

05 - 06 Kasım 2025  
Ankara

## Jeotermal Enerji Santrali Net Güç Üretim Tahmininde Çeşitli Regresyon Modellerinin Performans Karşılaştırması

Fatih MERCAN

0096



### Özet

Bu çalışmada, Batı Anadolu'da yer alan, üç flaşlı ve binary çevrimli bir jeotermal enerji santralinin ölçüm verileri kullanılarak net güç üretim tahmini yapan çeşitli regresyon analizi yöntemlerinin performansı karşılaştırılmıştır. Regresyon, sistem parametrelerin çıktılara olan etkisini tahmin etmek için kullanılan temel bir istatistik ve makine öğrenimi tekniğidir. Regresyon modellerinin temel amacı, santralin performansını etkileyen jeotermal akışkan sıcaklığı, akışkan debisi ve hava sıcaklığı gibi parametrelerin net elektrik üretimine etkisini istatistiksel olarak tahmin etmektir. Santralden temin edilen 1460 adet günlük ortalama veri, öncelikle Skewness-Kurtosis normallik testine tabi tutulmuş ve aykırı değerler çıkarılarak normallik koşulları sağlanmıştır. Çoklu doğrusal, polinomal, karar ağaçları, rastgele orman ve gradyan artırma regresyon modelleri oluşturulmuştur. Bazı modellerde aşırı uyum tespit edildiğinden, bu modellere 5 ve 10 katlı çapraz doğrulama yöntemleri uygulanmıştır. Hata ölçüm parametreleri olan R<sup>2</sup>, MAE, MSE ve RMSE değerlerine göre, en başarılı yöntemin 10 katlı çapraz doğrulama ve en başarılı modelin ise Gradyan Artırma olduğu belirlenmiştir.

### Yöntem

Çalışmada kullanılan veriler, Kızıldere jeotermal alanında bulunan, 165 MW kurulu güce sahip üç flaşlı ve binary kombine çevrimli jeotermal elektrik santralinden alınmıştır. Çalışma kapsamında, santralden Ekim 2022 – Eylül 2023 tarihleri arasındaki bir yıla ait günlük ortalama 1460 adet veri temin edilmiştir. Bu veri setindeki jeotermal akışkan sıcaklığı, akışkan debisi ve hava sıcaklığı bağımsız değişkenler olarak, net güç ise bağımlı değişken olarak kullanılmıştır. Veri setindeki bağımsız değişkenlerin normal dağılıma uygunluğu, Skewness-Kurtosis testi ile değerlendirilmiştir. Regresyon analizleri, Google Colaboratory ortamında Python programlama dili kullanılarak gerçekleştirilmiştir. Sklearn, Pandas, NumPy, Matplotlib ve Seaborn gibi kütüphaneler veri analizi, model oluşturma ve görselleştirmeler için kullanılmıştır.

Santralden alınan veri setleri, %70 eğitim, %30 test olacak şekilde ikiye ayrılmıştır. GridSearchCV fonksiyonu kullanılarak modeller için en iyi hiperparametre ayarları seçilmiştir. Belirlenen hiperparametrelerle çoklu doğrusal (ÇDR), polinomal (PR), karar ağaçları (KAR), rastgele orman (ROR) ve gradyan artırma (GAR) regresyon modelleri eğitilmiş ve eğitim ile test setleri üzerinde tahmin verileri oluşturulmuştur. Her modelin performansı R<sup>2</sup>, MAE (Hataların Mutlak Değerinin Ortalaması), MSE (Hataların Karelerinin Ortalaması) ve RMSE (Hataların Karelerinin Ortalamasının Karekökü) gibi hata ölçüm parametreleri ile değerlendirilmiş ve aşırı uyum kontrolü sağlanmıştır. Aşırı uyum tespit edilen modellere k-katlı çapraz doğrulama yöntemi uygulanarak aşırı uyum ortadan kaldırılmıştır.

### Bulgular

Modellerde aşırı uyum olup olmadığını değerlendirmek amacıyla hata ölçüm parametreleri eğitim ve test verileri için ayrı ayrı hesaplanmıştır. KAR, ROR ve GAR modellerinde eğitim seti değeri test setinden büyük olduğundan aşırı uyum tespit edilmiştir.

5 ve 10 katlı çapraz doğrulama yöntemi uygulanarak aşırı uyum ortadan kaldırılmıştır. Tablo 1'e göre bütün modellerin hata ölçüm parametreleri karşılaştırıldığında 10 katlı çapraz doğrulama yönteminde; GAR modeli R<sup>2</sup>, MSE ve RMSE parametrelerine göre sırasıyla 0,86, 4,22 ve 2,01 değerleri ile en yüksek skorlara sahipken ROR modelinin ise 1,38 ile en yüksek MAE skoruna sahip olduğu görülmüştür.

Tablo 1. Bütün modellerin hata ölçüm parametreleri

Model	R <sup>2</sup>	MAE	MSE	RMSE
ÇDR	0,71	2,45	9,09	3,01
PR	0,74	2,31	7,93	2,81
KAR5	0,7	1,81	9,98	2,95
KAR10	0,75	1,67	7,66	2,65
ROR5	0,83	1,44	4,91	2,19
ROR10	0,85	1,38	4,29	2,02
GAR5	0,82	1,53	5,25	2,26
GAR10	0,86	1,43	4,22	2,01

### Tartışma

ÇDR ve PR modellerinin performansı, diğer modellere göre düşük kalmıştır. Bu durum, değişkenler arasındaki ilişkinin basit doğrusal veya polinomal bir yapıda olmadığını göstermektedir.

KAR modeli ise doğrusal ilişkinin olmadığı ve büyük veri setlerinde iyi performans gösterdiği için tercih edilmiştir fakat bu çalışmada düşük performans göstermiştir. Bu durumun veri setimizin bu model için yeterli büyüklükte olmadığı ve değişkenlerin daha karmaşık ilişkilere sahip olduğundan kaynaklandığı düşünülmektedir.

En yüksek performansı ağaç tabanlı topluluk (ensemble) yöntemleri olan GAR ve ROR modelleri göstermiştir. Bu modellerin başarısı, santral verisindeki karmaşık ve doğrusal olmayan ilişkileri yakalayabilme yeteneklerinden kaynaklanmaktadır. Özellikle GAR modelinin en başarılı olmasının nedeni, algoritmasının zayıf öğrencileri (karar ağaçları) ardışık olarak eğiterek her adımda bir önceki modelin hatalarını en aza indirmeye odaklanmasıdır. Bu iteratif yaklaşım, veri setindeki karmaşık etkileşimleri daha hassas bir şekilde modelleyerek daha yüksek doğruluk sağlamıştır. Bu çalışmada GAR en yüksek performansı gösterse de, ROR modeli de yüksek başarı göstermiş olup yorumlanması daha basit bir alternatif sunmaktadır.

### Sonuç

- Bu çalışmada jeotermal santral net güç tahmini için literatürdeki klasik regresyon yöntemleri olan çoklu doğrusal regresyon (ÇDR) ve polinomal regresyona (PR) ek olarak karar ağaçları regresyonu (KAR), rastgele orman regresyonu (ROR) ve gradyan artırma regresyonu (GAR) gibi daha gelişmiş yöntemler kullanılmıştır.
- Bu çalışmanın bulguları, jeotermal sistemlerin modellenmesinde makine öğrenmesi yöntemlerinin, özellikle de topluluk algoritmalarının ve derin öğrenmenin, klasik regresyon yöntemlerinden daha iyi performans gösterdiği ortaya koyan literatürdeki benzer çalışmalarla uyumludur.