

Modelleme ve Tahmin Amaçlı Veri Ön İşleme Yöntemlerinin Ürün Kurutma Örneği ile Açıklanması

Explaining Data Preprocessing Methods for Modeling and Forecasting with the Example of Product Drying


Cem KORKMAZ^{1*}, İlyas KACAR²

Öz

Regresyon geleneksel bir veri işleme yöntemi olsa da, makine ve derin öğrenme yöntemleri hem modelleme hem de tahmin için son yıllarda literatürde yaygın olarak kullanılmaktadır. Ancak, bu yöntemlerin verimli bir şekilde kullanılabilmesi için veri türünün anlaşılmasına yönelik bir ön değerlendirme yapılması önem arz etmektedir. Bu nedenle, bu çalışmada ön değerlendirme prosedürleri açıklanmıştır. Kurutma deney düzeneğinde kullanılan ölçüm cihazları ve sensörlerindeki ölçüm belirsizliklerini belirlemek için deneysel belirsizlik analizi yapılmıştır. Veri setindeki değişkenler arasındaki anlamlı ve anlamsız ilişkiler Pearson korelasyon matrisi ile belirlenmiştir. Kurutma verilerindeki zaman serisi gecikmesini belirlemek için otokorelasyon ve kısmi otokorelasyon fonksiyonları kullanılmış ve 5 gecikmeli bir AR(5) serisi belirlenmiştir. Kurutma sürecinin doğal davranışından kaynaklanan ham verilerdeki tepe ve dip noktalar nedeniyle verilerin değişken varyansa sahip olduğu görülmüştür. Ham verileri bozmadan gerçekleştirilen normalizasyon ön değerlendirme işlemi ile modelleme başarısı elde edilmiştir. Böylece geleneksel modellere göre daha iyi modeller elde edilebileceği gösterilmiştir. Makine öğrenmesi yönteminde gizli katman ve nöron sayısını belirlemek için kullanılan deneme yanılma yöntemindeki, gereksiz zaman ve hesaplama maliyetlerinden kaçınmak için, literatürde önerilen çeşitli formüller karşılaştırılmıştır. Korelasyon katsayısının, modelin iyiliğini belirlemede tek başına yeterli olmadığı gösterilmiştir. Bu çalışmadaki verilerin modellenmesinde, NARX modelinin ANFIS ve LSTM modellerine göre daha hızlı ve daha az hata ile istenen değere yakınsadığı görülmüştür. Döner tamburlu bir kurutucunun simülasyonunda, ağ bağımsızlık analizi ile optimum ağ elemanı sayısı 1137 olarak belirlenmiştir. Bu sayede gereksiz aşırı hesaplamaların önüne geçilmiştir. Elbette tüm bu yöntemler istatistik biliminde zaten mevcuttur. Ancak bu çalışmada özellikle bu alanın dışında olan genç araştırmacıların hız kazanması ve kolay kavrayabilmesi için modelleme ve tahmin amaçlı kullanılacak yöntemler, özenle seçilmiş ve örneklerle açıklanmıştır.

Anahtar Kelimeler: Veri işleme, Regresyon, Pearson, Oto korelasyon fonksiyonu, Kısmi oto korelasyon fonksiyonu, Heteroscedasticity, Yakınsama, Doğrulama

^{1*}**Sorumlu Yazar/Corresponding Author:** Cem Korkmaz, Çukurova Üniversitesi, Ziraat Fakültesi, Tarım Makineleri ve Teknolojileri Mühendisliği Bölümü, Adana, Türkiye. E-mail: ckorkmaz@cu.edu.tr  ORCID: 0000-0003-1062-4581

²İlyas Kacar, Niğde Ömer Halisdemir Üniversitesi, Mühendislik Fakültesi, Mekatronik Mühendisliği Bölümü, Niğde, TÜRKİYE E-mail: ikacar@gmail.com  ORCID: 0000-0002-5887-8807

Atf: Korkmaz, C., Kacar, İ. (2024). Modelleme ve tahmin amaçlı veri ön işleme yöntemlerinin ürün kurutma örneği ile açıklanması. *Tekirdağ Ziraat Fakültesi Dergisi*, 21(2): 482-500.

Citation: Korkmaz, C., Kacar, İ. (2024). Explaining data preprocessing methods for modeling and forecasting with the example of product drying. *Journal of Tekirdag Agricultural Faculty*, 21(2): 482-500.

©Bu çalışma Tekirdağ Namık Kemal Üniversitesi tarafından Creative Commons Lisansı (<https://creativecommons.org/licenses/by-nc/4.0/>) kapsamında yayımlanmıştır. Tekirdağ 2024

Abstract

Although regression is a traditional data processing method, machine and deep learning methods have been widely used in the literature in recent years for both modelling and prediction. However, in order to use these methods efficiently, it is important to perform a preliminary evaluation to understand the data type. Therefore, pre-evaluation procedures are described in this study. Experimental uncertainty analysis was performed to determine the measurement uncertainties in the measurement devices and sensors used in the drying experimental setup. Significant and insignificant relationships between variables in the data set were determined by Pearson correlation matrix. Autocorrelation and partial autocorrelation functions were used to determine the time series lag in the drying data and an AR(5) series with 5 lags was determined. The data were found to have variable variance due to peaks and troughs in the raw data resulting from the natural behaviour of the drying process. Modelling success was achieved with the normalisation pre-evaluation process performed without distorting the raw data. Thus, it has been shown that better models can be obtained compared to traditional models. In order to avoid unnecessary time and computational costs in the trial and error method used to determine the number of hidden layers and neurons in the machine learning method, various formulas proposed in the literature were compared. It is shown that the correlation coefficient alone is not sufficient to determine the goodness of the model. In modelling the data in this study, the NARX model was found to converge to the desired value faster and with less error than ANFIS and LSTM models. In the simulation of a rotary drum dryer, the optimum number of mesh elements was determined as 1137 by mesh independence analysis. In this way, unnecessary over-calculations were also prevented. Of course, all these methods are already available in statistical science. However, in this study, the methods to be used for modelling and prediction purposes are carefully selected and explained with examples, especially for young researchers who are outside this field to gain speed and easy comprehension.

Keywords: Data processing, Regression, Pearson, Autocorrelation function, Partial autocorrelation function, Heteroscedasticity, Convergence, Validation

1. Giriş

Modelleme/tahmin yöntemleri, tüm bilimsel çalışmalarda olduğu gibi tarımsal alanda yapılan araştırmalarda da yaygın olarak kullanılmaktadır. Özellikle kurutma çalışmalarında bu yöntemlere sıkça başvurulmaktadır. Bu amaçla yıllardır kullanılan ve yaygın bir yöntem olan regresyonun (Çelen, 2016; Karacabey ve ark.,2020) günümüzde makine ve derin öğrenme yöntemleri ile birlikte daha hassas modeller üretilebildiği görülmüştür. Modelleme yöntemleri; *model esaslı* ve *veri esaslı* olmak üzere iki kategoriye ayrılmaktadır (Tobon-Mejia et al., 2012; Amiroh et al., 2021). Model esaslı olan yöntemler, ilgili bilim alanındaki yasalara dayanmakta ve kesin çözüm vermektedirler. Ayrıca, model esaslı yöntemler, tahminleri artık (kalıntı) kullanarak yapmaktadırlar. Bununla birlikte, gerçek sistemlerin karmaşık doğası nedeniyle bu tür matematiksel modellerin oluşturulması nispeten zordur. Gerçek sistemlerin karmaşık yapısı kabuller ile basitleştirilse de ne kadar çok kabul yapılırsa sonuçlar üzerinde o kadar çok hata oluşabilmektedir. Veri esaslı yöntemler, sistemden elde edilen veri setinin istatistiksel özelliklerini öğrenerek tahmin gerçekleştirmektedir. Bu yöntemler genellikle karmaşık sistemlerin tanımlanması için tercih edilmektedir. Her iki yöntemin derinlemesine araştırılmaları, karşılaştırılmaları yapılmış ve veriye dayalı yöntemlerin daha avantajlı olduğu belirlenmiştir (Dash and Venkatasubramanian, 2000; Venkatasubramanian et al., 2023). Veriye dayalı yöntem ayrıca iki kategoriye ayrılmaktadır. Birincisi, sinir ağı ve bulanık mantık tarafından temsil edilen *yapay zekâ* yöntemidir. En sık kullanılan yapay zekâ yöntemleri; nonlineer otomatik gerilemeli harici girişli model (NARX), çok katmanlı algılayıcı (MLP), uzun kısa süreli bellek (LSTM) ve uyarlamalı ağı tabanlı bulanık çıkarım sistemi (ANFIS) modellerdir. Son yıllarda hibrit modeller de incelenmektedir. Bununla birlikte daha pek çok yapay zekâ yöntemi bulunmaktadır. Veriye dayalı ikinci yöntem ise *istatistiksel öğrenme* yöntemleridir (Dong et al., 2020). *Mevcut veriyi* temsil etmek için ARIMA, SARIMA, SARIMAX, Poisson otomatik gerileyen, doğrusal vb. gibi uygun bir regresyon modelleri kullanılmaktadır. Mevcut *verinin dışındaki* noktaların tahmininde ise destek vektör makinesi (SVM), gizli Markov modeli (HMM) vb yöntemler kullanılmaktadır.

Her ne kadar yapay zekâ yöntemleri, veri setine bir ön işlem yapmayı gerektirmese de kullanıcının, verinin doğasını anlaması, uygun makine öğrenme parametrelerini seçebilmesi ile elde edilen modelin doğrulanmasında kolaylık sağlamaktadır. Yöntemlerin esası istatistik bilimi olmakla birlikte, uzmanlık alanı daha farklı disiplinler olan araştırmacıların, böylesi genel yöntemleri kendi alanında uygulayabilmesi için, tüm istatistik bilimini öğrenmek yerine, hızlı bir başlangıç olarak ilgili yöntemleri kavraması, alan dışı araştırmacılara çok büyük zaman kazancı sağlayacaktır. Çalışmanın ikinci bölümünde, yöntemlerde kullanılmak üzere zaman serisi hâlinde deneysel veri toplanması, belirsizlik analizi, hata kriterleri, korelasyon matrisi, model randıman değerlendirme, zaman gecikmesi tayini, gizli katman ve nöron sayılarının tespiti için kullanılan usul ve yöntemler verilmiştir. Üçüncü bölümde bir modele varyans, regresyon, hata yakınsama analizi yapılması ve model doğrulama yöntemleri verilmiştir. Dördüncü ve son bölümde ise çalışma sonunda elde edilen temel çıkarımlar sunulmuştur.

2. Materyal ve Metod

Bu çalışmada kullanılan veriyi toplamak üzere *Şekil 1*'de verilen deney düzeneği kullanılmıştır. Isı aktarımı, fırın içerisindeki fan kullanılmadığında, doğal konveksiyon ile veya fan kullanılarak zorlanmış konveksiyon yollarıyla gerçekleşebilmektedir. Fırın, nem, sıcaklık, güç, kütle ve süre ölçüm sensörleri ile donatılmıştır. Kurutma işleminde leonardit organik maddeli organomineral toprak düzenleyici ticari gold sülfür gübresi kullanılmıştır (Agrotime, 2023). Ürün 70 °C'de kurutulmuştur. Kurutma süresince 35741 adet veri toplanmıştır. Bu deneysel işlem neticesinde hesaplanan nem oranı, difüzyon katsayısı değerleri kullanılarak döner tamburlu bir fırındaki kurutma işleminin simülasyonları yapılmıştır (Kacar ve Korkmaz, 2021).

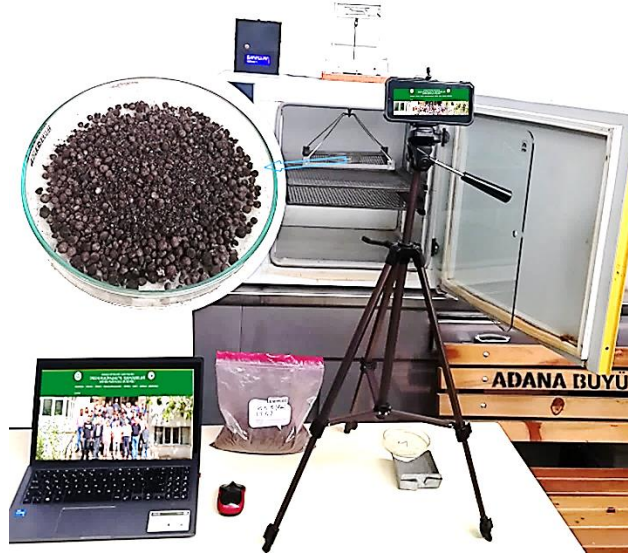


Figure 1. Drying experiment setup

Şekil 1. Kurutma deney düzeneği

2.1. Belirsizlik Analizi

Deneyisel çalışmalarda kullanılan ölçüm aletlerinin, sensörlerin belirli ölçüm aralıkları ve çözünürlükleri mevcuttur. Ölçüm aletlerinin hassasiyetlerindeki kabalıklar, sonuç üzerinde ufak da olsa belirli bir hataya sebep olmaktadır. Bu hata miktarını görebilmek için belirsizlik analizi yapılmaktadır. Deneyisel belirsizlik analizi Eşitlik (1) kullanılarak hesaplanmaktadır (Khanlari et al., 2020).

$$U_R = \sqrt{\left(\frac{\partial M}{\partial x_1} U_1\right)^2 + \left(\frac{\partial M}{\partial x_2} U_2\right)^2 + \dots + \left(\frac{\partial M}{\partial x_n} U_n\right)^2} \quad (\text{Eş. 1})$$

Burada U_R toplam belirsizliktir. M , ölçülen boyuttur (nem kütlesi). x_n (zaman, sıcaklık), ölçülen bu boyutları etkileyen n adet bağımsız değişkenden oluşan bir kümedir. U_n , bağımsız değişkenleri ölçmek için kullanılan her bir cihazın kendi ölçüm hatasıdır. Ölçüm cihazlarının hata aralıkları ve hesaplanan belirsizlikleri *Tablo 1*'de verilmiştir.

Tablo 1. Kurutma sırasında parametre ölçümlerindeki hatalar ve hesaplanan belirsizlikler

Table 1. Errors in parameter measurements and calculated uncertainties during the drying process

Ölçüm Cihazları	Birim	Hata
Fırın sıcaklığı	°C	±0.19
Kronometre	sn	±0.1E-9
Terazi	gr	±0.18
Belirsizlik	Birim	Değer
Sıcaklık ölçümündeki	%	±0.0027
Zaman ölçümündeki	%	±3.3E-11
Kütle ölçümündeki	%	±0.017

2.2. Korelasyon Matrisi

Korelasyon matrisi, iki veya daha fazla değişkenin birbirleri ile karşılıklı olarak ne kadar güçlü ve ne yönde lineer ilişkili olduğunu gösteren istatistiksel bir araçtır. Ön araştırma için pratik bir yoldur. En sık kullanılanları, Spearman ve Pearson korelasyon matrisleridir. Bu matrise bakarak, ilgisiz parametrelerin, daha baştan hesaplamaya dışı bırakılması mümkündür. Her ne kadar günümüzde kullanılan bilgisayarlar ile hızlı hesaplamalar yapılabiliyor olsa da bütün parametrelerin hesaplamada kullanılması çözümü uzatmakta ve daha da önemlisi yakınsama problemlerine neden olabilmektedir. Bu açıdan bakıldığında, daha az parametre ile hesaplama yapılması;

bilgisayar kaynak gereksinimi ve hesaplama sürelerini (maliyet) azaltacaktır. Ayrıca ilgisiz verinin hesaplama dışı bırakılması sayesinde hesaplamada olası bir saptırma etkisinin önüne de geçilebilecektir. *Tablo 2*'de Pearson korelasyon matrisi verilmiştir. Matristeki değerler -1 ve 1 aralığındadır. 1 değeri çok kuvvetli doğrusal ilişkiye işaret ederken -1 değeri ise çok kuvvetli ters ilişkiye işaret etmektedir. 0 ise ilişki olmadığı anlamına gelmektedir. Matrise bakılarak ısıtma hızının kütle ile yüksek doğrusal ilişkili ($R = 0.66$) olduğu, fan hızıyla ters doğrusal ilişkili ($R = -0.23$) olduğu, nem ile ilişkisiz olduğu ($R = 0.05$) sonucu çıkartılabilmektedir. Diğer parametreler için de benzer tarzda çıkarımlar yapabilmek mümkündür.

Tablo 2. Pearson korelasyon matrisi

Table 2. Pearson correlation matrix

Parametreler	Sıcaklık	Kütle	Nem	Fan hızı	Isıtma gücü	Süre	Salınım	Isıtma hızı	Soğutma hızı	Rotasyon	Bekletme
Sıcaklık	1.00	-0.01	0.06	0.01	0.00	-0.09	-0.14	-0.70	-0.47	-0.91	-0.91
Kütle	-0.01	1.00	-0.02	0.02	-0.02	-0.17	0.02	0.66	0.59	0.05	0.05
Nem	0.06	-0.02	1.00	0.02	0.02	0.04	0.01	-0.05	-0.04	-0.05	-0.05
Fan hızı	0.01	0.02	0.02	1.00	-0.05	0.09	0.07	-0.23	-0.65	-0.01	-0.01
Isıtma gücü	0.00	-0.02	0.02	-0.05	1.00	-0.08	-0.04	-0.01	0.01	0.01	0.01
Süre	-0.09	-0.17	0.04	0.09	-0.08	1.00	0.26	-0.07	-0.13	0.04	0.04
Salınım	-0.14	0.02	0.01	0.07	-0.04	0.26	1.00	0.11	0.03	0.07	0.07
Isıtma hızı	-0.70	0.66	-0.05	-0.23	-0.01	-0.07	0.11	1.00	0.88	0.63	0.63
Soğutma hızı	-0.47	0.59	-0.04	-0.65	0.01	-0.13	0.03	0.88	1.00	0.45	0.46
Rotasyon	-0.91	0.05	-0.05	-0.01	0.01	0.04	0.07	0.63	0.45	1.00	1.00
Bekletme	-0.91	0.05	-0.05	-0.01	0.01	0.04	0.07	0.63	0.46	1.00	1.00

1	0.94	0.82	0.7	0.58	-0.56	-0.8	-0.86	-0.92	-0.98	-1
---	------	------	-----	------	-------	------	-------	-------	-------	----

2.3. Modellerin Randımanlarının Ölçülmesi

Modellerin verimliliğini (başarısını) değerlendirmenin en iyi yolu, veri seti ile modelin tahmin değerleri arasındaki farkı karşılaştırmaktır. Böylesi bir karşılaştırma için *hata analizi* yapmak yaygın bir yöntemdir. Eşitlik (2)'de sık kullanılan hata fonksiyonları verilmiştir.

$$MSE = \frac{1}{N} \sum_{t=1}^N |p(t) - o(t)|^2 \quad (\text{Eş. 2.a})$$

$$RMSE = \sqrt{MSE} \quad (\text{Eş. 2.b})$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |p(t) - o(t)| \quad (\text{Eş. 2.c})$$

$$MAPE = \frac{100}{N} \sum_{t=1}^N \left| \frac{p(t) - o(t)}{o(t)} \right| \quad (\text{Eş. 2.d})$$

$$R = \frac{\sum_{t=1}^N (p(t) - \bar{p})(o(t) - \bar{o})}{\sqrt{\sum_{t=1}^N (p(t) - \bar{p})^2} \cdot \sqrt{\sum_{t=1}^N (o(t) - \bar{o})^2}} \quad (\text{Eş. 2.e})$$

$$Hata = p(t) - o(t) \quad (\text{Eş. 2.f})$$

$$Oransal Sapma (\%) = 100 * [o(t) - p(t)] * o(t)^{-1} \quad (\text{Eş. 2.g})$$

$$\sigma = \sqrt{\left(\sum_{t=1}^N [p(t) - \bar{p}]^2 * N^{-1}\right)} \quad (\text{Eş. 2.h})$$

Burada $o(t)$ eldeki veri kümesindeki değerler olup $p(t)$ ise bunların modeller tarafından tahmin edilen karşılıklarıdır. \bar{p} , modelden elde edilen değerlerin ortalaması \bar{o} , deneysel verilerin ortalaması, t değeri zaman adımı, σ ise standart sapma olup verilerin ortalamasına göre yayılımının (sapmanın) istatistiksel bir ölçüsüdür. R ve R^2 sırası ile korelasyon ve determinant katsayılarıdır. MSE; hata karelerinin ortalaması olup birimi, incelenen veri biriminin karesidir. RMSE ise MSE' nin kareköküdür. MAE; mutlak hatanın ortalamasıdır. RMSE ve MAE incelenen veri ile aynı birimdedir. MAPE mutlak hata ortalamasının yüzde değeri olup yüzde birimindedir.

Genel bir kural olarak; 1'e en yakın R (veya R^2) değeri ve sıfıra yakın MSE, RMSE, MAPE, MAE değerleri hedeflenmektedir. Bu değerlerinin sıfıra yakın olması, tahmindeki ortalama hatanın azlığını göstermektedir. R için 0.01-0.29 arasındaki değerler düşük düzeyde ilişki, 0.3-0.7 arasında orta düzeyde ilişki, 0.71-0.99 değeri yüksek düzeyde ilişki anlamına gelmektedir. Negatif değerler, ters bir ilişki anlamına gelmektedir (Köklü ve ark., 2006).

2.4. Zaman serilerinde zaman gecikmesinin tayini

Belli zaman adımlarında toplanan veriye *zaman serisi* (ARIMA, SARIMA, NARX, LSTM vb.) denmektedir. Günlük hava sıcaklıkları, saniyelik kütle ölçümleri vb. örnek olarak verilebilir. Bir veriyi modellemek için hem istatistiksel hem de yapay zekâ yöntemleri mevcuttur. Ancak bu yöntemlerde gecikme değerinin baştan bilinmesi gerekmektedir. Eğer zaman serisinin tahmin ve modellemesinde terim değerlerinin, kendisinin önceki değerlerine olan bağımlılığı ölçülecekse *otokorelasyon* (OKF) ve *kısmi otokorelasyon fonksiyonları* (KOKF), farklı serilerin gecikmelerinden, birbirlerine olan bağımlılığı ölçülecekse *çapraz korelasyon* (ÇK) fonksiyonları kullanılmaktadır. Bunlar bir sürecin rastgeleliğini ve veri düzensizliğini değerlendirmek için iyi birer kriterdir. Sinyal işleme yöntemi de, biçim belirlemek için güçlü bir araçtır. OKF ve KOKF grafiklerini birlikte değerlendirmek daha anlamlıdır. Grafikler dikey eksenin tek tarafında çizdirilebileceği gibi sağ-sol iki tarafında da çizdirilebilmektedir. Bu grafiklere *korelogram* da denmektedir. ÇK korelasyonu Eşitlik (3) ile hesaplanmaktadır. Korelogramlar, R' den farklı olup 1 veya -1 aralığının dışında da değer alabilmektedirler.

$$R(\tau) = \int_{-\infty}^{+\infty} x(t) * y(t + \tau) * dt \quad (\text{Eş. 3})$$

Burada x ve y fonksiyonları, t anındaki zaman serisi hâlindeki veri kümeleridir. τ negatif, sıfır veya pozitif olabilen *zaman gecikmesidir*. $R(\tau)$ ise τ zaman gecikmesinin bir fonksiyonu olan çapraz korelasyondur. x serisi, y serisine eşit olduğunda çapraz korelasyon, otokorelasyon hâline gelmektedir. Periyodik bir sinyal, OKF'da periyodik ve gittikçe azalan davranış göstermektedir.

- Eğer OKF'de azalma ve aynı zamanda KOKF'de ani kesilme varsa bu tür serilerin *otomatik gerileyen* (AR) türünde olduğu anlaşılmalıdır (*Şekil 2-a*). Bu durumda KOKF'deki kesilme sayısı, *gecikmeyi* verecektir.
- Eğer OKF'de ani kesilme ve aynı zamanda KOKF'de azalma varsa bu tür serilerin *hareketli ortalama* (MA) türünde olduğu anlaşılmalıdır (*Şekil 2-b*). Bu durumda OKF'deki kesilme sayısı *gecikmeyi* verecektir.
- Eğer OKF, sadece sıfır gecikme değerinde bir adet pik yapmış ancak geri kalanında beyaz gürültü olarak adlandırılan sıfıra yakın seviyede ise, seriler birbirlerine hiçbir gecikme değerinde bağlı değil, korelasyonsuzdur ve rastgele olarak ifade edilir. Sıfır gecikmedeki pikin varlığı, olması gereken sonuçtur. Böylelikle seriler tam üst üste çakışarak (sıfır gecikme) birbiri ile aynı olmaktadır (*Şekil 2-c*). Eğer iki farklı serinin ÇK grafiğinde hiçbir pik yok ve sadece beyaz gürültü varsa seriler birbirlerine hiçbir gecikme değerinde bağlı değildir ve korelasyonsuzdur. Sıfır gecikmede bile birbirinden farklıdır (*Şekil 2-d*).

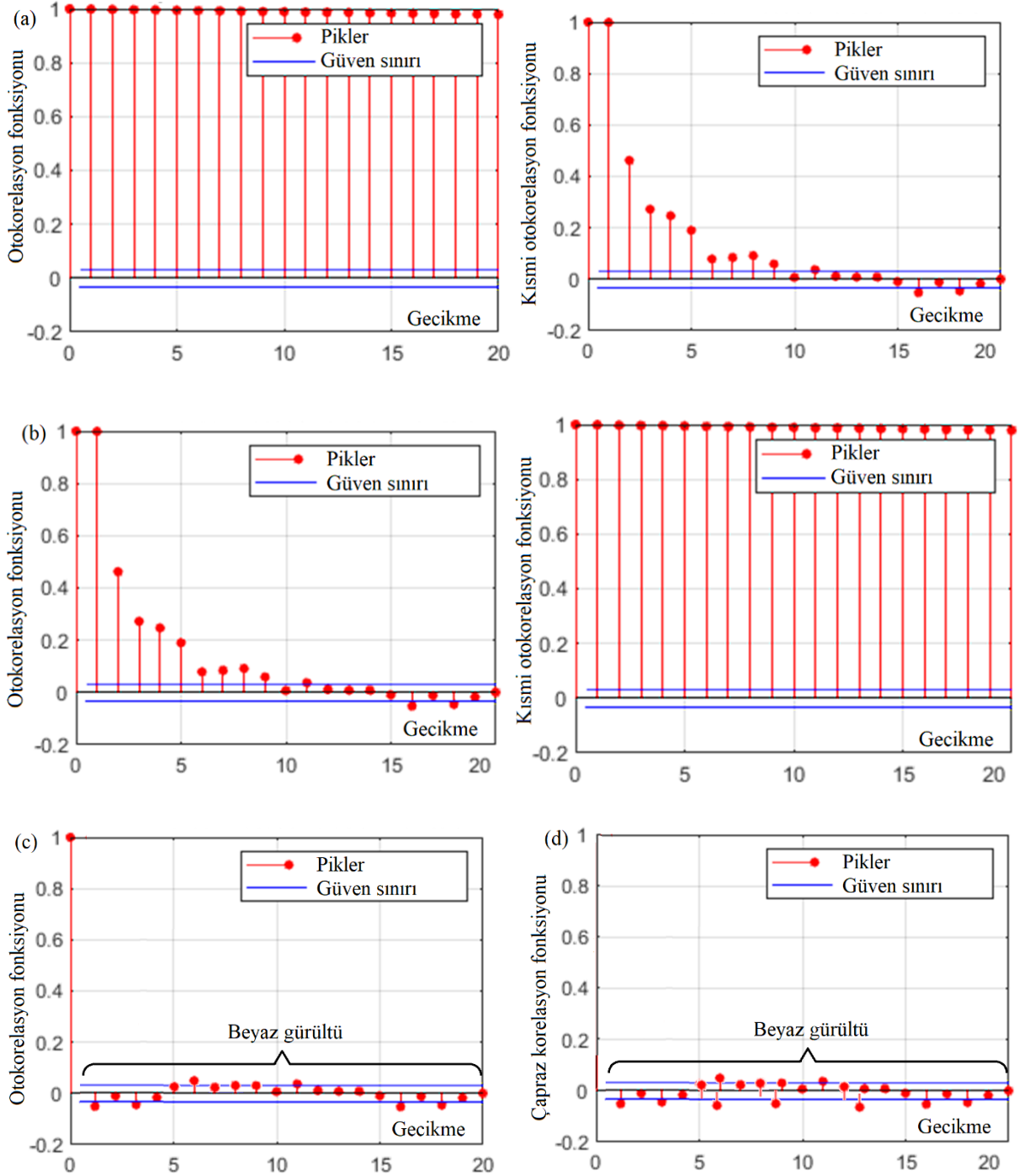


Figure 2. (a) ACF-PACF pattern in AR series (b) ACF-PACF pattern in MA series (c) an ACF pattern without delay, (d) an CC pattern without delay

Şekil 2. (a) AR serilerindeki OKF-KOKF deseni (b) MA serilerindeki OKF-KOKF deseni (c) gecikmesi olmayan bir OKF deseni, (d) gecikmesi olmayan bir ÇK deseni

Kurutma deneyinden elde edilen serinin korelogramları Şekil 2-a, b ile verilmiş olup buna göre veri rastgele değildir. Şekil 2-a'da OKF giderek azalmaktadır. OKF katsayıları sıfırdan önemli ölçüde farklıdır ve yavaş da olsa sıfıra doğru düşmektedir. KOKF'nin 5. gecikmeden sonra kesildiği görülmektedir. Beşten sonrası sıfıra yakındır ve bu nedenle *beyaz gürültü* olarak adlandırılmaktadır. Beyaz gürültü, bu çalışmada, \pm iki standart sapma ($\pm 2\sigma$) kadarlık oluşturulan *güven sınırlarının* hemen hemen içinde kalmaktadır. $\pm 2\sigma$ aralığı, %95.4 *güven seviyesine* denk gelmektedir.

$\pm 2\sigma$ ile gösterilen mavi renkli alt üst sınır çizgilerinin arasında kalan değerler, tüm değerlerin %95.4'lük kısmıdır (Datatab, 2024; The Data Scientist, 2024). Yani tüm verinin %95.4'ü, beyaz gürtlüdür. Bu durum ilk 5 gecikmenin, birbirleri ile yüksek ilişkili ve sonraki ardışık değerlerin ise ilişkili olmadığı anlamına gelmektedir. Bu davranış, serinin ilk beş gecikmeye bağlı ve otomatik gerileyen bir seri AR(5) olduğunu göstermektedir. Şekil 2-b'de ise KOKF giderek azalmakta ve OKF beşinci gecikmeden sonra kesilmektedir. Bu nedenle bu seri MA(5) türündedir. Şekil 2-c örnek olması açısından verilmiş olup verilerin rastgele olduğu iki aynı seriyi, Şekil 2-d ise verilerin rastgele olduğu iki farklı seri hâlini göstermektedir.

2.5. Gizli Katman ve Nöron Sayılarının Tespiti

Makine öğrenme yöntemlerinde giriş-çıkış oluşturmak üzere ağ modeli kullanılmaktadır. Ağ oluşturulurken; katman sayısı, katmanlardaki nöron sayısı, öğrenme hızı, momentum, aktivasyon fonksiyonu, eğitim algoritması, bias, iterasyon sayısı, ağırlıkların uygun seçilmesi önem arz etmektedir. Gizli katmanlarda çok az nöron kullanılması, *yetersiz uydurma* olarak adlandırılan bir duruma neden olurken, çok fazla nöron olması *aşırı uydurma* ile sonuçlanabilmektedir. Bu nedenle en uygun ağ parametrelerinin belirlenmesi gerekmektedir.

Maalesef, henüz her türlü ağ için kullanışlı olacak tek bir formül mevcut değildir. Veri seti ve model ile hesaplanan değerler arasındaki korelasyonun en yüksek veya değerler arasındaki farkın (hatanın) en düşük olacağı ağın, deneme-yanılma yoluyla tespit edildiği görülmektedir (Amini et al., 2021; Beigi and Toriki, 2021). Fakat deneme-yanılma metodu uzun zaman almakta, zaman, enerji ve hesaplama maliyeti randımanı açısından yetersiz kalmaktadır. Ayrıca bir YSA, eğitimi esnasında gerçekleştirilen iterasyonların gereğinden az veya fazla olması da (Öğündür, 2019) ağ hatası oluşturabilmektedir (Erenturk and Erenturk, 2007). YSA eğitimi için kullanılan pek çok algoritma arasında Levenberg-Marquardt algoritması, hız ve kararlılık sağladığı için çokça tercih edilmektedir (Kacar ve Korkmaz, 2022a).

Gizli katmanların sayısı problemin karmaşıklığına bağlıdır. Sinir ağlarındaki daha fazla sayıda gizli katman, daha karmaşık özellikleri temsil etmektedir (Rachmatullah et al., 2020). Sıfır gizli katman, doğrusal ayrılabilir fonksiyonlarda veya kararlarda iyidir. Bir adet gizli katman, bir sonlu uzaydan diğerine sürekli bir eşleme içeren herhangi bir fonksiyon için iyiye, iki gizli katman, rasyonel aktivasyon fonksiyonları sayesinde keyfi bir karar sınırını keyfi doğrulukla temsil edebilmede iyidir. Herhangi bir düzgün eşlemeyi herhangi bir doğrulukla yapabilmektedir (Heaton, 2008).

Gizli katmandaki nöron sayısının belirlenmesinde ise farklı öneriler mevcuttur. Önerilerin esası da deneme-yanılma yoluyla ortaya çıkan belirli formüllerdir. Bunlardan biri Eşitlik (4)'te verilmiştir. Aşırı uydurmayı önlemek için gizli katmandaki nöron sayısı bu eşitlik kullanılarak hesaplanan N_h değerinin altında olmalıdır (Heaton, 2008).

$$N_h \leq N_s * (\alpha * (N_i + N_o))^{-1} \quad (\text{Eş. 4})$$

Burada N_i giriş nöronlarının sayısıdır, N_o çıkış nöronlarının sayısı, N_s eğitim veri setindeki örnek sayısıdır. α , keyfi bir ölçeklendirme çarpanı olup genellikle 2-10 arasında alınmaktadır. Bu çalışmada $N_s = 35741$ ve $N_i = 11$ (sıcaklık, kütle, nem, fan hızı, ısıtma gücü, süre, salınım, ısıtma hızı, soğutma hızı, rotasyon, bekletme), $N_o = 1$ (nem) ve $\alpha = 10$ olup bu durumda gizli katmandaki nöron sayısı $N_h \leq 298$ olmalıdır. İlave olarak gizli katman(lar)daki nöron sayısı, giriş katmanı boyutu ile çıkış katmanı boyutu arasında olmalıdır. Yani $11 > N_h > 1$ olmalıdır. Bir başka yaklaşıma göre gizli katmandaki nöron sayısı, giriş katmanı boyutunun $2/3$ 'ü ile çıkış katmanı boyutunun $2/3$ 'ünün toplamı kadar olmalıdır. Yani $N_h = 8$ olmalıdır. Başka bir yaklaşıma göre gizli nöronların sayısı, giriş katmanı boyutunun iki katından az olmalıdır. Bu durumda $N_h \leq 22$ olmalıdır. Masters'a (1993) göre tek gizli katmanlı bir sinir ağı için gizli nöron sayısı $N_h = \sqrt{N_i * N_o}$ 'tür. Yukarıdaki örnek için $N_h \approx 4$ olacaktır. Bir örüntüye ve öznelik çıkarımına giderek daha fazla yaklaşmak ve hedef sınıfı belirlemek için sonraki katmanlarda gizli nöronların sayısı azalmaya devam etmelidir (Masters, 1993). Bir çalışmada, $N_h = N_i - 1$ ile bir gizli katman formülü kullanmıştır (Sheela and Deepa, 2013). Bu durumda $N_h = 10$ olmalıdır. Bir sinir ağı iki gizli katmana sahipse, birinci katmandaki nöron sayısı $N_{h1} = N_o * r^2$ ve ikinci katmandaki nöron sayısı,

$N_{h2} = N_o * r$ ile hesaplanmaktadır. Burada $r = \sqrt[3]{N_i * N_o^{-1}}$ olmaktadır (Rachmatullah et al., 2020). Bu formüle göre $N_{h1} = 3$ ve $N_{h2} = 3$ olmalıdır. Tamura ve Tateishi (1997) iki gizli katman kullanmış, karşılık gelen gizli katmanın nöron sayısını $N_h = 3 + N_i * 2^{-1}$ formülünü kullanarak hesaplanmışlardır (Heaton, 2008).

Bu formüle göre $N_{h1} = 9$ ve $N_{h2} = 8$ olmalıdır. Tüm bu formüller bir bütün hâlinde değerlendirildiğinde nöron sayısının 22'yi aşmaması ve 3'ten büyük olması gerektiği sonucu çıkmaktadır.

2.5. Normalizasyon

Alt ve üst limiti çok büyük değerlerde olan verilerin alt ve üst limitlerini -1 ve 1 veya 0 ve 1 aralığına getirmeye yarayan işlemlere *normalizasyon* denmektedir. Bu sayede verinin şeklinde, davranışında herhangi bir değişiklik olmadan standart bir aralık içerisine getirilmesi sağlanmaktadır. Bu sayede normalize edilmiş diğer serilerle de kıyaslama imkânı elde edilmektedir. İlâveten, eğer normalize edilmiş veri, ilk hâline göre daha küçük olacağından dolayı hatalar da daha küçülecektir. Bu nedenle istatistiksel işlemde önce verilerin normalize edilmesine sıklıkla rastlanmaktadır. Farklı normalize metotları mevcut olup örneğin; min-max normalizasyon metodu ile veriler 0 ile 1 aralığına getirilmektedir.

2.6. Tanımlayıcı ve Çıkarımsal İstatistik

Tanımlayıcı istatistik, bir veri setinin özelliklerini belirlemeyi amaçlayan bir istatistik türüdür. Bu amaçla, ortalama, mod, medyan, standart sapma, varyans, çarpıklık, aralık gibi pek çok araç kullanılmaktadır. *Çıkarımsal istatistik*, daha ufak bir örneklem grubu kullanarak büyük popülasyonlar hakkında çıkarımlar yapmaya yaramaktadır. Ülkedeki tüm nüfusun seçimlerde kime oy vereceğini, daha az sayıda insana uygulanan anketlerden elde edilen veri sayesinde tahmin etme işlemi bu türe örnek olarak verilebilmektedir. Bu amaçla regresyon analizleri, hipotez testleri, güven aralıkları gibi araçlar kullanılmaktadır.

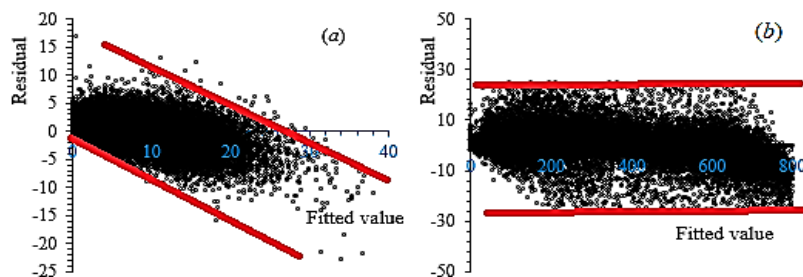
3. Sonuçlar ve Tartışma

Makine öğrenme yoluyla elde edilmiş olan bir modele varyans, regresyon ve hata yakınsama analizi yapılmalıdır. Hesaplamanın teyidi için işlemlerin doğrulanması ve sonuçların geçerliliklerinin ispatı gereklidir. Bunları yapabilmeyi sağlayan çeşitli metrikler ve araçlar mevcuttur. Bu bölümde bunlar izah edilmiştir.

3.1. Varyans

Model tahminine karşın hata grafiğine bakıldığında eğer değerler sabit bir varyansa sahip ise eş varyans (homoscedasticity) olduğu anlaşılabilir ve sifıra yakın rastgele dağılım gözlemlenmektedir. Heteroscedasticity ise değişken varyans olup tahmine karşın hata grafiklerinde dağılım sabit olmayıp koni biçiminde giderek artan veya giderek azalan şeklinde kendisini göstermektedir. Tanımları izah edebilmek üzere öncelikle *Şekil 3*'te bu tür grafikler verilmiş olup bu tür varyanslar gösterilmiştir. *Şekil 3-a* ve *b*'deki gibi iki paralel çizgi arasında kalan sabit ve rasgele dağılım, eş varyansa işaret etmektedir (Kacar ve Korkmaz, 2022a). *Şekil 3-c* ve *d*'deki gibi koni şekilli dağılımlar ise değişken varyans olup giderilmesi gerekmektedir.

Bununla birlikte ham dataya uygulanan Engle's ARCH testi de varyans hakkında bilgi veren bir başka yöntemdir. Ancak ham verilerin değişken varyans içermesi pek muhtemeldir. Bu değişim eğer hata değil ise verinin doğal davranışının bir sonucu da olabilecektir. Ancak bu değişim regresyonlarda eğri uydurmayı zorlaştırdığı için herhangi bir regresyondan önce araştırmacılar, deneysel ham verilerdeki tepe ve çukurları ortadan kaldırmak, veri eğrisini yumuşatmak için sıklıkla bazı ön işlemler uygulamaktadırlar (HosseinTabari et al., 2012; Shen et al., 2020; Taheri et al., 2021). Ancak hangi şiddette uygulanırsa uygulansın, bu şekilde bir ön işlem, verinin doğallığını azaltmakta, olayı temsil yeteneğini düşürmektedir.



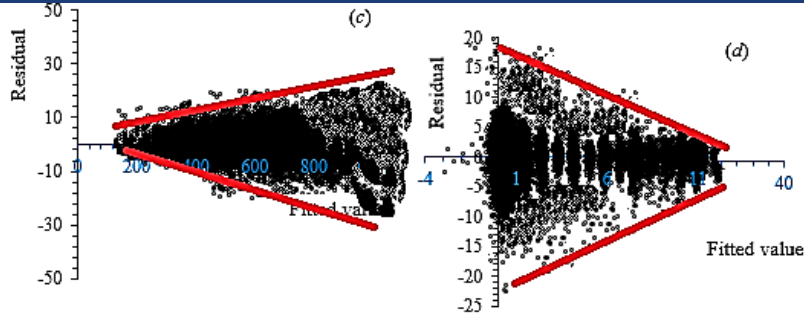


Figure 3. Error values versus the model's forecast; (a) constant, (b) constant, (c) variable, (d) variable variance (Kacar ve Korkmaz, 2022a, 2022b)

Şekil 3. Modelin tahminlerine karşın hata değerleri; (a) sabit, (b) sabit, (c) değişken, (d) değişken varyans (Kacar ve Korkmaz, 2022a, 2022b)

Şekil 4'te ticari gold sülfür gübresinin kurutulması deneyinden elde edilmiş veri gösterilmiştir. Verinin doğallığını bozmadan değişken varyansı gidermek için uygulanan bir strateji; ham veriye logaritma, karekök ve küp kök gibi bazı ön işlemler uygulamaktır. Bu işlemleri yapmadaki temel fikir şudur: Bu fonksiyonlar mevcut veriyi küçülttüğünden dolayı, hatanın da daha küçük olması sağlanmaktadır. Model tahmin yaptıktan sonra da bulunan sonuca, bu işlemlerin tersi uygulanarak mevcut boyuta geri getirilmesi sağlanabilmektedir. Dolayısıyla, bu işlemler veriyi bozmamaktadır.

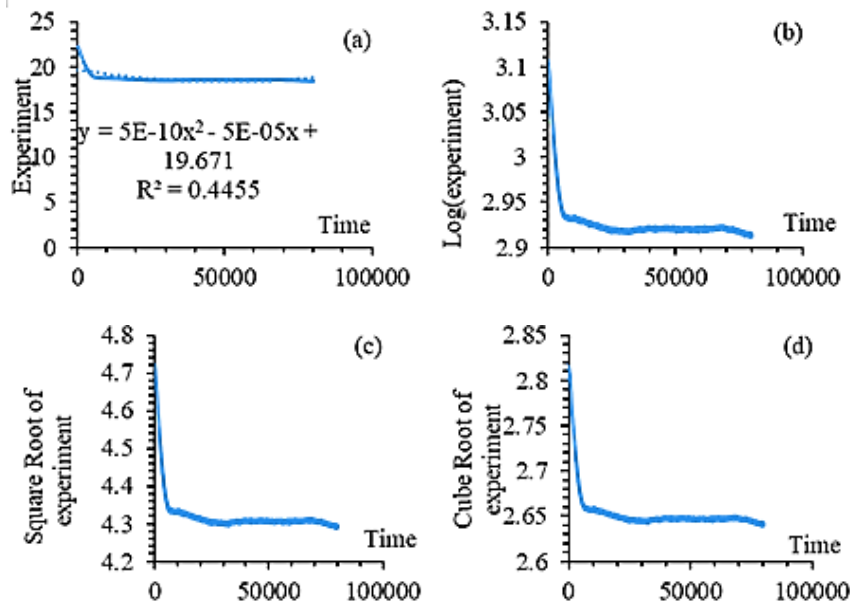


Figure 4. For gold sulfur (a) Experimental data, (b) logarithm of experimental data, (c) square root of experimental data, (d) cube root of experimental data

Şekil 4. Gold sülfür gübresine ait (a) Deneysel veriler, (b) deneysel verilerin logaritması, (c) deneysel verilerin karekökü, (d) deneysel verilerin küp kökü

Varyansı göstermenin bir yolu da modelden elde edilen tahmin değeri ile karşılık gelen hatayı grafik olarak çizmektir. Önceki şekildeki (Şekil 4) gösterilen logaritma, karekök ve küpkök işlemlerinin sonucunda elde edilen modellerin, tahminleri esnasında oluşan varyans Şekil 5'te verilmiştir. Şekillerden görülebileceği gibi her ne kadar dağılımın şekli değişmemişse de hem hata hem de varyans bu işlemler sayesinde azaltılmıştır.

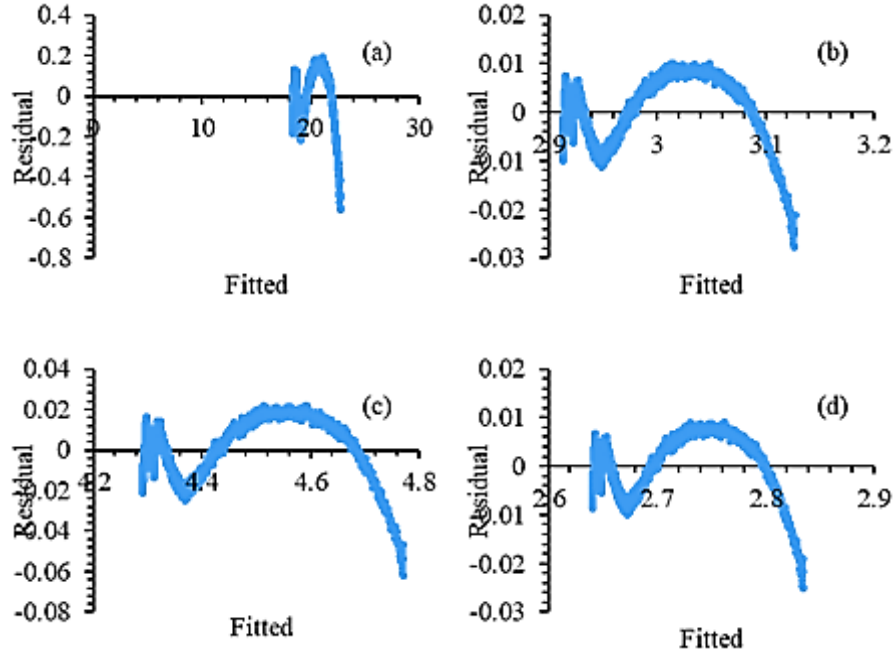


Figure 5. (a) The variance in the experimental data for gold sulfur, (b) the variance over the logarithm of the experimental data, (c) the variance over the square root of the experimental data, (d) the variance in the cubic root of the experimental data

Şekil 5. (a) Gold sülfür gübresine ait deneysel verilerdeki varyans, (b) deneysel verilerin logaritması üzerindeki varyans, (c) deneysel verilerin karekökü üzerindeki varyans, (d) deneysel verilerin kübik kökündeki varyans

3.2. Regresyon Analizi

Modeller için iyilik belirleme araçlarından biridir. Regresyon tıpkı interpolasyon gibi bir eğri uydurma yöntemidir. Kacar and Korkmaz (2022b)'ın kurutma modellemesinde kullandığı ağlardan elde edilen tahminler ve ham veri üzerine yapılan bir regresyon analizi Şekil 6-a'da verilmiştir. Bir ekseninde veri setindeki değerler diğer ekseninde de modelin tahmin ettiği değerler bulunmaktadır. Arzu edilen (idealde olması beklenen) durum *Model tahmini = Ham veri* olmasıdır. Bu durumda $R = 1$ olacaktır ve tüm noktalar tam olarak aynı doğrultu üzerinde olacaktır.

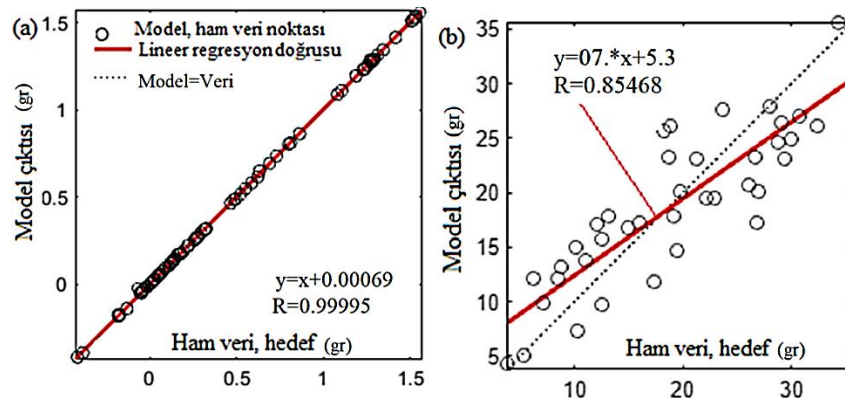


Figure 6. Comparison of model forecast and raw data; (a) $R=0.99994$, (b) $R=0.845555$

Şekil 6. Model tahminleri ile ham verinin kıyaslanması; (a) $R=0.99994$, (b) $R=0.845555$

Hâlbuki gerçekte, model tahminleri, ham veriye çok yakın olmakla birlikte aralarında belli bir ufak fark olabileceği için 45° doğrusunun etrafında ufak saçılmalar gözlemlenebilecektir.

Şekil 6-a'ya bakıldığında lineer regresyon doğrusunun neredeyse 45° doğrusunun tam üzerinde olduğu görülmektedir. Ancak gözle bir karar vermek yerine bu uyumun bir göstergesi olarak *korelasyon katsayısını* kullanmak çok daha güvenilirdir. Şekil 6-a'da $R = 0.99995$ ve (b) ile de $R = 0.845555$ olan iki ayrı model gösterilmiştir. (a)'daki modelin tahmin yeteneği çok iyidir. Lineer regresyon fonksiyonu ile model fonksiyonu birbirleri ile karıştırılmamalıdır! Model fonksiyonları genellikle çok terimli nonlineer fonksiyonlardır. Lineer regresyon fonksiyonu ise modelin tahmini ile karşılık gelen veri arasındaki lineer eğri uydurmazdır.

Ancak R veya R^2 değerleri bir karar vermek için tek başına kesinlikle yeterli değildir! Zira Tablo 3'te verilen değerlerin regresyon doğrusu Şekil 7'de verilmiş olup tablodan görüleceği üzere ham veri ve model tahminleri arasında çok büyük fark mevcut olmasına rağmen Şekil 7'deki regresyon grafiğinden görüleceği üzere $R^2 = 1$ çıkmaktadır! Sadece regresyon grafiğine bakılarak yorum yapılırsa, sanki ham veri ile tahmin değerleri birbirlerine eşitlermiş gibi bir yanlış algı oluşabilmektedir. Bu duruma karşı başka hata ölçütlerini de değerlendirmek önerilmektedir.

Tablo 3. Bir örnek veri seti ve model tahmini

Table 3. An example dataset and model prediction

Ham veri	Tahmin
-3	-6
-2	-4
-1	-2
0	0
1	2
2	4
3	6

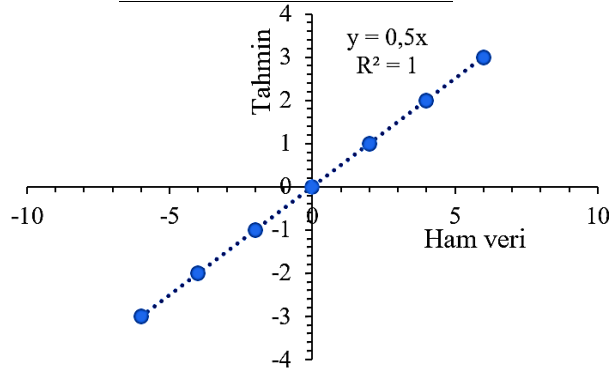


Figure 7. Graph of a regression applied to the data given in Table 3 (Saplıoğlu Akademi, 2020)

Şekil 7. Tablo 3'te verilen veriye uygulanmış bir regresyona ait grafik (Saplıoğlu Academy, 2020)

Modellerin iki boyutlu grafiklerine *doğru*, *eğri*, üç boyutlu grafiklerine ise *yüzey* denmektedir. Gerek doğrular, eğriler olsun gerekse de yüzeyler olsun *ani tepe veya çukurluklar* modeldeki veya hesaplamadaki bir hataya işaret edebilmektedir. Eğer bu durum beklenen bir davranış ise bu davranışın nedeni o şeklin açıklandığı metin içerisinde muhakkak izah edilmelidir. Aksi hâlde hata olarak düşünülebilecektir. Böyle bir eğri ve yüzey Şekil 8'de örnek olarak verilmiştir. Eğrideki veya yüzeydeki beklenmedik, açıklanamayan, doğal olmayan tepe ve çukurlar modelin güvenilirliğini azaltmaktadır.

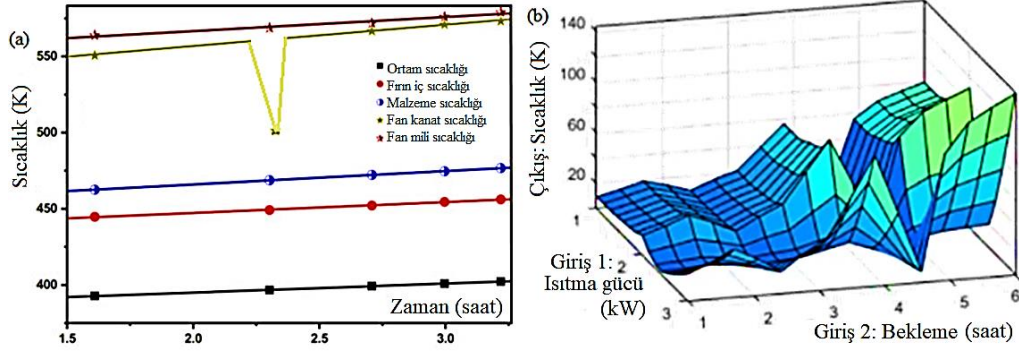


Figure 8. Model; (a) curve, (b) surface

Şekil 8. Model; (a) eğrisi, (b) yüzeyi

Bununla birlikte kurutma deneyi ile elde edilen ham verilerin zamana göre değişiminin grafiği detaylı olarak incelendiğinde, pek çok tepe ve çukurların var olduğu görülmektedir. Sebebi ise kapalı bir fırın içerisinde, kimyasal içeriğe sahip olan gübre tanelerinden sıcak havaya ve havadan gübre tanelerine olan kütle geçişleridir. Difüzyon olarak adlandırılan bu olay bu gibi malzemelerin kurutulmasında görülebilen bazı termo-kimyasal reaksiyonların neticesinde ortaya çıkmaktadır. Bu nedenle beklenen bir davranış olup kurutma işleminin doğal bir sonucudur (Şekil 9). Bu nedenle kullanılacak modelin, genel veri eğrisini yakalayabilmesinin yanı sıra eğrideki tepe ve çukurları da yakalayabilmesi, bu modelin ayırt edici özelliklerinden birisidir.

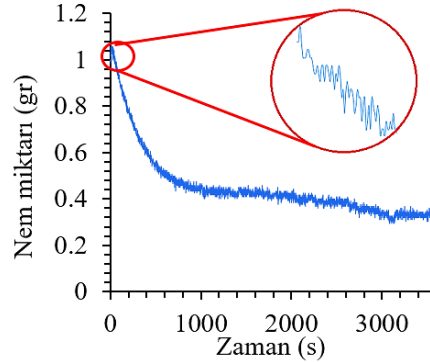


Figure 9. An example of the drying curves

Şekil 9. Kuruma eğrilerine bir örnek

3.3. Hata Yakınsama

Regresyon, makine ve derin öğrenme yöntemleri, simülasyonlar teknik itibarıyla iteratif olarak gerçekleştirilen matematiksel işlemlerdir. İterasyonların amacı, her seferinde istenen değere yaklaşmak, yani hata farkını istenilen değere kadar azaltabilmektir. Bu amaçla *yakınsama* eğrileri olarak adlandırılan grafikler kullanılmaktadır. Yakınsama kelimesi ile kastedilen aslında hatanın sifıra yakınsamasıdır. Hedeflenen ise iterasyonlar sonunda hatanın sifıra tam olarak eşitlenmiş olmasıdır. Ancak sayısal yöntemlerin doğasında yuvarlama hatası, kesme hatası, bilgisayar kaynaklı hatalar başta olmak üzere çeşitli hatalar mevcut olup bunların birikmesi nedeni ile tam değere ulaşmak yerine, iyi bir farkla yaklaşmak mümkündür. Bunun takibi için *hata yakınsama eğrileri* (*kalıntı*, *artık*) kullanılmaktadır. Eğer gittikçe sifıra yaklaşan ve umulmadık yükselişler yapmayan (yani stabil) bir eğri mevcutsa, iterasyonların sonuçlarına güvenileceği anlamı çıkartılabilecektir. Şekil 10'da üç adet hata yakınsama eğrisi verilmiştir. Dikey eksen logaritmiktir. RMSE hatası takip edilmiştir. Bütün modellerin stabilitesini koruduğu söylenebilmektedir. Ancak NARX modelinin, ANFIS ve LSTM'ye göre sifıra daha hızlı yaklaştığı için daha hatasız model olduğu söylenecektir.

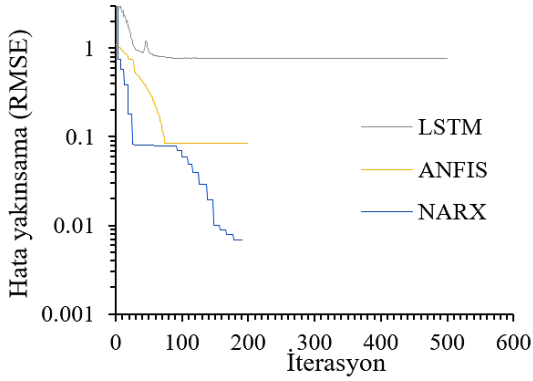


Figure 10. Error convergence curves

Şekil 10. Hata yakınsama eğrileri

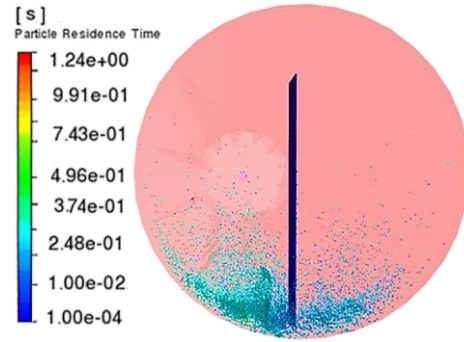


Figure 11. Coupled simulation of the discrete element method and computational fluid mechanics for the fertilizer drying process in a rotary drum dryer

Şekil 11. Döner tamburlu kurutucuda gübre kurutulması sürecinin hesaplamalı akışkanlar mekaniği ile ayrık eleman metodunun birleşik simülasyonu

Şekil 11' de döner tamburlu kurutucuda gübre kurutulması sürecinin hesaplamalı akışkanlar dinamiği ve ayrık eleman metodlarının birleşik simülasyonuna ait bir ekran görüntüsü verilmiştir (Kacar ve Korkmaz, 2021). Şekil 12' de ise bu simülasyon esnasında kaydedilmiş olan hata yakınsama (kalıntı, artık) eğrileri verilmiştir. Ancak bunlar, yakınsamadan emin olmak için tek başlarına yeterli değildir. Monitör oluşturmak, akış yolunu izlemek de gereklidir. Analiz sonunda hız, basınç, sıcaklık, nem, debi vs. gibi değerler elde edilmektedir. Fakat bu değerleri analiz esnasında eş zamanlı olarak takip edebilmek için daha analize başlamadan önce bunların bulunduğu monitör olarak adlandırılan pencerelerin oluşturulması gerekmektedir. Dinamik analizlerde her bir zaman adımında eğrilerin periyodik olması beklenen bir durumdur. Bu durum yakınsamada sıkıntı teşkil etmemekle birlikte her bir zaman adımının içerisinde eğrideki değişim ise düz (zikkaksız) ve kararlı olması gereklidir.

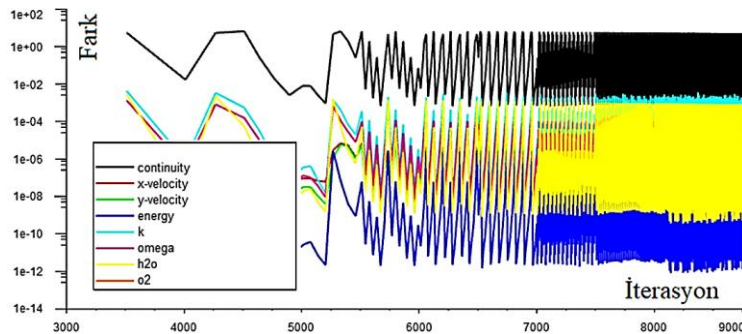


Figure 12. Error convergence(residual) curves of equations during the transient calculation steps

Şekil 12. Transient hesaplama adımları esnasında denklemlerin sifıra yakınsama (artık) grafikleri

Bu grafik sayesinde, hesaplama esnasında süreklilik, hız, türbülans, enerji, kütle transferi vs. denklemlerinin sonuçları takip edilerek her denklemin sonucunun sifıra ne denli yakın olduğu görülebilmektedir. Her bir denklemin sifıra eşit olması gerektiği için, hesaplama neticesinde de elde edilen değer en fazla 1e-3 kadar sifıra yakın olması istenmiştir. Analiz transient (zamanla değişen) olduğundan dolayı birden fazla zaman adımı mevcut olup her zaman adımının içerisinde de en fazla 20 iterasyon yaptırılmasına müsaade edilmiştir. Hesap adımlarının her birinin içerisinde en fazla 12 iterasyon sonunda yakınsama olduğu görülmüş, bu nedenle de 20 iterasyon sınırlandırılmasının yeterli olduğu anlaşılmıştır.

Grafik incelendiğinde, her bir zaman adımından sonra yeni bir hesap yapıldığı görülmüştür. Bu durum *Şekil 12*'de zikzak şeklinde görülmektedir. Zaman adım sayısının birden fazla olması da analizin zamana bağlı tipte yapıldığının bir başka göstergesidir. Her zaman adımı içerisinde (iki tepe arasında) değişim stabildir, zik zak içermemektedir.

3.4. Doğrulama

Doğrulama(lar) sadece veri işleme gerektiren hâllerde değil aynı zamanda da bütün çalışmalarda yapılmalıdır. Herhangi bir işin, analizin vb. doğruluğundan emin olmak için, izlenen yönteminin ve sonucun doğruluğunun ispatına yönelik olarak sağlama, kıyaslama, kanıtlama, teyit ve geçerlilik kontrolü sağlanmalıdır. Bu işlemler, İngilizce dilinde “verification” ve “validation” olarak ayrı ayrı ifade edilmektedir (Kacar, 2023). Türkçe’de bu ikisine “doğrulama” denildiği için aradaki fark, anlatılamamakta ve anlaşılammamaktadır. Farkı belirtmek üzere burada orijinal kelimeler kullanılmak zorunda kalmıştır.

Verification; *çözüm esnasındaki hesaplamaların* hatasız, eksiksiz yapıldığını teyit etmek için yapılan doğrulama işlemleridir. *Şekil 10, 12 ve 13*'te verilen grafikler verification için kullanılan yöntemlere birer örnektir. Daha pek çok farklı yöntemler de mevcuttur. Validation ise *çözüm sonunda bulunan sonuçların* hedeflere, isteklere, standartlara, şartnameye, müşteri talebine uygun olduğunun ispatını, teyidini yapmaktır. Hedef(ler) ise bulunan sonucun en az bir kriteri sağladığının ölçülebilir randıman ve fiziksel değerlerdir. Verification sonunda yanlış olduğu ortaya çıkan bir hesaplama daha baştan yanlıştır, dolayısıyla validation yapmakla vakit kaybetmeye gerek bile yoktur! Validation yapmak için aşağıdaki yöntemler kullanılabilir:

- Mevcut ise analitik formüllerden elde edilen sonuçlarla kıyaslama. Bu en geçerli yöntemdir. Ancak analitik formüller her türlü problem için mevcut olmayabilir!
- Deney yaparak elde edilen deneysel sonuçların mevcut model sonuçları ile kıyaslanması: Bu yöntem de çok güçlüdür ancak pahalı olması nedeni ile deney düzeneği, deneysel veri temini her zaman mümkün olmayabilmektedir.
- Benzer çalışmalarla kıyaslamak: Çalışmanızda ürettiğiniz model, sunduğunuz yöntem vs. ne ise, ona girdi olarak, literatürde bulunan o alandaki çalışmalardaki şartları, girdi değerlerini verdiğinizde, o çalışmalarda sunulan sonuçları aynen sağlamalıdır. Arada fark varsa nedenleri makul olarak açıklanabiliyor olmalıdır. Elbette ki kendi modelinizin, mevcut çalışmalardakilerle kıyaslandığında birebir kopya olmaması için, ilave bir yenilikçi özelliği/özellikleri de olsa daha iyi olur!
 - Varsa standart, tasarım dokümanları, şartname isteklerini sağlamalı.
 - Mevcut ise başka hesaplama yöntemi veya simülasyonlar kullanılarak, sonuçları birbirleri ile kıyaslama.
 - Mevcut tasarımın, referans olarak alınabilecek, aynı amaçla kullanılan başka tasarımlarla mukayesesi.

Sonlu eleman simülasyonlarında verification için yapılan işlemlerden biri yakınsama analizidir (ağdan bağımsızlık analizi). Ağ eleman sayısı değiştirilerek elde edilen sonucun bir değere doğru yakınsamasını gösteren bir grafik olup bir örnek *Şekil 13*'te verilmiştir. Bu grafiğe göre eleman sayısının 1137 adetten fazla olması durumunda incelenen değişkene (nem) ait sonuçta bir değişim olmamaktadır. Bununla birlikte artan eleman sayısı hesaplama süresini artırdığından dolayı *en az eleman* ve *en doğru sonuç* amaçlarına aynı anda uyan denge noktasının tam olarak 1137 eleman olduğu görülmektedir. Başta yapılan bu analiz sayesinde simülasyon esnasında eleman sayısının eksikliğinden kaynaklı bir hatanın olmayacağı ispat edilmiş olmaktadır. Ancak bu tek başına hatasız bir analiz için yeterli değildir!

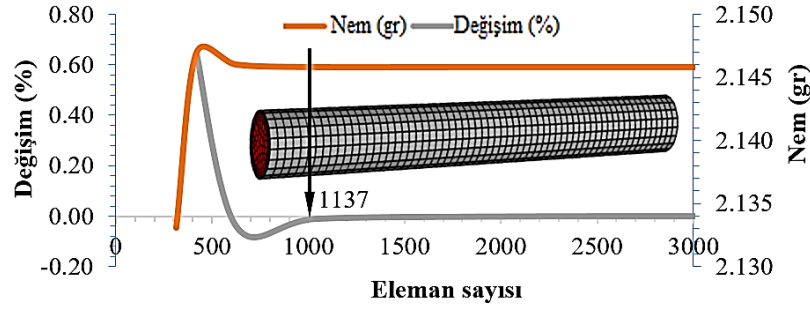


Figure 13. Convergence analysis and graph of percent change in results

Şekil 13. Yakınsama analizi ve sonuçlardaki değişim yüzdesi grafiği

3.5. Tartışma

Virüsün çoğalma-yayılma hızından, gıdaların kurutulmasının tahminine kadar pek çok alanda modelleme/tahmin teknikleri kullanılmaktadır. Esasen istatistiki olan bu yöntemlerin özü, varsa analitik formül kullanmak ya da veriyi modellemektir. Bununla birlikte kurutma alanında çalışan araştırmacılar, regresyon analizleri ile ürünün kuruma davranışını belirlemede literatürdeki çok yaygın matematik modelleri kullanmaktadır (Baran ve ark., 2015; Bayhan, 2021). Mısırın kuruma davranışını en iyi ifade eden matematiksel modelin Page denklemi olduğu ifade edilmiştir (Korkmaz, 2007). Kurutma işleminin doğası gereği nem belirli bir kritik seviyeye doğru (2.azalan bölgede) sürekli azalmaktadır (Kavheci, 1998; Hacıhafızoğlu, 2004). Kurutma deneyi ile elde edilen ham verilerin zamana göre değişiminin grafiği detaylı olarak incelendiğinde, nem miktarındaki azalma esnasında tepe ve çukur şeklinde malzeme davranışı oluştuğu, bunun da kurutma işleminin doğal sonucu olduğu anlaşılmaktadır. Bununla birlikte deneysel kuruma davranışını ifade etmek için kullanılan matematik model sonuçları ile deneysel sonuçlar arasında yüksek doğruluk elde etmek için ön işlemler yapılarak pik ve çukurlukların düzleştirildiği görülmüştür (Evcı ve Ülger, 2006; Aydın, 2016; Kars ve Ekberli, 2021). Zira regresyon modelleri genel eğri şeklini yakalayabilse de tepe ve çukurları istenen hassasiyette temsil edememektedir. Zaman serileri modeli ve yapay sinir ağı yöntemleri ile herhangi bir şekilde verilere ön hazırlık işlemi gerekliliği ortadan kalkmaktadır (Altan et al., 2021). Zira bu yöntemler tepe ve çukurları daha yüksek başarı ile modelleme yeteneğine sahiptir. Bu alanda yapay zekâ ile modellemeler, her geçen gün artmakta ve ham deneysel verilerin modellenmesinde kullanılan geleneksel modellere göre daha yüksek modelleme ve tahmin randımanı elde edilebilmektedir.

4. Sonuçlar

İstatistiksel pek çok veri işleme yöntemi ve değerlendirme metrikleri mevcuttur. Diğer disiplinlerde de ortak olarak kullanılan bu yöntemlere has hangi metriklerin kullanılacağı ve bunların yorumlanması istatistik biliminin haricindeki araştırmacılar, hele hele de genç araştırmacılar tarafından nadiren bilinmektedir. Pek çok metrik arasından modelleme ve tahmin amaçlı kullanılanlara özgü tekniklerin derli toplu bir sunumu bu çalışmada verilmiştir. Kadim bir yöntem olan regresyona kıyasla temsil etme yeteneğinin yüksekliği, veriyi düzeltmeye yönelik ön işlem gereksinimini ortadan kaldırması başta olmak üzere pek çok ilave avantajı bulunan makine öğrenme yöntemleri mevcuttur. Ancak ağırları uygun tasarlayabilmek ve sonuçların doğruluğundan emin olabilmek için bu yöntemlere özgü metrikler olarak regresyon analizi, korelasyon matrisi, oto korelasyon fonksiyonu, kısmi oto korelasyon fonksiyonu, değişen varyans, gizli katman, nöron sayılarının tespiti, hata metrikleri, yakınsama, belirsizlik analizi ve doğrulama teknikleri örneklerle izah edilmiştir.

Teşekkür

Yazarlar, bu çalışmanın inceleme ve değerlendirme aşamasında yapmış oldukları değerli katkılardan dolayı; editör, hakem ve emeği geçenlere içten teşekkür etmektedir.

Etik Kurul Onayı

Bu çalışma için etik kuruldan izin alınmasına gerek yoktur.

Çıkar Çatışması Beyanı

Makale yazarları olarak aramızda herhangi bir çıkar çatışması olmadığını beyan ederiz.

Yazarlık Katkı Beyanı

Bütün yazarlar eşit katkıda bulunmuşlardır.

Kaynakça

- Agrotime (2023). Organomineral toprak düzenleyici; Agrotime Gold Sulphur gübre. <https://www.agrotimegubre.com.tr/-1.html> (Erişim Tarihi: 08.02.2024)
- Altan, A. D., Diken, B. and Kayışoğlu, B. (2021). Prediction of Photovoltaic Panel Power Outputs Using Time Series and Artificial Neural Network Methods. *Tekirdağ Ziraat Fakültesi Dergisi*, 18(3), 457-469.
- Amini, G., Salehi, F., and Rasouli, M. (2021). Drying kinetics of basil seed mucilage in an infrared dryer: Application of GA-ANN and ANFIS for the prediction of drying time and moisture ratio. *Journal of Food Processing and Preservation*, 45(3), e15258. <https://doi.org/10.1111/jfpp.15258>
- Amiroh, K., Rahmawati, D., and Wicaksono, A. Y. (2021). Intelligent System for Fall Prediction Based on Accelerometer and Gyroscope of Fatal Injury in Geriatric. *Jurnal Nasional Teknik Elektro*, 10(3): 155-159. <https://doi.org.10.25077/jnte.v10n3.936.2021>
- Aydın, A. (2016). Precision Feeding in Laying Hens by Sound Technology. *Journal of Agricultural Machinery Science*, 13(3): 134-142.
- Baran, M. F., Dalmiş, İ. S. ve Kayışoğlu, B. (2015). Kanola Bitkisinin Parçalanmaya Yönelik Bazı Mekanik Özelliklerinin Belirlenmesi. *Avrupa Bilim ve Teknoloji Dergisi*, 2(5): 143-148.
- Bayhan, Y. (2021). İkinci ürün ayçiçeği üretiminde farklı toprak işleme ve doğrudan ekim yöntemlerinin enerji kullanım etkinliğinin karşılaştırılması. *Tarım Makinaları Bilimi Dergisi*, 13(2): 102-109.
- Beigi, M. and Torki, M. (2021). Experimental and ANN modeling study on microwave dried onion slices. *Heat and Mass Transfer*, 57: 787–796.
- Çelen, S. (2016). Mikrodalga bantlı kurutucunun gıda kurutmada kullanılabilirliği ve modellenmesi. *Tekirdağ Ziraat Fakültesi Dergisi*, 13(4): 71-83.
- Dash, S. and Venkatasubramanian, V. (2000). Challenges in the industrial applications of fault diagnostic systems. *Computers & Chemical Engineering*, 24(2-7): 785-791.
- Datatab (2024). Confidence Interval [Simply explained]. <https://www.youtube.com/watch?v=ENnISlvQH00> (Accessed Date: 10.02.2024).
- Dong, L., Wang, J., Tseng, M.-L., Yang, Z., Ma, B. and Li, L.-L. (2020). Gyro Motor State Evaluation and Prediction Using the Extended Hidden Markov Model. *Symmetry*, 12(11): 1750, 1-21. <https://doi.org/10.3390/sym12111750>
- Erenturk, S. and Erenturk, K. (2007). Comparison of genetic algorithm and neural network approaches for the drying process of carrot. *Journal of Food Engineering*, 78, 905-912.
- Evcı, G., ve Ülger, P. (2006). Çeltiğin Makineyle Kurutulmasında, Hasat Nemi İle Kurutma Nemi Arasındaki İlişkinin, Maksimum Kırksız Pirinç Randımanı Yönünden Belirlenmesi. *Tarım Makinaları Bilimi Dergisi*, 3(3), 253-258.
- Hacıhafızoğlu, O. (2004). *Çeltiğin kuruma davranışının teorik olarak incelenmesi ve sonuçların elipsoid modele uygulanması*. (Doktora Tezi), Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Makine Mühendisliği Anabilim Dalı, Edirne.
- Heaton, J. (2008). Introduction to Neural Networks with Java. Heaton Research, Incorporated; 2nd edition, Washington.
- HosseinTabari, Kisi, O., Ezani, A., and Talae, P. H. (2012). SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. *Journal of Hydrology*, 444-445: 78-89. doi:10.1016/j.jhydrol.2012.04.007
- Kacar, İ. (2023). Mekanik Tasarım ve Analizin Bilimsel Esasları. Akademisyen Kitabevi, Ankara, Birinci baskı, 36-40.
- Kacar, İ., ve Korkmaz, C. (2021). Döner tamburlu kurutucularda gübre kurutma simülasyonu. *Tarım Makinaları Bilimi Dergisi*, 17(3): 94-100.
- Kacar, İ., ve Korkmaz, C. (2022a). N25P5K5 türü yeni nesil bir bitki besininin kuruma kinetiğinin regresyon ve makine öğrenme yöntemleri ile belirlenmesi, *Çukurova Üniversitesi Mühendislik Fakültesi Dergisi*, 37(4): 989-1004. <https://doi.org/10.21605/cukurovaumfd.1230921>
- Kacar, İ., ve Korkmaz, C. (2022b). Çok katmanlı algılayıcı ağı, uzun kısa süreli bellek ağı ve regresyon yöntemleri ile tarımsal kurutma tahmini *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, 12(4): 1188-1206. <https://doi.org/10.17714/gumusfenbil.1110463>
- Karacabey, E., Aktaş, T., Taşeri, L., ve Seçkin, G. U. (2020). Sultani çekirdeksiz üzüm çeşidinde farklı kurutma yöntemlerinin kurutma kinetiği, enerji tüketimi ve ürün kalitesi açısından incelenmesi. *Tekirdağ Ziraat Fakültesi Dergisi*, 17(1): 53-65.
- Kars, N., ve Ekberli, İ. (2021). Soya bitkisinin verim parametreleri ile bazı kimyasal toprak özellikleri arasındaki pedotransfer modellerin uygulanabilirliği. *Tarım Makinaları Bilimi Dergisi*, 18(3): 494-507.
- Kavheci, K. (1998). *Çeltiğin yığın halinde kuruma davranışı*. (Doktora Tezi) Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Makine Mühendisliği Anabilim Dalı, Edirne.
- Khanlari, A., Güler, H. O., Tuncer, A. D., Sirin, C., Bilge, Y. C., Yılmaz, Y., and Güngör, A. (2020). Experimental and numerical study of the effect of integrating pluss shaped perforated baffles to solar air collector in drying application. *Renewable Energy*, 145: 1677–1692.
- Köklü, N., Büyüktürk, Ş., ve Çokluk-Bökeoğlu, Ö. (2006). *Sosyal Bilimler İçin İstatistik*. Pegem Yayıncılık, Ankara.

- Korkmaz, C. (2007). *Mısırın kuruma davranışının deneysel olarak belirlenmesi ve değişik teori modellerle simülasyonu*. (Yüksek Lisans Tezi), Trakya Üniversitesi Fen Bilimleri Enstitüsü, Makine Mühendisliği Anabilim Dalı, Edirne.
- Masters, T. (1993). *Practical Neural Network Recipes in C++*. Elsevier Inc. : Academic Press.
- Öğündür, G. (2019). Overfitting, underfitting and bias-variance contradiction. <https://www.medium.com> (Erişim Tarihi: 18.04.2023).
- Rachmatullah, M. I. C., Santoso, J. and Surendro, K. (2020). A Novel Approach in Determining Neural Networks Architecture to Classify Data With Large Number of Attributes. *Ieee Access*, 8, 204728-204743. <https://doi.org/10.1109/ACCESS.2020.3036853>
- Saplıoğlu Akademi. (2020). Matlab ile Anfis Uygulaması. <https://www.youtube.com/watch?v=syiqS2o7sKc> (Erişim Tarihi: 20.04.2023).
- Sartori, M. A., and Antsaklis, P. J. (1991). A simple method to derive bounds on the size and to train multilayer neural networks. *IEEE Transactions on Neural Networks*, 2(4): 467-471. <https://doi.org/10.1109/72.88168>
- Sheela, K.G. and S.N. Deepa. (2013). Review on Methods to Fix Number of Hidden Neurons in Neural Networks. *Mathematical Problems in Engineering*, p. 425740.
- Shen, L., Zhu, Y., Liu, C., Wang, W., Liu, H., Kamruzzaman, and Zheng, X. (2020). Modelling of moving drying process and analysis of drying characteristics for germinated brown rice under continuous microwave drying. *Biosystems Engineering*, 195: 64-88.
- Taheri, S., Brodie, G. and Gupta, D. (2021). Optimised ANN and SVR models for online prediction of moisture content and temperature of lentil seeds in a microwave fluidised bed dryer. *Computers and Electronics in Agriculture*, 182: 106003. <https://doi.org/10.1016/j.compag.2021.106003>
- The Data Scientist (2024). Comparing Significance Level, Confidence Level, And Confidence Interval, <https://thedata scientist.com/comparing-significance-level-confidence-level-and-confidence-interval/> (Accessed Date: 10.02.2024).
- Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N. and Tripot, G. (2012). A data-driven failure prognostics method based on mixture of gaussians hidden Markov Models. *IEEE Transactions on Reliability*, 61(2): 491-503. <https://doi.org/10.1109/TR.2012.2194177>
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N., and Yin, K. (2003). A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering*, 27(3): 327-346. [https://doi.org/10.1016/S0098-1354\(02\)00162-X](https://doi.org/10.1016/S0098-1354(02)00162-X)