



Bulanık çıkarım sistemlerinde kullanılan küme sayılarının K-ortalamlar ile belirlenmesi ve baraj hacmi modellenmesi: Kestel barajı örneği

Determination of the number of clusters used in fuzzy inference systems by means of K-means and modeling of dam volume: Kestel dam example

Tülay Suğra KÜÇÜKERDEM¹, Murat KİLİT², Kemal SAPLIOĞLU^{3*}

^{1,3}İnşaat Mühendisliği Bölümü, Mühendislik Fakültesi, Süleyman Demirel Üniversitesi, Isparta, Türkiye.

tulaykucukerdem@sdu.edu.tr, kemalsaplioglu@sdu.edu.tr

²İnşaat Mühendisliği Bölümü, Mühendislik Fakültesi, Afyon Kocatepe Üniversitesi, Afyon, Türkiye.

mkilit@aku.edu.tr

Geliş Tarihi/Received: 05.04.2019, Kabul Tarihi/Accepted: 10.09.2019

* Yazışılan yazar/Corresponding author

doi: 10.5505/pajes.2019.99223

Özel Sayı Makalesi/Special Issue Article

Öz

Giderek azalan su kaynaklarının etkili biçimde kullanılması ve gelecek için su kaynaklarının doğru planlanması önemlidir. Su kaynaklarının planlanması çalışmalarında akım modelleri ve akım tahminleri yapmak çalışmaların temelini oluşturmaktadır. Bu çalışmada Sandıklı Kestel barajına ait 1986-2008 yılı verileri ile ANFIS modeli kullanılarak aylık hacimlerin tahmini yapılmaya çalışılmıştır. Sistemde girdi olarak önceki aylara ait hacimler, hazneye giren ve çıkan hacimler ve buharlaşma miktarı kullanılmıştır. ANFIS yönteminde girdiler için kullanılan küme sayıları ise K-ortalamlar yöntemi ile elde edilmiştir. K-ortalamlar yönteminden elde edilen küme sayıları ile oluşturulan farklı kümeler ANFIS'te modellenmiş ve sonuçlar karşılaştırılmıştır. Her bir girdi değeri için en uygun küme sayıları belirlenmiş ve bu doğrultuda modelleme yapılmıştır. Sonuç olarak uygun küme sayılarına göre yapılan modellerin rastgele oluşturulan modellere göre daha düşük hata yüzdesine sahip sonuçlar verdiği belirlenmiştir.

Anahtar kelimeler: K-ortalamlar, ANFIS, Hidrolojik modelleme

Abstract

Correct planning of water resources is important for the efficient use of rapidly decreasing water resources in the future. Flow modeling and flow estimations in the planning of water resource are the basis of studies. In this study, it is aimed to estimate monthly volumes by using ANFIS model based on the data of 1986-2008 for Sandıklı Kestel dam. In the system, the volume of the previous months, the volume of the incoming and outgoing volumes and the amount of evaporation were used as input variables. In ANFIS method, the number of clusters used for the inputs was obtained by the method of K-means. Different clusters formed by K-averages were modeled in ANFIS and the results were compared. The optimal number of clusters for each input value is determined. Models have been established in this way. As a result, it has been found that the models made according to the optimal number of clusters yield results with lower error percentage compared to randomly generated models.

Keywords: K-means, ANFIS, Hydrological modeling

1 Giriş

İnsan yaşamının her döneminde gerekli olan suyun varlığı ve kalitesi oldukça önemlidir. Bununla beraber özellikle son yıllarda meydana gelen küresel iklim değişikliği ile birlikte bu konu daha da önem kazanmıştır. Kullanılabilir suyun kayda değer bir biçimde azalması nedeniyle önlem alınması gerekliliği kaçınılmaz olmuştur. Bu sebeple mevcut su kaynaklarının potansiyelinin bilinmesi ve daha verimli kullanılmasına yönelik birtakım tedbirler alınması için gerekli olan gösterilmesi gereklidir. Bütün bu sebeplerden ötürü, su kaynakları tasarımında ve işletme çalışmalarında mevcut su potansiyelinin yanı sıra ileriye dönük güvenilir tahminlerin yapılması da çok önemlidir. Bu yüzden elde edilen veriler ile akım modellerinin kurulması veya geleceğe yönelik tahminler yapılması çalışmaların temelini oluşturur [1]-[4].

İleriye yönelik tahminler yapmada model kurulması su kaynakları çalışmaları içerisinde hem ülkemizde hem de dünyada kendisine geniş çapta yer bulmuş bir konudur. Bu konuda ülkemizde ve dünyada bazı çalışmalar yapılmıştır. Okkan ve Dalkılıç, Kemer Barajı havzasına ait aylık akımları tahmin etmede radyal tabanlı sinir ağı modeli geliştirmişlerdir. Yağış, sıcaklık ve bir önceki aydaki yağış değerini girdi olarak alan modelde aylık akımlar tahmin edilmeye çalışılmıştır. Elde ettikleri sonuçlara göre kurulan modelin test ettikleri verilerle

büyük oranda tutarlı sonuçlar verdiğini ifade etmişlerdir [5]. Uysal ve Şorman, Çamlıdere Barajı'na ait aylık verilerinin tahmini için çok katmanlı ağ modeli geliştirmişlerdir [6]. Kızılaslan ve arkadaşları, Sakarya Nehri'nin akım miktarını ileri beslemeli geriye yayımlı yapay sinir ağları kullanarak modellemişlerdir [7]. Okkan ve Mollamahmutoğlu, Yiğitler Çayı için günlük akımları yapay sinir ağları ve çoklu regresyon modeli ile tahmin etmeye çalışmış ve karşılaştırmışlardır. Yapay sinir ağları modelinden elde ettikleri sonuçların çoklu regresyon modelinden elde ettikleri sonuçlara göre daha başarılı olduğunu ifade etmişlerdir [8]. Terzi ve Özcanoğlu, Göksu Nehri'nin akım tahminini genetik programlama ile modellemişlerdir. Modellemede seçilen istasyonun akım verileri ve havzadaki diğer istasyonların akım verilerinden faydalanmışlardır. Sonuç olarak gen ifade programlama yönteminin günlük akım tahmininde başarılı sonuçlar verdiğini belirtmişlerdir [9].

Modelleme çalışmaları incelendiğinde ANFIS yöntemi ile modeller kurmak sıklıkla kullanılan bir konu olmuştur. ANFIS ile uzman bilgisine gereksinim duyulmadan, hızlı bir şekilde güvenilir modellerin kurulabilmesi bu sistemin tercih edilmesine sebep olmaktadır. Anusree ve Varghese, Karuvannur nehri havzasında yapay sinir ağları, ANFIS ve çoklu nonlineer regresyon modeli kullanarak günlük akım verilerini tahmin etmeye çalışmışlar ve ANFIS modelinin diğer modellere

göre daha iyi sonuçlar verdiğini ifade etmişlerdir [10]. Altunkaynak ve Başakın, Columbia Nehri'nin günlük akım verilerini tahmin etmek için sinirsel bulanık çıkarım sistemi (ANFIS), yapay sinir ağları, nonlinear otoregresif model ve otoregresif hareketli ortalama modellerini geliştirmişler ve model sonuçları için karşılaştırma yapmışlardır. Bu karşılaştırmaya göre bulanık mantık yöntemlerinin, özellikle de ANFIS yöntemiyle yapılan modellerin daha iyi sonuç verdiğini belirtmişlerdir [11]. Fırat ve diğerleri, aylık su kullanımını tahmin ettikleri bir çalışmada çeşitli yöntemlerle oluşturdukları modeller içinden en iyi yöntemin uyarlamalı sinirsel bulanık çıkarım sistemi (ANFIS) olduğunu belirlemişlerdir [12]. Yine benzer bir çalışmada; Yurdusev ve Fırat, İzmir ili için su tüketim modellemesi çalışmışlar ve su tüketimini etkileyen parametreleri girdi olarak kullanarak çeşitli yöntemlerle birlikte bir de ANFIS modeli oluşturmuşlardır. Elde ettikleri sonuçlarda İzmir ilinin aylık su tüketimini tahmin etmede ANFIS yönteminin yeterli düzeyde olduğunu belirtmişlerdir [13].

ANFIS ile modelleme yapılırken veriler belirli sayıda üyelik fonksiyonlarına ayrılırlar. Üyelik fonksiyonunun sayısı ve şekli modellemenin başarısını etkilemektedir. Bu çalışmada verilere ait üyelik fonksiyonu sayısı K-ortalamar yöntemi ile kümeleme yapılarak belirlenmiştir. K-ortalamar yönteminden elde edilen küme sayıları ANFIS modelinde üyelik fonksiyonu olarak girilmiş ve karşılaştırılması yapılmıştır. K-ortalamar ile kümeleme de hidrolojik çalışmalar içinde kendisine yer bulmuştur. Ay ve Kişi, çalışmalarında kimyasal oksijen ihtiyacını tahmin etmede K-ortalamar kümeleme yöntemini çok katmanlı çıkarım yapan bir sistemle birleştirerek model kurmuşlar ve modelin kullanılabilirliğini ifade etmişlerdir [14]. Kim ve Parnichkun, su kalitesi parametrelerinin tahmininde ANFIS ve K-ortalamar yöntemini birlikte kullanmışlar ve bu yöntemin sadece ANFIS ve yapay sinir ağları ile kurulan modellerden daha başarılı sonuçlar verdiğini belirtmişlerdir[15].

Bu çalışmada ise Sandıklı Kestel barajının 1986-2008 yılı verileri kullanılarak baraj işletmesi için hacim tahmini yapmada ANFIS yöntemi kullanılmış, ANFIS yönteminde seçilen küme sayıları ise K-ortalamar yöntemi ile belirlenmiştir.

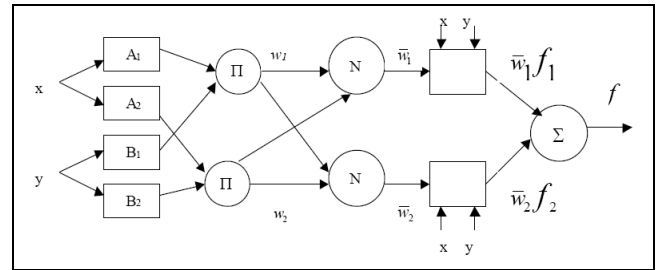
2 Materyal ve yöntem

2.1 Çalışma alanı ve kullanılan veriler

Sandıklı Kızılca barajı Afyon il sınırları içerisinde Kızılca yerleşim merkezinin güneyindeki Kestel Çayı üzerindedir. Kil çekirdekli kum-çakıl dolgu tipindeki barajın gövde hacmi 714290 m³, yüksekliği 59 m, normal su seviyesinde göl alanı 0,532 km²'dir. Çalışma için seçilen baraja ait 1986-2008 yılları işletme verileri ile çalışılmıştır. Çalışmada kullanılan verilere ait bilgiler Tablo 1'de verilmiştir. Tablodaki analiz sonuçlarına göre verilerin modellenebileceği görülmüştür.

2.2 Adaptif sinir ağına dayalı bulanık çıkarım sistemi (ANFIS)

ANFIS (Adaptive Neuro-Fuzzy Inference System) lineer olmayan sistemlerin tahmininde yapay zekâyı kullanarak çözüme ulaşmayı amaçlar. Özellikle tahmin modellerinde çok kullanılan yapay sinir ağlarının öğrenme özelliği ile uzmanlığa dayalı bulanık mantığın çıkarım yapma özelliğinin bir sistemde birleşimi şeklindedir. Yapay sinir ağlarındaki gibi veriye ihtiyaç duyarken aynı zamanda bulanık mantıktaki kümeleme işlemini yapmaktadır. Girdi ve çıktı değerlerinin bilindiği durumlarda ANFIS tüm kuralları kendisi belirleyerek uzman gereksinimini ortadan kaldırmaktadır. Veri sayısının çokluğu ve homojenliği modelin çalışabilmesi açısından son derece önemlidir. Verilerin küme biçimleri ve sayıları problem tipine ve karmaşıklığına göre değişebilmektedir [16]. Şekil 1'de ANFIS mimarisi temsil edilmiştir.



Şekil 1: ANFIS mimarisi.

Tablo 1: Çalışmada kullanılan eğitim ve test verilerine ait istatistiksel analiz.

	EĞİTİM					TEST					
	Ay Başında Rezervuar Hacmi 10 ⁶ m ³	Gelen Hacim 10 ⁶ m ³	İhtiyaç Hacmi 10 ⁶ m ³	Net Buharlaşma mm	Hacim 10 ⁶ m ³	Ay Başında Rezervuar Hacmi 10 ⁶ m ³	Gelen Hacim 10 ⁶ m ³	İhtiyaç Hacmi 10 ⁶ m ³	Net Buharlaşma mm	Hacim 10 ⁶ m ³	
Ortalama	6.291	0.995	0.660	45.34	6.638	Ortalama	5.792	0.754	0.660	44.76	5.901
Standart Hata	0.187	0.096	0.061	4.23	0.219	Standart Hata	0.367	0.134	0.117	7.96	0.389
Ortanca	6.268	0.382	0.065	5	6.331	Ortanca	5.811	0.185	0.065	5	5.814
Standart Sapma	2.742	1.412	0.899	62.27	3.215	Standart Sapma	2.842	1.042	0.905	61.64	3.011
Örnek Varyans	7.520	1.994	0.809	3877.9	1.034	Örnek Varyans	8.075	1.085	0.819	3800.4	9.068
Basıklık	-141.706	1.877	-0.331	-0.92	-0.898	Basıklık	-122.463	5.648	-0.275	-0.92	-100.017
Çarpıklık	0.016	3.514	1.122	0.91	0.352	Çarpıklık	0.047	2.242	114.262	0.91	0.209
Aralık	9.339	11.398	2.490	159.9	1.420	Aralık	9.631	5.163	2.490	156	1.226
Sayı	216	216	216	216	216	Sayı	60	60	60	60	60
Güvenilirlik Düzeyi (95 %)	0.368	0.189	0.121	8.35	0.431	Güvenilirlik Düzeyi (95 %)	0.734	0.269	0.234	15.92	0.778

2.3 K-ortalamlar ile kümeleme

J. MacQueen tarafından 1967'de geliştirilen K-ortalamlar yöntemi bir veri setini belirlenen sayıdaki kümelere bölme işlemidir [17]. Bu yöntemde istenen küme sayısı kadar, kümelerin özelliklerini taşıyan rastgele nesnelere seçilir. Seçilen nesnelere kümelerin merkez noktalarıdır. Veri setindeki elemanların her biri küme merkezlerine olan uzaklıklarına göre bir kümeye dâhil olurlar. Uzaklıkların hesaplanmasında Denklem (1) kullanılır [18],[19].

$$J = \sum_{j=1}^k \sum_{i=1}^M ||Y_i^{(j)} - C_j||^2 \quad (1)$$

Burada Y_i veri değerleri, C_j ise küme merkezinin değeridir.

Daha sonra bütün elemanlar ile birlikte kümelerin merkez noktaları yeniden hesaplanır. Kümelerin merkez noktalarını belirlemek için Denklem (2) kullanılır. Kümelerin merkez noktaları arasındaki fark azalana kadar hesaplamaya devam edilir [20].

$$Z_j = \frac{1}{n_j} \left(\sum_{Y_p \in C_j} Y_p \right) \quad (2)$$

Burada, n herbir kümeye ait eleman sayısı Y_p kümeye ait verilerin değerleridir.

K-ortalamlar yöntemi her bir verinin yalnızca bir kümeye ait olmasına izin verir. Bu yüzden küme ayrımları bellidir [21]. Bu yöntemde iki temel özellik vardır. Bunlardan birincisi birbirine benzeyen elemanların aynı kümede bulunmasının amaçlanmasıdır. Yani aynı kümede bulunan elemanlar arasında benzerliklerin en fazla olması istenir. Diğeri ise oluşturulan kümeler arasındaki benzerliklerin en az olmasının istenmesidir [22].

K-ortalamlar yönteminin değerlendirilmesi için genellikle karesel hata kullanılmaktadır. Kümeleme sonuçlarından en düşük karesel hataya sahip olan en iyi sonuç olarak kabul edilmektedir [23].

3 Bulgular

Su kaynaklarının planlanmasında ve işletilmesinde gelecek tahminleri yapmak oldukça önemlidir. Bu sebeple baraj hacimlerinin tahmin edilmesi de işletme çalışmaları açısından gereklilik arz etmektedir. Bu çalışmada ise baraj hacmini tahmin etmede adaptif ağ tabanlı bulanık çıkarım sistemi kullanılarak modelleme yapılmıştır. Modellemede girdi olarak sırasıyla mevcut hacim (G1), gelen hacim (G2), çıkan hacim (G3) ve buharlaşma verileri (G4) kullanılarak oluşacak hacim miktarı tespit edilmeye çalışılmıştır.

Modellemeye aracı olan ANFIS gibi yapay zeka elemanlarının en büyük sorunu verileri ezberlemesi ve sonuçları bu doğrultuda vermesidir. Verilerin ezberlenmesinden veya öğrenmeden kaynaklı olup olmadığının anlaşılabilmesi için modelleme sırasında mevcut verilerden 1986-2003 yılları arasındaki veriler eğitimde kullanılmış ve 2004-2008 yılları arasındaki veriler de test için ayrılmıştır.

Çalışmada öncelikle mevcut veriler K-ortalamlar yöntemi kullanılarak kümelendirilmiştir. Kümeleme iki şekilde yapılmıştır. Bunlardan birincisi verilerin kendi içinde kümelendirilmesidir. Doğrusal kümeleme olarak isimlendirilmiştir. Diğeri ise girdi

değerlerine karşılık gelen çıktı konumlarının kümelendirilmesi şeklindedir.

K-ortalamlar yöntemi ile kümelemeye ait sonuçlar Tablo 2'de sunulmuştur.

Elde edilen sonuçlara bakıldığında; elemanların homojen dağıldığı kümelerin G1 için dört küme, G2 için beş küme, G3 için beş ve altı küme ve G4 için dört küme olduğu görülmektedir. Tablo 3'ün model bölümünde bu kümeler işaretlenmiştir.

Çalışmada K-ortalamlar ile oluşturulan ANFIS modellerinin yanısıra rastgele oluşturulan kümelemelerle yapılan ANFIS modelleri oluşturulmuştur. Ayrıca süreklilik denklemi ile elde edilen sonuçlar bu modellemelerle kıyaslanmıştır.

Tablo 3'te küme sayıları belirlenen girdi verileri ile kurulan modellere ait sonuçlar verilmiştir. Tablo 3 incelendiğinde, K-ortalamlar ile belirlenen en uygun küme sayıları ile kurulan modellerin, rastgele oluşturulan küme sayılarına göre kurulan modellerden daha iyi sonuçlar verdiği görülmüştür. Şekil 2-5'te en iyi sonuçların elde edildiği modellere ait R² grafikleri verilmiştir. Ayrıca çalışmadaki üç girdili modeller buharlaşma verilerinin girdi olarak baraj hacminin tahmininde kullanılmadığı modellerdir. Tablo 3 incelendiğinde, buharlaşmanın girdi olarak kullanıldığı modellerin hata yüzdelilerinin daha düşük olduğu tespit edilmiştir. Buradan hareketle, bu çalışmaya ek olarak sızma, yüzeyaltı suları vb. verilerin modellere eklenmesi durumunda daha gerçekçi sonuçlar alınabileceği düşünülmektedir. Ancak bu parametrelerin tespitinin zor olması ve yeterli verinin bulunmamasından dolayı bu çalışmada modeller bahsedilen verilerle sınırlandırılmıştır. Ayrıca K-ortalamlar ile oluşturulan ANFIS modellerinin süreklilik denklemi ile elde edilen sonuçlardan da daha iyi sonuçlar verdiği tespit edilmiştir.

Çalışmada elde edilen sonuçlara bakıldığında anfis modellerinin, süreklilik denklemi ile elde edilen sonuçlara çok yakın değerler aldığı hatta zaman zaman da daha iyi olduğu görülmüştür. Çok hızlı ve pratik bir şekilde modellenebilen anfis modellerinin de su hacmi tahmininde kullanılabileceği düşünülmüştür. Ancak klasik olarak anfis modelleri oluşturulurken girdi parametrelerinin alt küme sayıları deneme yanılma yöntemi ile yapılmaktadır. Bu yöntem ise çözüm süresini uzatmaktadır. Bu çalışmada alt küme sayılarının K-ortalamlar sınıflandırma yöntemi ile bulunması düşünülmüş ve girdi parametrelerinin alt küme sayıları bu yöntemle oluşturulmuştur. Alt küme sayıları bu şekilde bulunan modellerin diğer sonuçları biraz daha iyileştirdiği söylenebilir. Örneğin Şekil 2' de 4 er alt küme sayısına ait modelin K-ortalamlar yöntemi ile ikinci ve üçüncü girdi sayılarının 5 e çıkarılması durumu (Şekil 3) incelenmiş ve regresyonu arttırdığı tespit edilmiştir. 3 girdi ile yapılan diğer modellerde de benzer durum olup olmadığı araştırılmıştır. Ancak bu iyileşme sadece eğitim verilerinde gözlemlenmiştir (Şekil4-Şekil 5).

4 Sonuçlar

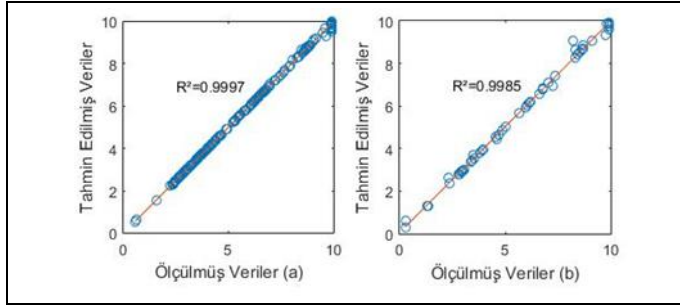
Su kaynakları planlaması ve işletmesi için gelecek tahmini önemlidir. Bu amaçla ANFIS kullanılan bu çalışmada, K-ortalamlar yöntemi ile kümeler belirlenmiştir. Kümeleme yapılırken veriler hem doğrusal kümeleme ile hem de girdi değerlerine karşılık gelen çıktı konumlarının kümelendirilmesi şeklinde gerçekleştirilmiştir.

Tablo 2: K-ortalamlar yöntemi ile yapılan kümelemeye ait bilgiler.

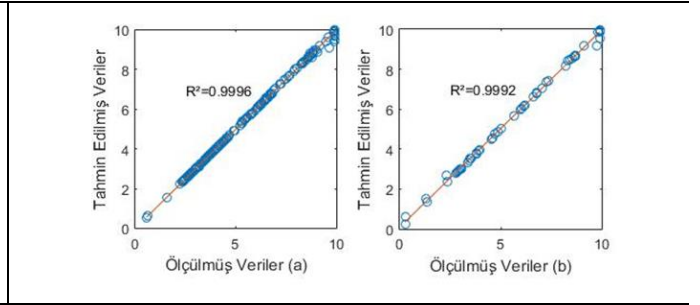
	Küme sayısı	Küme 1		Küme 2		Küme 3		Küme 4		Küme 5		Küme 6	
		Eleman Sayısı	Ortalama	Eleman Sayısı	Ortalama	Eleman Sayısı	Ortalama	Eleman Sayısı	Ortalama	Eleman Sayısı	Ortalama	Eleman Sayısı	Ortalama
G1	2	143	3.89	133	8.64	-	-	-	-	-	-	-	-
	3	96	3.17	89	6.29	91	9.24	-	-	-	-	-	-
	4	84	3.14	58	5.19	57	7.42	77	9.31	-	-	-	-
	5	6	1.06	91	3.32	70	6.38	34	6.63	75	9.67	-	-
	6	26	2.23	67	3.55	35	4.59	36	7.08	43	7.46	69	9.75
G1 _D	2	135	3.79	141	8.68	-	-	-	-	-	-	-	-
	3	112	3.3	67	6.32	97	9.41	-	-	-	-	-	-
	4	52	2.54	65	4.04	64	6.47	95	9.44	-	-	-	-
	5	6	0.76	106	3.44	60	6.18	40	8.37	64	9.84	-	-
	6	22	1.98	67	3.37	36	4.68	52	6.53	38	8.56	61	9.87
G2	2	148	0.41	128	1.55	-	-	-	-	-	-	-	-
	3	129	0.41	119	1.4	28	10.39	-	-	-	-	-	-
	4	112	0.12	73	0.32	68	0.97	23	1.66	-	-	-	-
	5	53	0.16	63	0.47	73	0.86	64	0.96	23	4.18	-	-
	6	7	0.3	105	0.61	36	0.87	60	0.95	50	3.26	18	10.39
G2 _D	2	233	0.49	43	3.35	-	-	-	-	-	-	-	-
	3	218	0.41	56	2.66	2	10.39	-	-	-	-	-	-
	4	180	0.25	68	1.51	26	3.48	2	10.39	-	-	-	-
	5	174	0.23	59	1.27	30	2.53	11	4.31	2	10.39	-	-
	6	163	0.23	55	1.26	29	2.41	18	3.59	9	4.83	2	10.39
G3	2	129	0.37	147	0.91	-	-	-	-	-	-	-	-
	3	96	0.35	113	0.79	67	0.86	-	-	-	-	-	-
	4	68	0.11	96	0.68	63	0.72	49	1.09	-	-	-	-
	5	44	0.05	87	0.34	95	0.35	26	2.31	24	2.33	-	-
	6	39	0.05	64	0.24	63	0.36	39	1.23	44	2.32	27	2.33
G3 _D	2	184	0.07	92	1.83	-	-	-	-	-	-	-	-
	3	184	0.07	46	1.25	46	2.41	-	-	-	-	-	-
	4	161	0.07	23	1.13	46	1.37	46	2.41	-	-	-	-
	5	161	0.07	23	1.13	23	1.37	23	2.29	46	2.53	-	-
	6	161	0.04	23	0.27	23	1.13	23	1.37	23	2.29	23	2.53
G4	2	184	3.06	92	129.9	-	-	-	-	-	-	-	-
	3	184	3.06	46	101.75	46	158.05	-	-	-	-	-	-
	4	138	0	46	12.25	46	101.75	46	158.05	-	-	-	-
	5	138	0	23	10	23	14.5	46	101.75	46	158.05	-	-
	6	138	0	23	99.8	46	12.25	23	103.7	23	156.2	23	159.9
G4 _D	2	184	3.06	92	129.9	-	-	-	-	-	-	-	-
	3	184	3.06	46	101.75	46	158.05	-	-	-	-	-	-
	4	138	0	46	12.25	46	101.75	46	158.05	-	-	-	-
	5	138	0	23	10	23	14.5	46	101.75	46	158.05	-	-
	6	138	0	23	99.8	46	12.25	23	103.7	23	156.2	23	159.9

Tablo 3: Kurulan modellere ait R² ve hata değerleri.

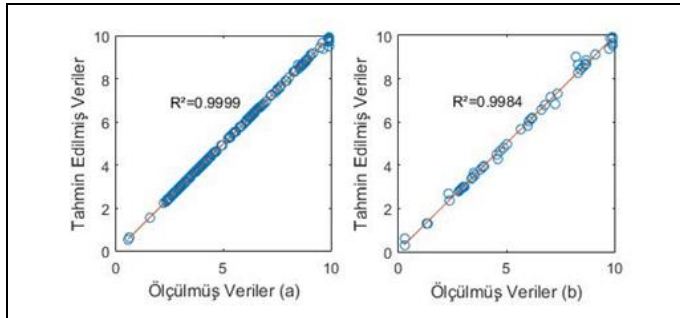
MODEL	EĞİTİM			TEST		
	R ²	Mutlak Hata(%)	Karesel Hata	R ²	Mutlak Hata(%)	Karesel Hata
Süreklilik Modeli	0.9993	0.68	0.07	0.9989	2.24	1.03
2-2-2-2	0.9968	4.66	1.76	0.9957	7.99	1.98
3-3-3-3	0.9995	0.77	0.09	0.9959	2.35	1.18
4-4-4-4	0.9997	0.34	0.04	0.9985	2.27	0.72
5-5-5-5	0.9999	0.22	0.02	0.9949	2.25	1.64
6-6-6-6	0.9999	0.13	0.02	0.9895	2.33	3.66
3-5-5-4	0.9994	1.00	0.12	0.9985	3.17	1.01
4-5-5-4	0.9999	0.25	0.02	0.9984	2.19	0.67
4-5-6-4	0.9999	0.25	0.02	0.9986	2.18	0.67
6-5-5-4	0.9999	0.17	0.01	0.9852	3.11	5.57
2-2-2	0.9963	0.49	1.28	0.9943	11.06	5.44
3-3-3	0.9993	0.94	0.13	0.9986	2.22	0.83
4-4-4	0.9996	0.57	0.07	0.9992	2.03	0.83
5-5-5	0.9999	0.23	0.02	0.9962	2.36	1.30
6-6-6	1	0.14	0.01	0.9751	4.27	5.79
3-5-5	0.9997	0.35	0.05	0.9986	1.92	0.79
4-5-5	0.9997	0.35	0.04	0.9988	1.89	0.78
4-5-6	0.9999	0.22	0.02	0.9975	2.31	1.07
6-5-5	0.9999	0.18	0.01	0.9829	3.72	4.16



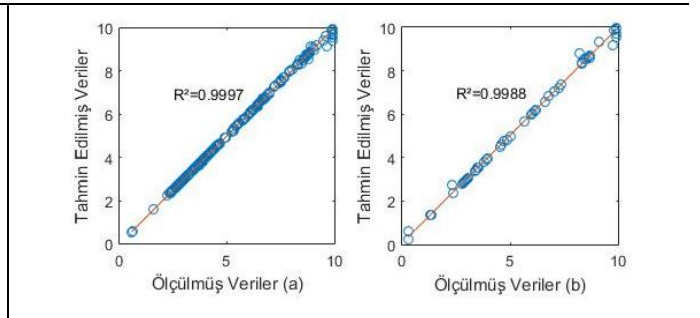
Şekil 2: Dört girdili ve 4-4-4-4 alt kümeye sahip ANFIS modeli.



Şekil 4: Üç girdili ve 4-4-4 alt kümeye sahip ANFIS modeli.



Şekil 3: Dört girdili ve 4-5-5-4 alt kümeye sahip ANFIS modeli.



Şekil 5: Üç girdili ve 4-5-5 alt kümeye sahip ANFIS modeli.

Kümelemenin iki şekilde yapılmasının sebebi girdi verilerinin çıktı verilerini doğrudan veya dolaylı olarak etkilemesinden kaynaklanmaktadır.

Girdi verilerinin çıktı verileri ile dolaylı olarak ilişkili olduğu durumlarda doğrusal kümeleme, doğrudan ilişkili olduğu durumlarda ise girdi verilerinin çıktı verilerine karşılık gelen konumlarının kümelemesi şeklinde yapılmıştır.

Bu çalışmada K-ortalamlar ile kümelemenin, yapay zekâ için gerekliliği araştırılmıştır. Bu çalışmada elde edilen sonuçlara göre ANFIS ile oluşturulacak üyelik fonksiyonu sayısının rastgele yapılması yerine K-ortalamlar ile kümeleme yapıldıktan sonra oluşturulmasının daha iyi sonuçlar verdiği görülmüştür. Yapay zekâ uygulamalarında çok fazla model üretilmesi sağlıklı sonuçların elde edilebilmesi için önemlidir.

Ancak bunun hangi aşamada kesileceği veya kaç tane yapılacağı bilinmemektedir. Ancak K-ortalamlar ile kümeleme yapıldığı takdirde modelleme sayısının minimuma düşürüleceği düşünülmektedir. Bu çalışmada elde edilen bu bulgunun daha sonra yapılacak olan çalışmalarda da denenmesi halinde model sayısı minimuma düşürülebilecektir.

Çalışmada K-ortalamlar ile seçilen küme sayılarından üç girdili modeller için 4-5-5, dört girdili modeller için de 4-5-6-4 küme sayısına sahip modellerin daha iyi sonuçlar verdiği tespit edilmiştir. Ayrıca dört girdili modellerin de üç girdili modellere göre daha iyi sonuç verdiği görülmektedir. Bununla birlikte, K-ortalamlar kümelemesi ile oluşturulan üç girdili modellerin de rastgele kümelenen ANFIS modellerinin büyük bir kısmından daha iyi sonuçlar verdiği gözlemlenmiştir. Ayrıca hem üç girdili hem de dört girdili (K-ortalamlar ile oluşturulmuş) modeller süreklilik denkleminde daha iyi sonuçlar vermiştir.

5 Kaynaklar

- [1] Sattari MT, Salmasi F, Öztürk F. "Sulama amaçlı hazne kapasitesinin belirlenmesinde çeşitli yöntemlerin karşılaştırılması". *Ankara Üniversitesi Ziraat Fakültesi Tarım Bilimleri Dergisi*, 14(1), 1-7, 2008.
- [2] Simonovic SP. "Reservoir system analysis: Closing gap between theory and practice". *Journal of Water Resources Planning and Management*, 118(3), 262-280, 1992.
- [3] Srinivasan K, Neelakantan TR, Narayan PS, Nagarajukumar C. "Mixed-integer programming model for reservoir performance optimization". *Journal of Water Resources Planning and Management*, 125(5), 298-301, 1999.
- [4] Treiber B, Schultz GA. "Comparison of required reservoir storages computed by the Thomas-Fiering model and the Karlsruhe Model Type A and B". *Hydrological Sciences Journal*, 21(1), 177-185, 1976.
- [5] Okkan U, Dalkılıç HY. "Radyal tabanlı yapay sinir ağları ile Kemer Barajı aylık akımlarının modellenmesi". *İMO Teknik Dergi*, 5957-5966, 2012.
- [6] Uysal G, Şorman AÜ. "Monthly streamflow estimation using wavelet-artificial neural network model: A case study on Çamlidere Dam Basin, Turkey". *Procedia Computer Science*, 120, 237-244, 2017.
- [7] Kızılaslan MA, Sağın F, Doğan E, Sönmez O. "Aşağı Sakarya Nehri akımlarının yapay sinir ağları ile tahmin edilmesi". *Sakarya Üniversitesi Fen Bilimleri Dergisi*, 18(2), 99-103, 2014.
- [8] Okkan U, Mollamahmutoğlu A. "Yiğitler Çayı günlük akımlarının yapay sinir ağları ve regresyon analizi ile modellenmesi". *Dumlupınar Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23, 33-48, 2010.
- [9] Terzi Ö, Özcanoğlu O. "Gen ifade programlama ile Göksu Nehri'nin akım tahmini". *Mühendislik Bilimleri ve Tasarım Dergisi*, 5(3), 483-487, 2017.
- [10] Anusree K, Varghese K. "Streamflow prediction of Karuvannur River basin using ANFIS, ANN and MNL models". *Procedia Technology*, 24, 101-108, 2016.
- [11] Altunkaynak A, Başakın EE. "Zaman serileri kullanılarak nehir akım tahmini ve farklı yöntemlerle karşılaştırılması". *Erzincan Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 11(1), 92-101, 2018.
- [12] Fırat M, Yurdusev MA, Mermer M. "Uyarlamalı sinirsel bulanık mantık yaklaşımı ile aylık su tüketiminin tahmini". *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 23(2), 449-457, 2008.
- [13] Yurdusev MA, Fırat M. "Adaptive neuro fuzzy inference system approach for municipal water consumption modeling: An application to Izmir, Turkey". *Journal of Hydrology*, 365, 225-234, 2009.
- [14] Ay M, Kisi O. "Modelling of chemical oxygen demand by using ANNs, ANFIS and K-ortalamlar clustering techniques". *Journal of Hydrology*, 511, 279-289, 2014.
- [15] Kim CM, Parnichkun M. "Prediction of settled water turbidity and optimal coagulant dosage in drinking water treatment plant using a hybrid model of K-ortalamlar clustering and adaptive neuro-fuzzy inference system". *Applied Water Science*, 7, 3885-3902, 2017.
- [16] Jang JSR. "ANFIS: Adaptive-network-based fuzzy inference system". *IEEE Transactions on Systems, Man, and Cybernetics*, 23(3), 665-685, 1993.
- [17] MacQueen J. "Some methods for classification and analysis of multivariate observations". *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281-297, 1967.
- [18] Goyal MK, Gupta V. "Identification of homogenous rainfall regimes in northeast region on India using fuzzy cluster analysis". *Water Resources Management*, 28(13), 4491-4511, 2014.
- [19] Dikbas F, Fırat M, Koc AC, Gungor M. "Defining homogeneous regions for streamflow processes in Turkey Using a K-Means Clustering Method". *Arab J Sci Eng*, 38, 1313-1319, 2013.
- [20] Xu R, Wunsch D. "Survey of clustering algorithms". *IEEE Transactions on Neural Networks*, 16(3), 645-678, 2005.
- [21] Sarıman G. "Veri madenciliğinde kümeleme teknikleri üzerine bir çalışma: K-ortalamlar ve k-Medoids kümeleme algoritmalarının karşılaştırılması". *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(3), 192-202, 2011.
- [22] Dalton L, Balalrin V, Brun M. "Clustering algorithms: on learning, validation, performance, and applications to genomics". *Current Genomics*, 10(6), 430-445, 2009.
- [23] Tan PN, Steinbach M, Kumar V. *Introduction to Data Mining*. Pearson, 2016.