



ROBUST ARIMA MODELİ İLE YAPAY SİNİR AĞLARI MODELİNİN KIYASLANMASI: TURİZM ÖRNEĞİ

Selim DÖNMEZ^{1,*}, Özer ÖZAYDIN²

¹ Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Eskişehir, sdonmez3@gmail.com,

² Eskişehir Osmangazi Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, Eskişehir, oozaydin@ogu.edu.tr,

Geliş Tarihi:07.08.2017

Kabul Tarihi:28.04.2018

ÖZ

Yapay sinir ağları yöntemi 20. yüzyılın ilk yarısından sonra popülerleşmiş bir yeni nesil analiz yöntemidir. Örüntü sınıflandırma, ses tanımlama vs. gibi alanlarda başarılı uygulamaları olan yapay sinir ağları yöntemi zaman serileri analizinde de son yıllarda çok sıklıkla uygulanmaya başlanmıştır. Ancak bahsi geçen alanlardan farklı olarak zaman serileri analizinde başarıya etki eden etkenler çok farklıdır ve zaman serisi modellerindeki parametrelerin tahmin edilmesi, seride eğer varsa kırılma noktalar vs. öngörülerini etkilemektedir. Bu nedenle, yapay sinir ağları yönteminin zaman serileri analizinde başarı sağlaması çok zor olmaktadır. Bu sebeple yapay sinir ağında performansı geliştirecek yeni bir yöntem önererek öngörü yeteneği bu çalışmada geliştirilmiştir. Robust teknikler en küçük verilerde bile işe yarayacak etkili zaman serisi analizi prosedürleri geliştirmişlerdir. Bunlardan etkili olarak kabul edebileceğimiz bir yöntem, modifiye en çok olabilirlik yöntemidir. Bu yöntemde herhangi bir zaman serisi modeli üzerinden elde edilecek modifiye en çok olabilirlik yönteminden elde edilen parametre tahminleri ile öngörü yapılabilen ve hatalarının dağılımına göre etkin ve yansız tahmin ediciler elde edilmektedir. Bu çalışmada turizm verisi üzerinde modifiye en çok olabilirlik yöntemi ile performansı geliştirilmiş yapay sinir ağının kıyaslanması yapılacaktır.

Anahtar kelimeler: *Modifiye en çok olabilirlik yöntemi, Yapay sinir ağları, Zaman serisi analiz*

THE COMPARISON OF ROBUST ARIMA MODEL AND ARTIFICIAL NEURAL NETWORK MODEL: AN EXAMPLE OF TOURISM

ABSTRACT

Artificial neural network is an analysis procedure which became popular after the first half of twentieth century. Having many successful applications in areas such as Pattern Recognition, Sound Analysis and etc., the artificial neural network procedure has also been used in Time Series Analysis. Nevertheless, the factors that effect the success of time series analysis are different than the factors of success in the areas we mentioned previously and factors such as estimation of the time series' parameters and the existence of breaking points of time series can effect the forecasts of the model. That's why in our study we seek to develop the artificial neural network procedure. Robust Statistical Methods in time series analysis, have developed effective time series analysis procedures even though there can be a small amount of data. One of those methods is the modified maximum likelihood estimation method. In this method using any time series model, the forecasts are obtained using

modified maximum likelihood estimates of parameters and using the distribution of residuals we can obtain unbiased and efficient estimators. In this study, we will compare the modified maximum likelihood method and the enhanced artificial neural network procedure on a tourism data.

Keywords: *Modified maximum likelihood estimation method, Artificial neural networks, Time Series Analysis*

1. GİRİŞ

İstatistiksel analiz yöntemleri, “Sağlamcı” ve “Klasik” olmak üzere ayrı ayrı değerlendirilebilir. Bu ayrımın yapılmasındaki temel amaç, istatistiğin gelişiminin veri analizi yöntemlerini nasıl etkilediğini açıklamaktır. Örneğin klasik istatistik ya da başka bir deyişle sıklıkçı istatistik (frequentist statistics) hipotez testlerinin gerçekleştirilmesinde ve parametreler için güven aralıklarının oluşturulmasında önemli bir rol üstlenmiştir. Sağlamcı istatistik (robust statistics) varsayımlardan bağımsız olarak etkin istatistiksel tahmin etme teknikleri üzerinde durmaktadır ve diğer istatistiksel yaklaşımlara göre daha yeni bir yaklaşım türüdür.

Yapay sinir ağları, yapay zekâ ile beraber gelişen bir kavramdır. Bu kavrama ilişkin ilk eserlerden birini yazan Donald Hebb, 1949 yılında Organization of Behaviour (Davranışın organizasyonu) adlı eseriyle ilk defa öğrenebilen bir yapay sinir ağı modeli oluşturulmuştur. Bundan sonra 1957 yılındaki Rosenblatt’ın ortaya attığı tekli doğrusal algılayıcı ve 1959 yılında Bernard Widrow ile Marcian Hoff’un ortaya attıkları ADALINE ve MADALINE ağ modelleri ortaya çıkmıştır. 1970’lerde her ne kadar yapay sinir ağlarına ilişkin çalışmalar durma noktasına gelse de 1980’lerde bu durum değişmiş ve çok büyük hızla gelişmeler gerçekleşmiştir. Bugünlerde popüler konulardan biri, zaman serileri analizinin yapay sinir ağlarıyla nasıl yapılabileceğidir. Bu konuyla ilgili pek çok çalışma yapılmıştır.

Azadeh vd. [1], yapay sinir ağlarını, bulanık doğrusal regresyonu ve konvansiyonel doğrusal regresyonunu birleştirerek yeni bir yöntem oluşturmuştur. Önerilen yöntemin değişik verilere uygulanabilecek kadar esnek olduğu belirtilmektedir. Önerilen yöntem Amerika’daki ve İran’daki demir fiyatları verilerine uygulanmış ve olumlu sonuçlar elde edilmiştir.

Jhajharia vd. [2], sıtma vakaları üzerine yapay sinir ağları metodolojisini uygulamıştır. Bundaki temel amaç sıtma yayılımının hava koşullarından etkilenip etkilenmediğini test etmektir. Hindistan’ın kuzeyindeki çölde, değişik virüsler tarafından yayılan sıtma salgınındaki trendler Mann-Kendall testi ile test edilmiş sonrasında sıtma verisi ak gürlütlü hale getirilmiştir. Bundan sonra iç bağıntılılık problemi incelenmiş ve temel bileşenler analiziyle iç bağıntı problemi tamamen ortadan kaldırılmıştır. Analiz sonunda öngörü için yapay sinir ağları kullanılmış ve virüslerden biri olan P. Falciparum’un çöl şartlarından etkilendiği görülmüştür.

Khatibi vd. [3], Tebrizdeki hava kirliliği verisini kullanarak çoklu doğrusal regresyonu, yapay sinir ağlarını, genetik programlama, doğrusal olmayan yerel tahmin modeli ile ARIMA modelini kıyaslamıştır. İlk üç modelleme tekniğinde açıklayıcı değişkenler, sıcaklık, güneşin radyasyonu, çiğ oluşma derecesi ve rüzgâr hızı olarak alınmıştır. Ozon tabakasına ilişkin veri, 2010 Ağustosundan 2011 Martına kadarki saatlik verilerden oluşan bir zaman serisidir. Makalede çoklu doğrusal regresyon, yapay sinir ağları, genetik programlama modelleri etkili sonuçlar veremese de doğrusal olmayan yerel programlama ile ARIMA modeli etkili olmuştur. Bu çalışmada, parametreleri modifiye en çok olabilirlik yöntemi ile tahmin edilmiş ARIMA modeli ile sinaptik bağlantılar üzerinde kısıtlamalar getirilen yapay sinir ağı modeli karşılaştırılacaktır.

2. ARIMA MODELİ ÜZERİNDE MODİFİYE EN ÇOK OLABİLİRLİK YÖNTEMİNİN KULLANIMI

Temeli Tiku'nun makalelerine dayanan bu yöntem, en çok olabilirlik yönteminin uygulanmasına ilişkin zorluklarını telafi etmek için ortaya atılmıştır [4-9]. Bu yöntemin en ilginç ve en önemli yanı, belli koşullar altında elde edilen modifiye en çok olabilirlik tahmin edicilerinin en çok olabilirlik tahmin edicileriyle asimptotik olarak eşdeğer olmasıdır [10]. Bu özelliğinden ötürü, modifiye en çok olabilirlik tahmin edicileri en çok olabilirlik tahmin edicilerinin bütün özelliklerini elde etmektedir.

Zaman serileri analizinde basit bir otoregresif model Eş. 1 ve 2 ile ifade edilebilmektedir:

$$y_t = \mu + \delta x_t + e_t \quad (1)$$

$$e_t = \phi e_{t-1} + a_t \quad (2)$$

Burada y_t gözlenen rassal değişken y 'nin t . zamanda aldığı değer, x_t stokastik olmayan değişkenin önceden belirlenmiş değeri, a_t hata olarak adlandırılmaktadır. Burada a_t 'ler birbirinden bağımsız olarak aynı dağılıma sahip olmaktadır. Eş. 1 ve 2'deki formülün bir alternatifi aşağıdaki şekildedir:

$$y_t - \phi y_{t-1} = \mu + \delta(x_t - x_{t-1}) + a_t \quad (3)$$

Bu formüle göre, şayet $\phi = 0$ olursa elde edilecek denklem standart bir doğrusal regresyon denklemi olmaktadır ve $\phi \neq 0$ ve $\delta = 0$ olursa elde AR(1) modeli kalmaktadır. modifiye en çok olabilirlik tahmin edicileri dağılımlara göre değişik formüllerle ifade edilebilmektedir. Bu sebepten Akkaya ve Tiku [10] eserlerinde üç farklı dağılımlar ailesi tanımlamış ve bunun üzerinden modifiye en çok olabilirlik tahmin edicilerinin formüllerini çıkarmışlardır. Tanımlanan dağılımlar uzun kuyruklu simetrik dağılımlar ailesi, kısa kuyruklu simetrik dağılımlar ailesi ve çarpık dağılımlar ailesidir [10].

Bu dağılım ailelerine göre, değişik tahmin ediciler çıkarabilmektedir. Bunlara göre çıkartılan tahmin ediciler, çok kullanışlı özelliklere sahiptir. Bu durumda tahmin edicilerin kullanılmaları, analizde çok önemli sonuçlar elde etmekte önemli rol üstlenebilir. Bu dağılım ailelerinden uzun kuyruklu dağılıma sahip herhangi bir x değişkeni için $\sqrt{\frac{v}{2p-3}} \left(\frac{x-\mu}{\sigma} \right)$ değişkeni v serbestlik dereceli t dağılımına sahiptir [11], ancak integralin çözümü oldukça zordur ve Simpson yönteminin alternatif bir kullanımı gerekmektedir [12]. Hesaplama zorluklarından doğan tekniklerden ötürü t -dağılımından sayı üretilip $t_{(i)}$ hesaplanmıştır.

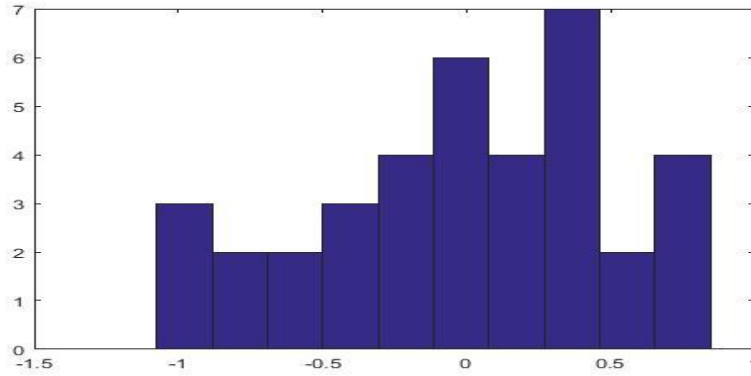
3. YAPAY SİNİR AĞLARI YÖNTEMİ

Yapay sinir ağları yöntemi, 1940'lı yıllarda ortaya çıkmış ve kısa zamanda çok popüler bir analiz yöntemi olmuştur. 1970'li yıllara gelindiğinde miladını doldurduğu düşünüldüğünde, sınıflandırma becerileri ortaya konulunca canlanmış [13]. Bu yöntem, oluşturduğu modelde hatalarla ilgili dağılımsal herhangi bir varsayıma ihtiyaç duymamaktadır. Yine de yapay sinir ağları yöntemi, modellemedeki hataların ak gürültüye sahip olmasına ihtiyaç duyar. Yapay sinir ağı yapısı içerisinde bulunan sinaptik ağırlıklar, ak gürültüye sahip hatalar aracılığıyla hata düzeltmeli öğrenme, Hebb'çi öğrenme, rekabete dayalı öğrenme, Boltzmann öğrenmesi güncellenir.

4. TURİZME YÖNELİK UYGULAMA

4.1. Türkiye’de Yurtdışından Gelip Sınırlardan Çıkış Yapan İlkokul Mezunu Turist Sayısının Öngörüsü

Veri olarak yurtdışından gelip sınırlardan çıkış yapan ilkököl mezunu turist sayısını kullandık. Uygulamada verinin önce doğal logaritması sonra farkı alınarak analiz edilmiştir. Yapay sinir ağı tekniği uygulanırken modifiye en çok olabilirlik tekniğiyle katsayılar hesaplanmış ve ona göre aralıklar belirlenmiştir. Bu aralıklardan rasgele seçim yapılarak sinaptik katsayılar belirlenmiştir. Modifiye en çok olabilirlik yönteminde elde edilen hata dağılımı Şekil 1’deki gibidir:



Şekil 1. Hataların histogramı.

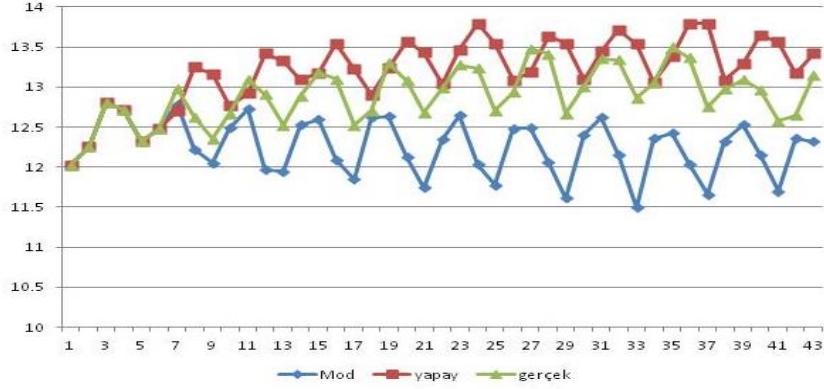
Hata dağılımında birden fazla mod göze çarpmaktadır ve hataların dağılımının $r > d > 0$ olan bir kısa kuyruklu dağılıma sahip olduğuna kanaat getirilmiştir [10]. Bunun yanında hatalarının dağılımının çarpıklığı ve basıklığı 0.1339 ile 1.4534 bulunmuştur ve $d > 0$ için basıklık nispeten daha düşük çıkmaktadır. Bu nedenle $r=4$ ve $d=-100$ parametrelili kısa kuyruklu dağılım seçilmiş ve hesaplanan olabilirlik fonksiyonunun logaritması 181.835 bulunmuştur. Bu seçilen dağılım sonucunda elde edilen hata kareler ortalamasının kökleri yapay sinir ağı tekniği için 0.161, modifiye en çok olabilirlik tekniği için 0.7591 bulunmuştur. Gerçek veride uygulanan yapay sinir ağı ve modifiye en çok olabilirlik modellerinin hata kareler ortalamasının kökleri 75842 ile 244370 bulunmuştur. Modeller sırasıyla yapay sinir ağı ve modifiye en çok olabilirlik tekniği için Eş. 4 ve 5 ile oluşturulmuştur:

$$\Delta Y_t = -0.0115 - 0.4731\Delta Y_{t-1} + 0.8999\Delta Y_{t-4} + 0.425\Delta Y_{t-5} + e_t \quad (4)$$

$$\Delta Y_t + 0.4815\Delta Y_{t-1} = -0.0092 + 0.9099(\Delta Y_{t-4} + 0.4815\Delta Y_{t-5}) + e_t \quad (5)$$

Burada Δ zaman serisinin farkını göstermektedir. Bunun sonucunda gerçek verilerle beraber Şekil 2’deki grafik elde edilmiştir:

Burada Mod serisi modifiye en çok olabilirlik yöntemiyle elde edilen tahminleri, yapay serisi yapay sinir ağı yöntemiyle elde edilen tahminleri ve gerçek serisi ise gerçek değerleri ifade etmektedir.



Şekil 2. MEÇO ve yapay sinir ağları yöntemleriyle elde edilen öngörüler ve orjinal veri.

Grafikten görüldüğü üzere yapay sinir ağları tekniğinin modifiye en çok olabilirlilik tekniğine göre daha iyi sonuç verdiği söylenebilir. Elde edilen modellerle gelecek 4 dönemin öngörülere çizelge 1'deki gibidir:

Çizelge 1. İki yönteme göre günümüzdeki dört döneme ilişkin öngörüler.

Dönemler	Modifiye en çok olabilirlilik tekniği için	Yapay sinir ağları tekniği için	Gerçek değerler
2013-IV	146666	501655	483354
2014-I	138869	289346	279854
2014-II	313527	325405	362548
2014-III	151511	366923	468684

Bu öngörülerin tamamında gerçek değerlere en yakın öngörülerin yapay sinir ağları tekniğiyle elde edildiği görülmektedir.

5. SONUÇ VE ÖNERİLER

Yapay sinir ağlarını oluştururken getirdiğimiz esaslar, daha iyi bir model kurulmasına hizmet etmektedir. Ayrıca dinamik yapısı sayesinde oluşturulan model, kendi kendini veriye uyarlayacaktır. Modelimiz istatistiksel yöntemle elde edilen tahminlerin daha iyi bir model kurulumuna hizmet ettiğini göstermiştir ve metodolojimizim başarısı burada yatmaktadır. Bir veri olarak Türkiye'de yurtdışından gelip sınırlardan çıkış yapan ilkökul mezunu turist sayısı verisi iki yöntemin kıyaslanması için son derece önemli olmuştur.

KAYNAKÇA

- [1] Azadeh, A., Neshat, N., Mardan, E., Saberi, M. (2013). Optimization of steel demand forecasting with complete and uncertain economic inputs by an integrated neural network–fuzzy mathematical programming approach. *International Journal of Advanced Manufactured Technology*, 65, 833–841.
- [2] Jhajharia, D., Chattopadhyay, S., Choudhary, R. R., Dev, Vas S., Vijay, P., Lal, S. (2013). Influence of climate on incidences of malaria in the Thar Desert for Northwest India. *International Journal of Climatology*, 33(2), 312-325.
- [3] Khatibi, R., Naghipour, L., Ghorbani, M. A., Smith, M. S., Karimi, V., Farhoudi, R., Delafrouz, H., Arvanaghi, H. (2013). Developing a predictive tropospheric ozone model for Tabriz. *Tabriz Atmospheric Environment*, 68, 286–293.
- [4] Tiku, M. L. (1967a). Estimating the mean and standard deviation from a censored normal sample. *Biometrika*, 54, 155-165.
- [5] Tiku, M. L. (1967b). A note on estimating the location and scale parameters of the exponential distribution from a censored sample. *Australian Journal of Statistics*, 9, 49-53.
- [6] Tiku, M. L. (1968a). Estimating the parameters of log-normal distribution from censored samples. *Journal of American Statistical Association*, 63, 134-140.
- [7] Tiku, M. L. (1968b). Estimating the parameters of normal and logistic distributions from censored samples. *Australian Journal of Statistics*, 10, 64-73.
- [8] Tiku, M. L. (1968c) "Estimating the mean and standard deviation from progressively censored normal samples" *J. Ind. Soc. Agric. Stat.*, 20, 20-25.
- [9] Tiku, M. L. (1970). Monte Carlo study of some simple estimators in censored normal samples. *Biometrika*, 57, 207-210.
- [10] Akkaya, A., Tiku, M.L. (2004). *Robust Estimation and Hypothesis Testing*. New Age International (P) Ltd, New Delhi.
- [11] Trawinski, B.J., Bechhofer, R.E., Tamhane, A.C., Tiku, M.L., Kumra, S. (1985). *Selected tables in mathematical statistics*. American Mathematical Society.
- [12] Rowland, J.H., Varol, Y.L. (1972). Exit criteria for Simpson's compound rule. *Math Computation*, 26, 119, 699-702.
- [13] Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation (İkinci Baskı)*, Prentice-Hall, New Jersey.