımız, öznitelik tabanlı hibrit yöntemimizin, incelenen diğer tüm yöntemlere kıyasla daha yüksek bir doğruluk performansı sağladığını göstermektedir.

#### KAYNAKLAR (REFERENCES)

1. Weigend, A, Time series prediction: forecasting the future and understanding the past, New York: Routledge, 2018.
2. S. Li and R. Li, Comparison of Forecasting Energy Consumption in Shandong, China Using the ARIMA Model, GM Model, and ARIMA-GM Model Sustainability, vol. 9, no. 7, p. 1181, Jul. 2017.
3. M. Valipour, M. E. Banihabib, and S. M. R. Behbahani, Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir, Journal of Hydrology, vol. 476, pp. 433–441, Jan. 2013
4. B. Siregar, E. B. Nababan, A. Yap, U. Andayani, and Fahmi, Forecasting of raw material needed for plastic products based in income data using ARIMA method, in 2017 5th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), 2017, pp. 135–139.
5. N. Apergis, Forecasting Credit Default Swaps (CDSs) spreads with newswire messages: Evidence from European countries under financial distress, Economics Letters, vol. 136, pp. 92–94, Nov. 2015.
6. V. Ediger, S. Akar, ARIMA forecasting of primary energy demand by fuelin Turkey, Energy Policy 35 (2007) 1701-1708. doi:10.1016/j.enpol.2006.05.009.
7. R. E. Soria-Guerra, R. Nieto-Gomez, D. O. Govea-Alonso, and S. Rosales-Mendoza, An overview of bioinformatics tools for epitope prediction: Implications on vaccine development, Journal of Biomedical Informatics, vol. 53, pp. 405–414, Feb. 2015.
8. A. V. Babkin, E. P. Karlina, and N. S. Epifanova, Neural Networks as a Tool of Forecasting of Socioeconomic Systems Strategic Development, Procedia-Social and Behavioral Sciences, vol. 207, pp. 274–279, Oct. 2015.
9. G. Cervone, L. Clemente-Harding, S. Alessandrini, and L. Delle Monache, Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble, Renewable Energy, vol. 108, pp. 274–286, Aug. 2017.
10. J. Kamruzzaman and R. A. Sarker, Forecasting of currency exchange rates using ANN: a case study, in International Conference on Neural Networks and Signal Processing, 2003, vol. 1, pp. 793-797 Vol.1.
11. G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and Systems (4) (1992) 455-455. doi:10.1007/BF02134016.
12. K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (5) (1989) 359-366. doi: 10.1016/0893-6080(89)90020-8.
13. P. Zhang, Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing 50, 159-175, Neurocomputing 50 (2003) 159-175. doi:10.1016/S0925-2312(01)00702-0.
14. W. R. Foster, F. Collopy, L. H. Ungar, Neural network forecasting of short, noisy time series, Computers Chemical Engineering 16 (4) (1992) 293-297. doi:10.1016/0098-1354(92)80049-F.
15. S. Aras, D. Kocako, A new model selection strategy in time series forecasting with artificial neural networks: IHTS, Neurocomputing 174 (2016) 974-987. doi:10.1016/j.neucom.2015.10.036.
16. M. Khashei, M. Bijari, A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series Forecasting, Appl. Soft Comput. 11 (2) (2011) 2664-2675. doi:10.1016/j.asoc.2010.10.015.
17. C. N. Babu, B. E. Reddy, A Moving-average Filter Based Hybrid ARIMA-ANN Model for Forecasting Time Series Data, Appl. Soft Comput. 23 (2014) 27-38. doi:10.1016/j.asoc.2014.05.028.
18. Ü. Ç. Büyükaşahin, Ş. Ertekin, Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition, arXiv:1812.11526 [cs], Jan 2019.
19. X. Wang, K. Smith, R. Hyndman, Characteristic-based clustering for time series data, Data Min. Knowl. Discov. 13 (3) (2006) 335–364.

20. X. Wang, K. Smith, R. Hyndman. (2005) Dimension Reduction for Clustering Time Series Using Global Characteristics. In: Sunderam V.S., van Albada G.D., Sloot P.M.A., Dongarra J. (eds) Computational Science – ICCS 2005. ICCS 2005. Lecture Notes in Computer Science, vol 3516. Springer, Berlin, Heidelberg
21. B. D. Fulcher and N. S. Jones. Highly Comparative Feature-Based Time-Series Classification. In: IEEE Transactions on Knowledge and Data Engineering 26.12 (Dec. 2014), pp. 3026–3037. issn: 1041-4347. doi: 10.1109/TKDE.2014.2316504.
22. RJ Hyndman, E Wang, N. Laptev, Large-scale unusualtime series detection. IEEE International Conference on Data Mining Workshop (ICDMW), 14-17 Nov. 2015. doi: 10.1109/ICDMW.2015.104.
23. N. Laptev, J. Yosinski, L. Erran, S. Smyl, Time-series Extreme Event Forecasting with Neural Networks at Uber, International Conference on Machine Learning, 2017.
24. E. T. Hibon M., To combine or not to combine: selecting among forecasts and their combinations. International Journal of Forecasting, International Journal of Forecasting 21 (2005) 15-24.
25. F. Aydin and Z. Aslan, Yapay öğrenme yöntemleri ve dalgacık dönüşümü kullanılarak nöro dejeneratif hastalıkların teşhisi, Journal of the Faculty of Engineering and Architecture of Gazi University, 32(3), 749-766, 2017.
26. D. Abidin, Ö. Öztürk, and T. Özacar Öztürk, Klasik Türk müziğinde makam tanıma için veri madenciliği kullanımı, Journal of the Faculty of Engineering and Architecture of Gazi University, 32(4), 1221-1232, 2017.
27. H. Rao et al., Feature selection based on artificial bee colony and gradient boosting decision tree, Applied Soft Computing, vol. 74, pp. 634–642, Jan. 2019.
28. R. Leardi, Application of a genetic algorithm to feature selection under full validation conditions and to outlier detection, Journal of Chemometrics, vol. 8, no. 1, pp. 65–79, 1994.
29. Friedman, J.H. Stochastic gradient boosting. Comput. Stat. Data Anal. 2002, 38, 367–378.
30. Friedman, J.; Hastie, T.; Tibshirani, R. The Elements of Statistical Learning; Springer Series in Statistics; Springer: Berlin, Germany, 2001; Volume 1.
31. Breiman, L., Friedman, J.H., Olshen, R.A., and Stone, C.I. (1984). Classification and regression trees. Belmont, Calif.: Wadsworth.
32. Intraday EPIAS, <https://www.epias.com.tr/en/intra-day-market/> introduction, 2018, (Erişim tarihi Kasım 12, 2018).