



Süleyman Demirel Üniversitesi
YEKARUM e-DERGI
(Journal of YEKARUM)



Cilt 9, Sayı 2, 108-127, 2024
E - ISSN: 1309-9388

Makine Öğrenimi Algoritmaları Kullanılarak Bina Enerji Verimliliğinin Doğrusal Regresyon Analizi

Hakan YÜKSEL^{1*}

^{1*} Isparta Uygulamalı Bilimler Üniversitesi, Teknik Bilimler Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Isparta, Türkiye, (ORCID: 0000-0003-2186-533X), hakanyuksel@isparta.edu.tr

(İlk Geliş Tarihi 01/10/2024 ve Kabul Tarihi 18/12/2024)

ÖZET:

İklim değişikliği günümüz dünyasında kritik bir sorun haline gelmiştir. Günümüzde binalar, küresel enerjinin büyük çoğunluğunu kullandığından, bina enerji verimliliği (ısıtma yükü ve soğutma yükü) düzenlemeleri giderek daha katı hale gelmektedir. Binaların, enerji tüketimini anlamak ve bina tasarım aşamasında enerji yükünün iyi bir tahminine sahip olmak ise oldukça önemlidir. Geleneksel enerji tahmin çözümleri zaman alıcı ve maliyetli iken, makine öğrenimi algoritmaları bu duruma alternatif bir yaklaşım sunabilmektedir. Bu çalışmada, binaların ısıtma ve soğutma yüklerini tahmin etmek için bir lineer regresyon modeli kullanılmıştır. Veri modeli için enerji verimliliği veri seti kullanılarak enerji verimliliğinin değerlendirilmesi gerçekleştirilmiştir.

Anahtar Kelimeler: *Makine öğrenmesi, Doğrusal regresyon analizi, Bina verimliliği*

Linear Regression Analysis of Building Energy Efficiency Using Machine Learning Algorithms

ABSTRACT

Climate change has become a critical issue in today's world. Since buildings consume the vast majority of global energy today, building energy efficiency (heating load and cooling load) regulations are becoming increasingly stringent. Understanding the energy consumption of buildings and having a good estimate of the energy load during the building design phase is very important. While traditional energy estimation solutions are time-consuming and costly, machine learning algorithms can offer an alternative approach. In this study, a linear regression model is used to estimate the heating and cooling loads of buildings. The energy efficiency dataset is used for the data model to evaluate energy efficiency.

Keywords: *Machine learning, Linear regression analysis, Building efficiency,*

* Sorumlu yazar E-mail: hakanyuksel@isparta.edu.tr

1. GİRİŞ

Enerji verimliliği, enerji tüketimini düşürme, sera gazı emisyonlarını azaltma ve iklim değişikliğinin etkilerini hafifletme ihtiyacıyla yönlendirilen küresel bir öncelik haline gelmiştir. Günümüzde binalar, toplam küresel enerji kullanımının yaklaşık %40’ını oluşturmakta ve bu da sürdürülebilirlik hedeflerine ulaşmak için inşa edilmiş çevrede enerji verimliliği iyileştirmelerini gerekli kılmaktadır. Binaların enerji verimliliğini artırmak, operasyonel maliyetleri önemli ölçüde azaltabilirken aynı zamanda çevresel faydalar da sağlamaktadır. Bu da onu politikacılar, mühendisler ve şehir plancıları için önemli bir alan haline getirmektedir. [1], [2], [3].

Binalardaki enerji tüketimi aslında oldukça karmaşıktır. Yapısal özellikler (boyut, tasarım, malzemeler), çevre koşulları (yerel iklim, mevsimsel değişiklikler) ve yapım davranışları (kullanım kalıpları, ısıtma/soğutma tercihleri) gibi çok sayıda faktörden etkilenmektedir. Bu faktörler son derece dinamik şekillerde etkileşime girerek enerji tüketiminin yalnızca geleneksel yöntemlerle doğru bir şekilde modellenmesini zorlaştırmaktadır. Tarihsel olarak istatistiksel yaklaşımlar bina enerji tüketimini tahmin etmek için kullanılmış olsa da bu yöntemler genellikle değişkenler arasındaki doğrusal olmayan ilişkileri yakalamada başarısız sonuçlar ortaya koymaktadır. Makine öğreniminin, bina enerji tüketimini tahmin etmek için kullanılması, karmaşık, yüksek boyutlu verileri daha büyük bir hassasiyetle analiz edebilen güçlü bir alternatif olarak görmektedir [4], [5], [6].

Makine öğrenimi (ML) algoritmaları, büyük veri kümelerini kullanarak ve bunlardaki kalıpları ortaya çıkararak binalardaki enerji tahmininin karmaşıklıklarını ele almak için oldukça uygundur. Özellikle ML tabanlı regresyon algoritmaları, bina enerji tüketimini doğru bir şekilde tahmin etme konusunda güçlü bir potansiyel göstermektedir. Enerji kullanımı ve diğer ilgili özellikler hakkındaki geçmiş verileri inceleyerek, regresyon tabanlı ML modelleri önemli tahmin faktörlerini belirleyebilir, bağımlı ve bağımsız değişkenler arasındaki ilişkileri modelleyebilir ve gelecekteki enerji ihtiyaçlarını tahmin edebilmektedir [7]. Bu yaklaşım, bina iyileştirmeleri ve enerji tasarrufu müdahaleleriyle ilgili kararları bilgilendirebilecek daha ayrıntılı ve veri odaklı içgörüler sağlamaktadır [8], [9], [10].

Bina enerji verimliliği için makine öğrenimi yaklaşımlarına odaklanan, temel yöntemleri, modelleri ve bulguları özetleyen karşılaştırmalı bir literatür incelemesi Tablo 1’ de verilmiştir. Bu tablo, her bir çalışmanın makine öğrenimi kullanılarak enerji verimliliğinin anlaşılmasına veya iyileştirilmesine nasıl katkıda bulunduğuna dair yapılandırılmış bir görünüm sağlamaktadır.

Tablo 1. Veri Setinin Özellikleri

Yazar(lar) ve Yıl	Amaç	Makine Öğrenimi Model(ler)	Temel Bulgular	Enerji Verimliliğine Odaklanma
Dong ve diğerleri (2005) [11]	Tropikal binalarda SVM uygulaması	Destek Vektör Makineleri	SVM, belirli iklim bağlamları için etkilidir ve esneklik sunar	ML'nin bölgeye özgü enerji tasarrufu potansiyelini gösterir
González & Zamarrero (2005) [10]	Yapay Sinir Ağları ile saatlik enerjiyi tahmini	Yapay Sinir Ağları	Yapay Sinir Ağları'nın geri bildirim mekanizması gerçek zamanlı verilerle doğruluğu artırır	Binalarda sürekli enerji verimliliği iyileştirmelerini destekler
Ekici & Aksoy (2009) [12]	Yapay Sinir Ağları ile enerji kullanımını tahmini	Yapay Sinir Ağları	Yapay Sinir Ağları, doğrusal olmayan durumları iyi bir şekilde ele alarak tahmin gücünü artırır	Enerji tahmin doğruluğu, verimli enerji tahsisine yardımcı olur
Yu ve diğerleri (2010) [8]	Enerji talebi	Karar Ağaçları	Karar Ağaçları yorumlanabilirlik sunar ancak aşırı uyum riski taşır; özellik önemi açısından yararlıdır	Enerji verimliliği politikalarında karar vermeyi destekler
Zhao & Magoulès (2012) [2]	Bina enerji tüketiminin tahmini	Doğrusal Regresyon, Sinir Ağları	Sinir Ağları karmaşık ortamlarda doğrusal modellerden daha iyi performans gösterir	Enerji israfını azaltmada doğru tahminin rolünü vurgular
Fouquier ve diğerleri (2013) [4]	Enerji performansı modellemesinin gözden geçirilmesi	ML ile entegre simülasyon	ML, geleneksel simülasyonları geliştirir; dinamik faktörleri ele alır	ML'yi optimize edilmiş enerji planlaması için değerli bir araç olarak tanımlar
Liang ve diğerleri (2016) [13]	Doluluk verileriyle temel modelleri geliştirme	Çoklu Regresyon	Doluluk, enerji doğruluğunu önemli ölçüde etkiler; eklenen veriler model doğruluğunu artırır	Doluluk odaklı enerji verimliliği önlemlerinin önemini vurgular
Mocanu ve diğerleri (2016) [14]	Enerji tahmini için derin öğrenmeyi uygulamalar	Derin Öğrenme	DNN'leri karmaşık kullanım modellerini yakalamada daha basit modellerden daha iyi	Daha iyi verimlilik planlaması için enerji tahminlerini iyileştirir

Reyna & Chester (2017) [15]	Kentsel bina stoku büyüme etkilerini inceleme	Simülasyon, İstatistiksel Modeller	performans gösterir	Kentsel büyüme yüksek enerji kullanımını sabitleyebilir; modeller erken müdahale ile verimlilik kazanımları gösterir	Uzun vadeli enerji verimliliği için proaktif tasarıma odaklanır
Wei ve diğerleri (2018) [16]	Veri odaklı enerji tüketim yöntemlerini inceleme	Regresyon, Karar Ağaçları, SVM	Karmaşık ML modelleri (ANN, SVM), yüksek doğruluklu tahmin için daha basit olanlardan daha iyi performans gösterir	Optimize edilmiş verimlilik için enerji sistemlerinin hassas kontrolünü destekler	
Amasyalı & El-Gohary (2018) [9]	Enerji tüketimi için veri odaklı modellerin incelenme	Karar Ağaçları, Regresyon Modelleri	Regresyon modelleri temeldir, ancak karmaşık yöntemler daha fazla faktörü ele alır	Verimli enerji yönetimi stratejileri için bir araç olarak ML’ye odaklanma	
Cuilla ve diğerleri (2018) [17]	Çoklu regresyonla bina enerjisini tahmin etme	Çoklu Doğrusal Regresyon	Regresyon güvenilir bir temel model sağlar; ek özellikler hassasiyeti artırır	Regresyon verimlilik değerlendirmeleri için temel model olarak kullanılır	
Hong ve diğerleri (2018) [1]	Bina enerji simülasyonundaki zorlukları analiz etme	Simülasyon tabanlı ML entegrasyonu	Simülasyon sınırlamalarla karşı karşıyadır; ML dinamik modelleme için umut verici geliştirmeler sunar	Karmaşık binalarda enerji verimliliğini ele almak için ML’nin uyarlanabilirliği	
Yang ve diğerleri (2019) [18]	Bina enerjisi için model tahminli kontrol	Regresyon, Model Tahminli Kontrol (MPC)	ML ile MPC, enerji kullanımını dinamik olarak azaltır	Verimlilik için gerçek zamanlı enerji yönetimini optimize eder	

Bu çalışma, bina enerji tüketimini tahmin etmedeki etkinlikleri açısından birkaç önemli makine öğrenimine ait doğrusal regresyon algoritmasını değerlendirmeyi amaçlamaktadır. Regresyon, giriş ve çıkış değişkenleri arasında doğrusal bir ilişki varsayan, genellikle bir temel model görevi gören temel bir tekniktir. Her ne kadar yorumlanabilir olsa da, performansı karmaşık, doğrusal olmayan veri kümeleri için sınırlı olabilmektedir [9]. Dolayısıyla regresyon analizinde ön plana çıkan değerlendirme ölçümlerinden olan Ortalama Karesel Hata (MSE), Kök Ortalama Karesel Hata (RMSE) ve Ortalama Mutlak Hata (MAE) değerleri çıktı sonuçları

kapsamında detaylı bir şekilde incelenmiştir. Yine doğrusal regresyon modeli, Karar Ağacı, Rastgele Orman, Gradient Boosting, Ada Boost, K-En Yakın Komşu, Destek Vektör Makineleri (SVR) performanslarının karşılaştırılması çıktıları üzerinde gerçekleştirilmiştir.

Bu araştırma, regresyon algoritmalarını inceleyerek bina enerji verimliliği alanına değerli içgörüler sağlamayı amaçlamaktadır. Bulgular, bina sahipleri, mühendisler ve şehir plancıları gibi paydaşların enerji tasarrufu stratejileri, tasarım optimizasyonları ve yenileme planlaması hakkında bilinçli kararlar almalarına yardımcı olmayı amaçlamaktadır. Bu araştırma, nihayetinde azaltılmış bir çevresel ayak izine ve önemli enerji tasarruflarına yol açacak olan enerji verimli binaların inşasını ve bakımını destekleyerek sürdürülebilir kalkınmanın daha geniş hedeflerine de uyum sağlamaktadır.

2. MATERYAL ve METOT

2.1. Veri Seti

Enerji verimliliği veri seti, bina parametrelerinin bir fonksiyonu olarak enerji verimliliğinin değerlendirilmesi için hazırlanmış ve Kaggle platformunda bulunan bir veri seti kümesidir [19]. Binaların ısıtma yükü ve soğutma yükü gereksinimlerini (yani enerji verimliliğini) bina parametrelerinin bir fonksiyonu olarak değerlendirmeyi ele almaktadır. Veri kümesi sekiz öznitelik (özellik) ve iki sonuç (çıkıtı) içermektedir. Amaç, sekiz özelliği kullanarak her iki yanıtı da tahmin etmektir. Veri setine ait özellik ve çıktılar şu şekildedir:

Relative_Compactness (Göreceli_Kompaktlık); binanın hacminin, binayı çevreleyen eşdeğer bir küboidin hacmine oranı olarak tanımlanan binanın görece kompaktlığıdır. Bu değişken 0.62 ile 0.98 arasında değişmektedir.

Surface_Area (Yüzey_Alanı); duvarlar, çatı ve pencereler dahil olmak üzere binanın toplam yüzey alanıdır. Bu değişken 514.5 ile 808.5 m² arasında değişmektedir.

Wall_Area (Duvar_Alanı); binanın duvarlarının toplam alanıdır. Bu değişken 245 ile 416,5 m² arasında değişmektedir.

Roof_Area (Çatı_Alanı); binanın çatısının toplam alanıdır. Bu değişken 110.25 ile 220.5 m² arasında değişmektedir.

Overall_Height (Genel_Yükseklik); binanın yüksekliğidir. Bu değişken 3.5 ile 7.0 metre arasında değişmektedir.

Orientation (Yönlendirme); 2: Kuzey, 3: Doğu, 4: Güney, 5: Batı: Bu, binanın yönüdür. 2, 3, 4 ve 5 değerleri sırasıyla Kuzey, Doğu, Güney ve Batı yönlerini temsil etmektedir.

Glazing_Area (Cam Alanı); binanın toplam cam alanıdır ve kat alanının yüzdesi olarak ifade edilmektedir. Bu değişken dört değerden birini almaktadır: %0, %10, %25 veya %40.

Glazing_Area_Distribution (Cam Alanı Dağılımı); cam alanının bina genelindeki dağılımıdır. 1, 2, 3, 4 ve 5 değerleri, düzgün bir dağılımı ve sırasıyla 2- Kuzey, 3- Doğu, 4- Güney ve 5- Batı dağılımlarını temsil etmektedir.

Heating_Load (Isıtma Yüğü); binanın ısıtma yüküdür ve metrekare başına kilovatsaat olarak ifade edilmektedir. Bu değişken 6.01 ila 43.1 kWh/m² arasında değişmektedir.

Cooling_Load (Soğutma Yüğü); binanın soğutma yüküdür ve metrekare başına kilovatsaat olarak ifade edilmektedir. Bu değişken 10.9 ila 48.03 kWh/m² arasında değişmektedir.

Veri seti, farklı binaların şeklini kullanarak enerji analizi gerçekleştirebilmek için kullanılmaktadır. Binalar, diğer parametrelerin yanı sıra cam alanı, cam alanı dağılımı ve yönlendirme açısından farklılık göstermektedir. Veri seti, iki gerçek değerli yanıtı tahmin etmeyi amaçlayan 768 örnek ve 8 özellikten oluşmaktadır.

2.2. Veri Ön İşleme

Veri seti, 768 kayıt ve 8 özellikten oluşmaktadır. Veri seti eksik ya da hatalı veri kapsamında pandas kütüphanesinin fonksiyonları tarafından gözden geçirilmiştir. Bu kapsamda `duplicate()` ve `isna()` fonksiyonlarıyla veri seti denetlenmiştir.

Yapılan incelemeler neticesinde veri setinde herhangi bir eksik değer ya da tekrar eden bir veriye rastlanılmamıştır. İlgili detaylar Tablo 2’de gösterilmiştir.

Tablo 2. Veri Setinin Özellikleri

Sıra	Özellik & Çıktı	Toplam Kayıt	Veri Tipi
0	relative_compactness	768	float64
1	surface_area	768	float64
2	wall_area	768	float64
3	roof_area	768	float64
4	overall_height	768	float64
5	orientation	768	int64
6	glazing_area	768	float64
7	glazing_area_distribution	768	int64
8	heating_load	768	float64
9	cooling_load	768	float64

2.3. Verilerin İncelenmesi

Enerji verimliliği veri seti’ ne ait ilk 5 kayıt Tablo 3’ te gösterilmektedir. Özellikler ve çıktı sütunlarının Tablo 1’ de verilen veri tipine ait değerler olduğu tabloda açıkça görülmektedir.

Tablo 3. Veri Setinin Değerleri

	Relative Compac tness	Surface _Area	Wall_ Area	Roof_ Area	Overall_ Height	Orient ation	Glazing _Area	Glazing_ Area_Distr ibution	Heating_ Load	Cooling_ Load
0	0.98	514.5	294.0	110.25	7.0	2	0.0	0	15.55	21.33
1	0.98	514.5	294.0	110.25	7.0	3	0.0	0	15.55	21.33
2	0.98	514.5	294.0	110.25	7.0	4	0.0	0	15.55	21.33
3	0.98	514.5	294.0	110.25	7.0	5	0.0	0	15.55	21.33
4	0.90	563.5	318.5	122.50	7.0	2	0.0	0	20.84	28.28

2.4. Lineer Regresyon

Doğrusal regresyon, özellikle girdi özellikleri ile hedef değişken arasındaki ilişki doğrusal olduğunda, öngörücü analiz için kullanılan makine öğrenimindeki (ML) en basit ve en popüler modellerden birisidir. Bu model, bir bağımlı değişken ile bir (basit doğrusal regresyon) veya birden fazla bağımsız değişken (çoklu doğrusal regresyon) arasındaki ilişkiyi, gözlemlenen verilere doğrusal bir denklem uygulayarak tahmin etmektedir [20], [21], [22].

2.4.1. Doğrusal Regresyonun Temel Bileşenleri

Basit bir doğrusal regresyon için, model tipik olarak şu şekilde gösterilir:

$$y=w \cdot x+b \quad (1)$$

y ; y bağımlı değişkendir (hedef).

x ; x bağımsız değişkendir (giriş özelliği).

w ; x için ağırlık veya katsayıdır ve doğrunun eğimini temsil eder.

b ; doğrunun y -kesişimini temsil eden önyargı terimidir.

n ; özelliği olan çoklu doğrusal regresyon için şu hale gelir:

$$y=w_1 \cdot x_1+w_2 \cdot x_2+\dots+w_n \cdot x_n+b \quad (2)$$

Doğrusal regresyonun amacı, tahmin edilen ve gerçek değerler arasındaki farkı (hatayı) en aza indirmektir. Bu fark genellikle, tahmin edilen ve gözlenen değerler arasındaki kare farklarının toplamını en aza indiren en küçük kareler kullanılarak en aza indirilir:

$$\text{Cost} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (3)$$

burada; y_i gerçek değer, \hat{y}_i öngörülen değer ve N veri noktası sayısıdır.

2.5. Değerlendirme Ölçütleri

Ortalama Mutlak Hata (MAE), Ortalama Karesel Hata (MSE) ve Kök Ortalama Karesel Hata (RMSE) genellikle modelin performansını değerlendirmek için kullanılmaktadır. Doğrusal regresyon, ekonomi (giderleri tahmin etme), sağlık (hastalıkların ilerlemesini tahmin etme), çevre bilimi (kirlilik seviyelerini tahmin etme) ve birçok farklı alanlarda uygulanmaktadır.

Makine öğrenimi modelleri, verilerdeki kalıpları anlamayı, tahminler yapmayı, sorulara yanıtlar vermeyi veya gizli kalıpların daha derin bir şekilde anlaşılmasını sağlamayı hedefler. Bu yinelemeli öğrenme süreci, modelin kalıpları edinmesini, yeni verilerle test etmesini, parametreleri ayarlamasını ve tatmin edici bir performans elde edene kadar tekrarlamasını içermektedir. Regresyon sorunları için olmazsa olmaz olan değerlendirme aşaması, kayıp fonksiyonlarını kullanmaktadır.

Kayıp fonksiyonları, modelin tahmin edilen değerlerini gerçek değerlerle karşılaştırır ve X (özellik) ile Y (hedef) arasındaki ilişkiyi haritalamadaki etkinliğini ölçmektedir. Tahmin edilen ve gerçek değerler arasındaki farklılığı gösteren kayıp, model iyileştirmesine rehberlik sunmaktadır. Daha yüksek bir kayıp, daha zayıf performansı belirtir ve optimum eğitim için ayarlamalar olduğu göstermektedir.

2.5.1. Ortalama Mutlak Hata (MAE)

Ortalama Mutlak Hata veya L1 kaybı, en basit ve kolay anlaşılır kayıp fonksiyonlarından ve değerlendirme ölçütlerinden biri olarak öne çıkmaktadır. Veri kümesi boyunca tahmin edilen ve gerçek değerler arasındaki mutlak farkların ortalamasını alarak hesaplanmaktadır. Matematiksel olarak, yönden bağımsız olarak yalnızca büyüklüklerine odaklanarak mutlak hataların aritmetik ortalamasını temsil etmektedir. Daha düşük bir MAE, üstün model doğruluğuna işaret eder.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4)$$

y_i = gerçek değer

\hat{y}_i = tahmin edilen değer

n = örneklem büyüklüğünü ifade etmektedir.

2.5.2. Ortalama Karesel Hata (MSE)

MSE, en yaygın regresyon kayıp fonksiyonlarından biri ve önemli bir hata metriğidir. L2 kaybı olarak da bilinen Ortalama Karesel Hata’da, tahmini değer ile gerçek değer arasındaki farkı kareleyerek ve veri kümesi boyunca ortalamasını alarak hata hesaplanmaktadır.

MSE, cezanın hataya değil, hatanın karesine orantılı olması nedeniyle Kuadratik kayıp olarak da bilinmektedir. Hatanın karesi, aykırı değerlere daha yüksek ağırlık verir ve bu da küçük hatalar için düzgün bir eğimle sonuçlanır.

Optimizasyon algoritmaları, en küçük kareler yöntemini kullanarak parametreler için optimum değerleri bulmaya yardımcı olduğu için büyük hatalar için bu cezalandırmadan faydalanır. Hatalar karelendiği için MSE asla negatif olmamaktadır. Hata değeri sıfırdan sonsuza kadar değişmektedir. MSE, hata arttıkça üstel olarak artar. İyi bir modelin MSE değeri sıfıra yakın olacaktır ve bu da verilere daha iyi uyum sağladığını gösterir.

$$MAE = \frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

2.5.3. Kök Ortalama Karesel Hata (RMSE)

Makine Öğrenmesinde Kök Ortalama Kare Hatası (RMSE), makine öğrenmesi ve istatistikte bir tahmin modelinin doğruluğunu ölçmek için kullanılan popüler bir ölçümdür. Tahmin edilen değerler ile gerçek değerler arasındaki farkları niceliksel olarak belirler, hataları kareler, ortalamayı alır ve ardından karekökünü bulur. RMSE, modelin performansı hakkında net bir anlayış sağlar ve daha düşük değerler, kök ortalama kare hatasına göre daha iyi tahmin doğruluğu göstermektedir.

MSE’nin karekökü alınarak hesaplanır. RMSE, Kök Ortalama Kare Sapması olarak da adlandırılır. Hataların ortalama büyüklüğünü ölçer ve gerçek değerden sapmalarla ilgilenir. Sıfır olan RMSE değeri, modelin mükemmel bir uyuma sahip olduğunu gösterir. RMSE ne kadar düşükse, model ve tahminleri o kadar iyidir. Makine öğreniminde daha yüksek bir göreceli kök ortalama kare hatası, kalıntıdan temel gerçeğe büyük bir sapma olduğunu gösterir. RMSE, özelliğın modelin tahminini iyileştirip iyileştirmediğini anlamaya yardımcı olduğu için farklı özelliklerle kullanılabilir.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

3. ARAŞTIRMA SONUÇLARI

Enerji verimliliği veri seti’ nin temel istatistik değerleri Tablo 4’ te verilmiştir. Veriler ait histogram grafikleri ise Şekil 1’ de gösterilmiştir.



Şekil 1. Histogram Grafikleri

Tablo 4. Veri Setine Ait Temel İstatistik Değerler

	RelativeCompactness	Surface_Area	Wall_Area	Roof_Area	Overall_Height	Orientation	Glazing_Area	Glazing_Area_Distribution	Heating_Load	Cooling_Load
count	768.0000	768.0000	768.0000	768.0000	768.0000	768.0000	768.0000	768.0000	768.0000	768.0000
mean	0.764167	671.7083	318.5000	176.604167	5.25000	3.50000	0.234375	2.81250	22.307201	24.587760
std	0.105777	88.086116	43.626481	45.165950	1.75114	1.118763	0.133221	1.55096	10.090196	9.513306
min	0.620000	514.5000	245.0000	110.25000	3.50000	2.00000	0.000000	0.00000	6.010000	10.900000
25%	0.682500	606.3750	294.0000	140.87500	3.50000	2.75000	0.100000	1.75000	12.992500	15.620000
50%	0.750000	673.7500	318.5000	183.75000	5.25000	3.50000	0.250000	3.00000	18.950000	22.080000
75%	0.830000	741.1250	343.0000	220.50000	7.00000	4.25000	0.400000	4.00000	31.667500	33.132500
max	0.980000	808.5000	416.5000	220.50000	7.00000	5.00000	0.400000	5.00000	43.100000	48.030000

Relative_Compactness; Binanın genel genişliği, uzunluğu ve yüksekliğinin çarpımının alınarak binanın toplam hacmine bölünmesiyle hesaplanan boyutsuz bir niceliktir. Bu sütun 0,62 ile 0,98 arasında değişir ve ortalaması 0.764, standart sapması 0.106'dır. Değer ne kadar düşükse bina o kadar kompakttır, daha yüksek değerler ise daha az kompakt bir binayı gösterir.

Surface_Area; Çatı, duvarlar ve pencereler dahil binanın yüzey alanını metrekare olarak gösteren sayısal veridir. Ortalama değeri 671.71 metrekare, standart sapması 88.09 metrekaredir. Minimum değeri 514.50 metrekare, maksimum değeri ise 808.50 metrekaredir. 25., 50. (ortanca) ve 75. yüzdeler sırasıyla 606.38, 673.75 ve 741.13 metrekaredir.

Wall_Area; Her binanın toplam duvar alanını metrekare olarak temsil eder. 245 ila 416.5 arasında değişen değerlere sahip, ortalama değeri yaklaşık 318.5 metrekare ve standart sapması yaklaşık 43.6 metrekare olan sürekli bir sayısal bir parametredir. Minimum değer 245 maksimum değer ise 416.5'tir. 25. yüzdeler değeri yaklaşık 294 iken 75. yüzdeler değeri yaklaşık 343'tür. Duvarlar, çatılar ve pencereler dahil olmak üzere binanın toplam yüzey alanını ölçen "Surface_Area" sütunuyla karşılaştırıldığında, "Wall_Area" sütunu yalnızca duvarların alanını ölçer. Bu nedenle, daha büyük toplam yüzey alanına sahip binalar daha büyük duvar alanlarına

sahip olabilir, ancak bu her zaman böyle değildir, çünkü bu binanın belirli tasarımına ve yönüne bağlıdır.

Roof_Area; binanın toplam çatı alanını metrekare olarak ifade eder. Bu sütundaki minimum değer 110.25 metrekaredir. Bu sütundaki maksimum değer 220.5 metrekaredir. Bu özelliğin ortalama değeri 176.604167 metrekaredir. Bu sütunun standart sapması 45.165950 metrekaredir. Bu özelliğin medyan değeri 183.75 metrekaredir.

Overall_Height; Binanın yüksekliğini temsil eder. Metre olarak ölçülür ve minimum 3.5 metreden maksimum 7 metreye kadar değişir. Veri setindeki binaların ortalama yüksekliği 5.25 metredir ve standart sapması 1.751 metredir. Binaların yükseklik dağılımı simetrik, çünkü ortalama ve medyan (aynı zamanda 5.25 metredir) eşittir.

Orientation; Binanın yönünü temsil eder. 2 ile 5 arasında tam sayı değerleri alır, burada: 2: Kuzeye bakan bina 3: Doğuya bakan bina 4: Güneye bakan bina 5: Batıya bakan bina. Bir binanın yönü, aldığı güneş radyasyonu miktarını etkileyebilir ve bu da ısıtma ve soğutma yüklerini etkileyebilir.

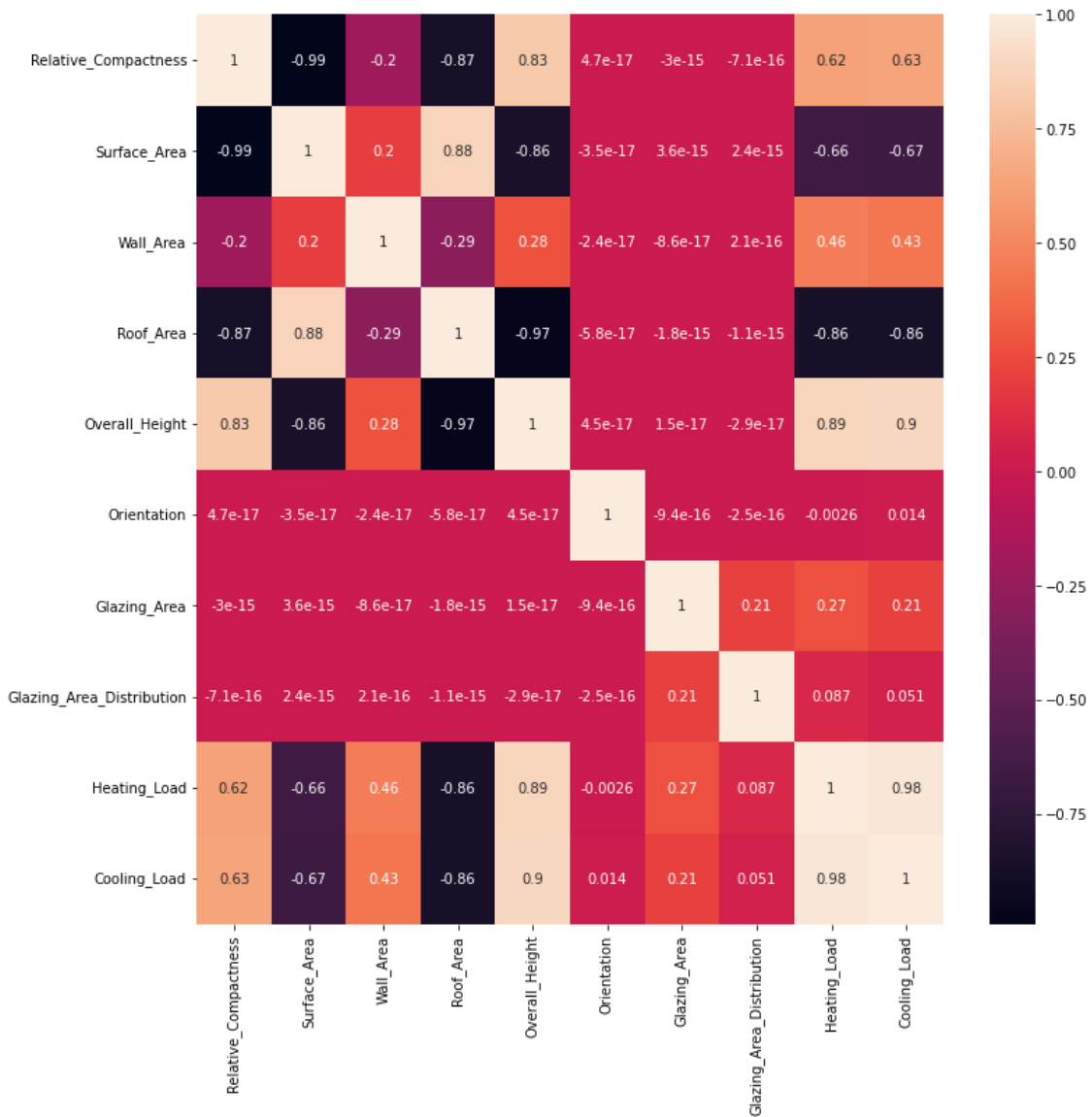
Glazing_Area; Toplam cam alanını binanın taban alanına oranı olarak temsil eder. 0.0 ile 0.4 arasında değişir, ortalaması 0.234 ve standart sapması 0.133'tür. Minimum değer 0.0'dır, yani binanın penceresi yoktur; maksimum değer ise 0.4'tür, yani taban alanının %40'ının pencerelerle kaplı olduğunu gösterir. Genel olarak, daha yüksek bir cam alanı daha fazla doğal ışık ve havalandırma sağlar ancak aynı zamanda daha yüksek ısıtma ve soğutma yüklerine de yol açabilir. Bu nedenle, optimum cam alanı iklim, bina yönü ve istenen enerji verimliliği seviyesi gibi çeşitli faktörlere bağlıdır.

Glazing_Area_Distribution; Binadaki cam alanının dağılımını temsil eden kategorik bir değişkendir. 0 ile 5 arasında bir değer aralığına sahiptir, burada: 0: cam olmadığını temsil eder 1: yalnızca kuzey tarafında cam olduğunu temsil eder 2: kuzey ve güney taraflarında cam olduğunu temsil eder 3: kuzey, güney ve doğu taraflarında cam olduğunu temsil eder 4: kuzey, güney, doğu ve batı taraflarında cam olduğunu temsil eder 5: doğu ve batı taraflarında cam olduğunu temsil eder. Bu sütun kategorik olduğundan, onu tanımlamak için ortalama veya standart sapma gibi ölçütler kullanılmamaktadır. Bunun yerine, her bir değer sıklığına bakarak her bir dağılımın ne kadar yaygın olduğuna dair bir fikir edinebiliriz.

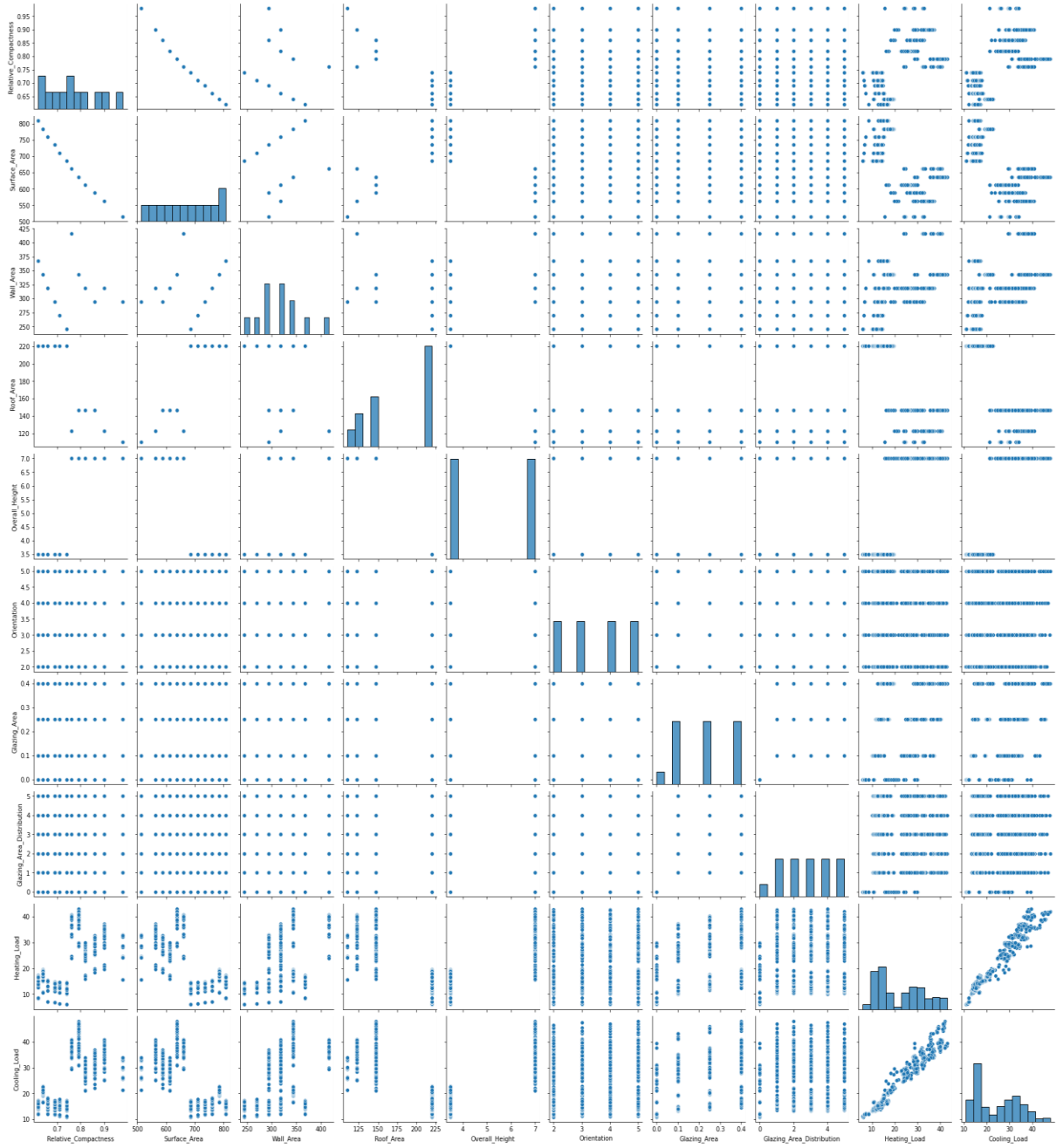
Heating_Load; Veri kümesindeki her bir bina için kWh/m² cinsinden ısıtma yükünü temsil eder. Sürekli bir sayısal değişkendir. Minimum ve maksimum değerler sırasıyla 6.01 ve 43.1Q'dir. Ortalama değer 22.31 kWh/m² ve standart sapma 10,09 kWh/m²'dir. Bu sütundaki değerler, bir binayı kışın konforlu bir sıcaklığa ısıtmak için gereken enerji miktarını temsil eder.

Cooling_Load; Bağımlı kompaktlık, yüzey alanı, duvar alanı, çatı alanı, genel yükseklik, yönelim, cam alanı ve cam alanı dağılımı gibi çeşitli özelliklere dayalı olarak her bir bina için soğutma yükünü (kWh/m² cinsinden) temsil eder. Soğutma yükü, konforlu bir iç mekan sıcaklığını korumak için bir binadan uzaklaştırılması gereken ısı miktarıdır. Bu sütun, 10.90 kWh/m²'lik minimum değerden 48.03 kWh/m²'lik maksimum değere kadar değişir ve ortalaması 24.59 kWh/m² ve standart sapması 9.51 kWh/m² dir.

Orientation ve glazing_area_distribution alanları kategori cinsinden değerlendirilmiştir. Verilere ait karşılaştırma matrisi Şekil 2’ de gösterilmiştir. Enerji verimliliği veri seti’ nin karşılaştırma matrisi Şekil 2’ de verilmiştir. Verilerin çift grafiği ise Şekil 3’ te verilmiştir.



Şekil 2. Karmaşıklık Matrisi



Şekil 3. Verilerin Çift Grafiği

Değişkenler arasındaki dağılımı ve ilişki kontrol edildiğinde `relative_compactness` ve `surface_area` özelliklerinin ilişkili olduğu ve `surface_area`, `relative_compactness`'tan daha yüksek diğer değişkenlerle korelasyona sahiptir, bu nedenle kolineariteyi önlemek için `surface_area`'yı düşürülmüştür.

`Orientation` ve `glazing_area_distribution` alanları kategorik değerlerden oluştuğu için `heating_load` ve `cooling_load` ile neredeyse hiç korelasyonu yoktur, bu nedenle boyutluluğu azaltmak için bu alanlar karşılaştırma sürecinin dışında tutulabilir.

3.1. Linear Regression Model

Makine öğrenimindeki doğrusal regresyon modelinin amacı, bağımlı değişken (hedef) ile bir veya daha fazla bağımsız değişken (özellik) arasındaki ilişkiyi, bu ilişkiyi en iyi şekilde temsil eden düz bir çizgi (veya çoklu boyutlarda hiper düzlem) uydurarak tahmin etmek veya açıklamaktır.

3.2. Tahmin Sonuçlarının Yorumlanması

Isıtma yükü modeli ve soğutma yükü modelinin ayarlanmış R karesi sırasıyla 0.9003 ve 0.8623'tür ve bu da modelin farklı binaların enerji verimliliğindeki farklılıkların sırasıyla %90.03'ünü ve %86.23'ünü açıklayabileceğini göstermektedir. Modeller verilere iyi uyum göstermekte ve bir konut binasının 6 metriğine dayalı olarak enerji verimliliğini tahmin etmede yararlı olabilir.

Uygun modellerin MAE'si (Ortalama Mutlak Hata) sırasıyla 2.3002 ve 2.4978'dir; bu da enerji verimliliğinin tahmin edilen değerlerinin gerçek değerlerden sırasıyla ortalama 2.3002 ve 2.4978 birim sapacağı anlamına gelir. Bu metrik, uygun modeller ve 6 özellik kullanılarak tahminler yapılırken beklenen hatayı gösterir ve hatanın kullanım durumunda kabul edilebilir olup olmadığını belirlememizi sağlamaktadır.

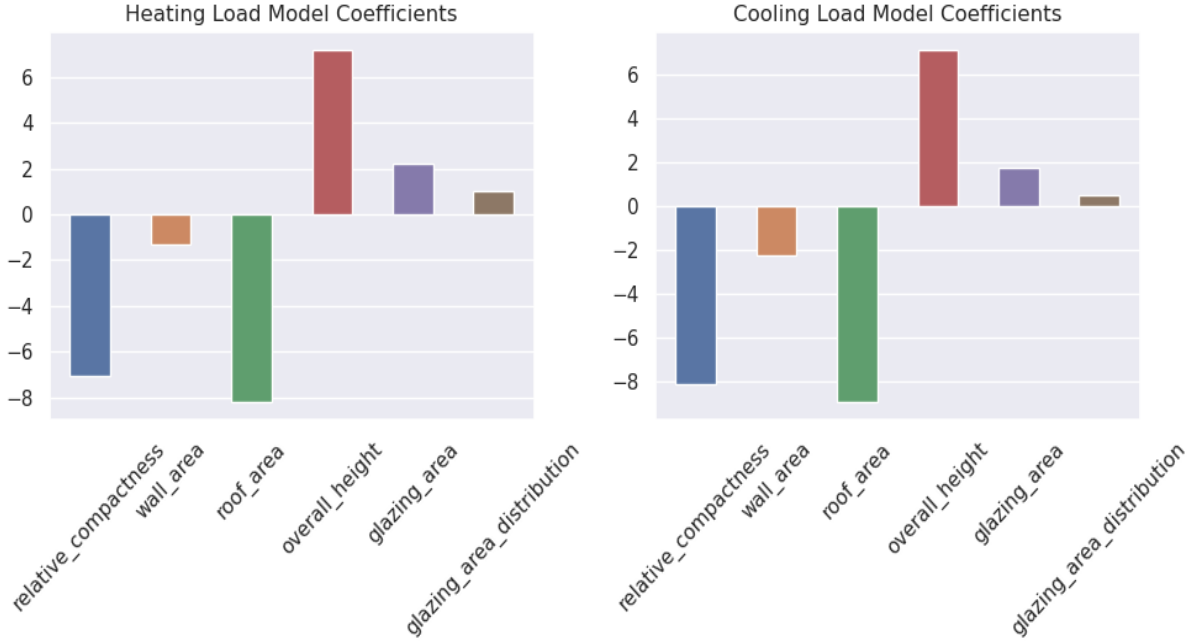
MSE (Ortalama Karesel Hata) sırasıyla 9.8933 ve 12.2353'tür. Bu metrik, her hatayı kare olarak büyük hataları cezalandırdığı için MAE'den daha büyüktür. MSE, farklı modellerin tahmin performansının istikrarsızlığını gösterebildiği için farklı modellerin performansını karşılaştırırken yararlı olabilir.

RMSE (Ortalama Karesel Hatanın Kökü) sırasıyla 3.1454 ve 3.4979'dur. Bu metrik, RMSE'nin birimini hedef değişkenin birimiyle aynı yapmak için ortalama karesel hatanın karekökünü alır, bu durumda enerji verimliliğidir. Yine de MAE'den daha büyüktür çünkü büyük hataları cezalandırır.

Tablo 5. Tahmin Sonuçları

	Adjusted R2	MSE	RMSE	MAE
Heating load	0.9003	9.8933	3.1454	2.3002
Cooling load	0.8623	12.2353	3.4979	2.4978

Modelin performansını değerlendirdikten sonra, bir konut binasının hangi özelliklerinin enerji verimliliğine en çok katkıda bulunduğunu görmek için özellik önemini kontrol edilmiştir. İlgili kontrol görseli Şekil 4’ te verilmiştir.



Şekil 4. Model Performansının Sıcak ve Soğuk Yük ’ün Karşılaştırması

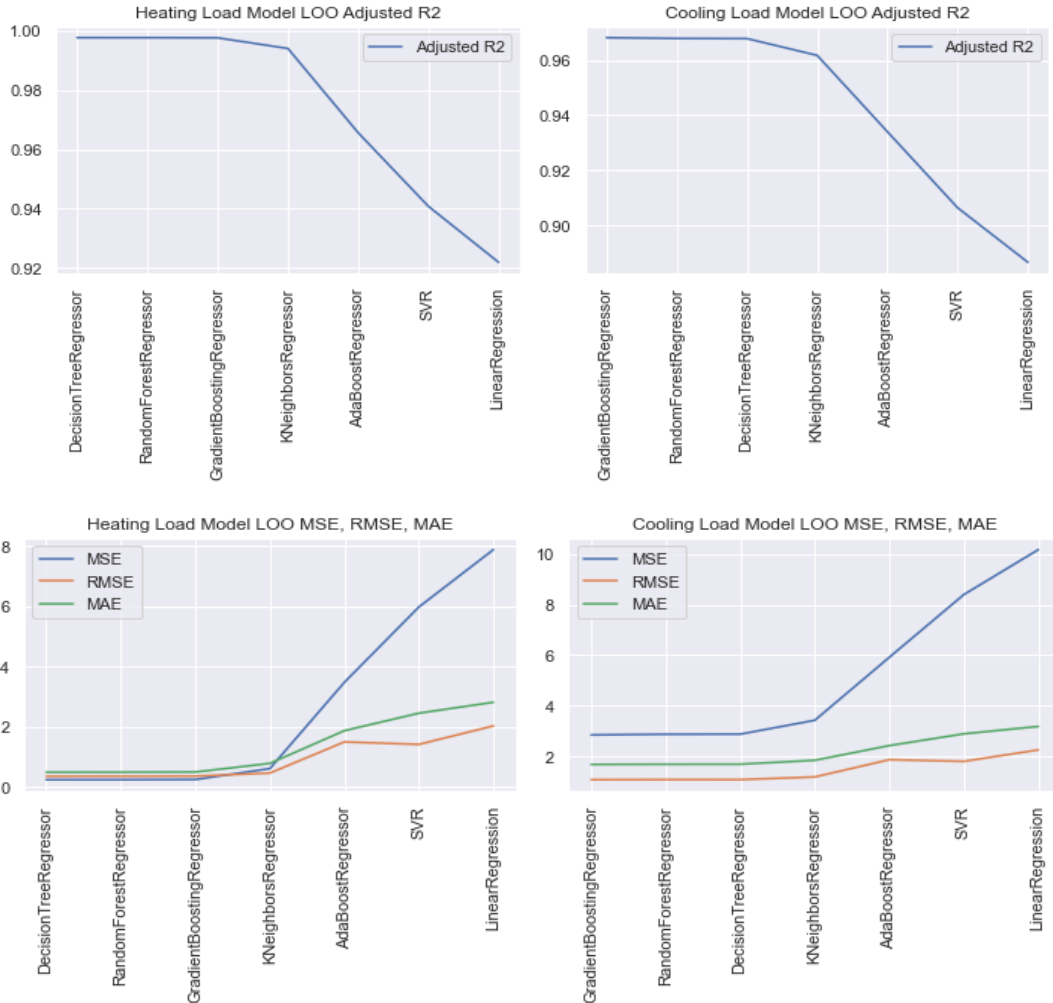
relative_compactness ve roof_area hem ısıtma yükü hem de soğutma yükü üzerinde önemli olumsuz etkiye sahiptir. Overall_height hem ısıtma yükü hem de soğutma yükü üzerinde önemli olumlu etkiye sahiptir. Bu üç özellik enerji verimliliği üzerinde en büyük etkiye sahiptir.

Tablo 6. Isıtma ve Soğutma Yükü

	Heating load model coefficients	Cooling load model coefficients
relative_compactness	-7.0958458730476455	-8.15611490605704
wall_area	-1.3159700922227593	-2.227514344870592
roof_area	-8.234793485188415	-8.970425365072838
overall_height	7.183266594217719	7.131722242939717
glazing_area	2.2416741152751656	1.7540712278958694
glazing_area_distribution	1.020804688621697	0.513058498698021
Heating load model intercept	21.91440595903166	24.170484171322162

Yukarıdaki çift diyagram, özellikler ve hedef değişkenler arasında çok fazla doğrusal ilişki olmadığını ve değişkenlerin normal dağılım göstermediğini gösteriyor, bu nedenle doğrusal regresyon modeli bu veri seti için en iyi seçim algoritma olmayabilir. Enerji verimliliğini tahmin etmek için daha iyi bir model olup olmadığını görmek için Karar Ağacı, Rastgele Orman, Gradient Boosting, Ada Boost, K-En Yakın Komşu, Destek Vektör

Makineleri (SVR) performansları incelenmiştir. Veri seti yalnızca 768 örnek içerdiğinden, çapraz doğrulama yöntemi olarak Leave-One-Out (LOO) kullanmak mümkündür.



Şekil 5. Farklı Algoritmalar Karşısında Tahmin Sonuçlarının Değerleri

Tablo 7. Isıtma Yüğü Model Sonuçları

Model	Heating	LOO_Adj_R2	LOO_MSE	LOO_RMSE	LOO_MAE
1	DecisionTreeRegressor	0.9977	0.2330	0.3457	0.4827
2	RandomForestRegressor	0.9977	0.2337	0.3470	0.4834
3	GradientBoostingRegressor	0.9976	0.2395	0.3514	0.4894
4	KNeighborsRegressor	0.9940	0.6040	0.4560	0.7771
5	AdaBoostRegressor	0.9656	3.4682	1.4889	1.8623
6	SVR	0.9408	5.9727	1.4068	2.4439
7	LinearRegression	0.9219	7.8788	2.0220	2.8069

Tablo 8. Soğutma Yüğü Model Sonuçları

Model	Cooling	LOO_Adj_R2	LOO_MSE	LOO_RMSE	LOO_MAE
1	GradientBoostingRegressor	0.9682	2.8537	1.0858	1.6893
2	RandomForestRegressor	0.9679	2.8797	1.0904	1.6970
3	DecisionTreeRegressor	0.9678	2.8842	1.0897	1.6983
4	KNeighborsRegressor	0.9624	3.3747	1.1850	1.8370
5	AdaBoostRegressor	0.9340	5.9163	1.8755	2.4323
6	SVR	0.9064	8.3925	1.8098	2.8970
7	LinearRegression	0.8866	10.1666	2.2634	3.1885

4. SONUÇ

Enerji verimliliği veri seti incelendiğinde, Isıtma yükünü tahmin ederken beklenen hata ortalama 0,3840 birimdir ve soğutma yükünü tahmin ederken beklenen hata ortalama 1.2718 birimdir. Bir konut binasının bu özellikleri, %1.7214’ lük hata yüzdesiyle ısıtma yükünü tahmin etmede daha faydalıdır, soğutma yükünü tahmin ederken ise hata yüzdesi %5.1724’tür.

Test veri kümesinde, Gradient Boosting, Karar Ağacı ve Rastgele Orman mükemmel performansa sahipken, K-En Yakın Komşu çapraz doğrulama sonucuna kıyasla daha kötü performansa sahiptir ve bu da KNN modelinin olası bir aşırı uyum sorununa işaret ettiği gözlenmektedir.

Sonuç olarak, ağaç tabanlı modeller, özellikler ve hedef değişkenler arasındaki doğrusal olmayan ilişkiler nedeniyle bu veri kümesinde doğrusal modellerden daha iyi performans gösterdiği görülmektedir.

Çıkar Çatışması Beyanı

Yazarlar arasında çıkar çatışması yoktur.

Araştırma ve Yayın Etiği Beyanı

Çalışma, araştırma ve yayın etiğine uygundur.

KAYNAKLAR

- [1] Hong, T., Langevin, J., & Sun, K. (2018, October). Building simulation: Ten challenges. In Building simulation (Vol. 11, pp. 871-898). Springer Berlin Heidelberg.
- [2] Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16(6), 3586-3592.

- [3] Reyna, J. L., & Chester, M. V. (2015). The growth of urban building stock: Unintended lock-in and embedded environmental effects. *Journal of Industrial Ecology*, 19(4), 524-537.
- [4] Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews*, 23, 272-288.
- [5] Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y., & Livingood, W. (2021). A review of machine learning in building load prediction. *Applied Energy*, 285, 116452.
- [6] Egwim, C. N., Alaka, H., Egunjobi, O. O., Gomes, A., & Mporas, I. (2024). Comparison of machine learning algorithms for evaluating building energy efficiency using big data analytics. *Journal of Engineering, Design and Technology*, 22(4), 1325-1350.
- [7] Forootan, M. M., Larki, I., Zahedi, R., & Ahmadi, A. (2022). Machine learning and deep learning in energy systems: A review. *Sustainability*, 14(8), 4832.
- [8] Yu, Z., Haghighat, F., Fung, B. C., & Yoshino, H. (2010). A decision tree method for building energy demand modeling. *Energy and Buildings*, 42(10), 1637-1646.
- [9] Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 1192-1205.
- [10] Gonzalez, P. A., & Zamarreno, J. M. (2005). Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and buildings*, 37(6), 595-601.
- [11] Dong, B., Cao, C., & Lee, S. E. (2005). Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*, 37(5), 545-553.
- [12] Ekici, B. B., & Aksoy, U. T. (2009). Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, 40(5), 356-362.
- [13] Liang, X., Hong, T., & Shen, G. Q. (2016). Improving the accuracy of energy baseline models for commercial buildings with occupancy data. *Applied energy*, 179, 247-260.
- [14] Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2016). Deep learning for estimating building energy consumption. *Sustainable Energy, Grids and Networks*, 6, 91-99.
- [15] Reyna, J. L., & Chester, M. V. (2015). The growth of urban building stock: Unintended lock-in and embedded environmental effects. *Journal of Industrial Ecology*, 19(4), 524-537.

- [16] Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., ... & Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82, 1027-1047.
- [17] Ciulla, G., & D'Amico, A. (2019). Building energy performance forecasting: A multiple linear regression approach. *Applied Energy*, 253, 113500.
- [18] Yang, S., Wan, M. P., Chen, W., Ng, B. F., & Dubey, S. (2020). Model predictive control with adaptive machine-learning-based model for building energy efficiency and comfort optimization. *Applied Energy*, 271, 115147.
- [19] <https://www.kaggle.com/datasets/ujjwalchowdhury/energy-efficiency-data-set/data>
- [20] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer.
- [21] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.
- [22] Hastie, T. (2009). *The elements of statistical learning: data mining, inference, and prediction*.