

Binalarda Enerji Verimliliğine Makine Öğrenmesi Temelli Hibrit Bir Kümeleme ve Tahminleme Yaklaşımı

Aslıhan SAĞIROĞLU^{1*}, Alev TAŞKIN²

Öz

Günümüzde enerji kullanımında verimlilik hem maliyetlerin yükselmesi hem de çevresel kirliliğin önlenmesi açısından çok önemli bir konumda bulunmaktadır. Enerji kullanımının önemli bir kısmı binalarda gerçekleştiği için binalarda enerji verimliliği ayrıca bir çalışma konusu olmuş, yeşil bina gibi kavramlarla birlikte binalar için enerji sınıfları oluşturulmuştur. Bu enerji sınıfları oluşturulurken dikkate alınan enerji tüketim türlerinden iki çeşidi; ısıtma yükü ve soğutma yüküdür. Binalarda aydınlanma ve havalandırma amaçlı enerji tüketimine kıyasla, ısıtma ve soğutma yükleri enerji tüketimi açısından daha fazla göze çarpmaktadır. Bu çalışmada ısıtma ve soğutma yükleri üzerinden binalarda enerji kullanım tahminleri makine öğrenmesi yöntemleriyle gerçekleştirilmiş, öncesinde yapılan kümeleme analizi ile tahminleme performansının iyileştirilmesi sağlanmaya çalışılmıştır. Analizler sırasında SPSS programı ve Python programlama dili kullanılmıştır.

Anahtar Kelimeler: Kümeleme, Tahminleme, Enerji verimliliği, Makine öğrenmesi.

A Machine Learning Based Hybrid Clustering and Prediction Approach to Energy Efficiency in Buildings

Abstract

Efficiency in energy usage has become increasingly important in today's World due to the rising costs and the need to prevent environmental pollution. As buildings are responsible for a major portion of energy consumption, energy efficiency in buildings has gained significant attention, which has led to the creation of energy classifications, including the concepts as green buildings. These classifications are based on two types of energy consumptions: heating load and cooling load, which are more significant than lighting and ventilation. This study utilized machine learning methods to predict energy usage in buildings by analyzing heating and cooling loads. Prior cluster analysis was performed to enhance the prediction performance. The study employed SPSS and Python programs for the analysis.

Keywords: Clustering, Prediction, Energy efficiency, Machine learning.

¹İstanbul Arel Üniversitesi, Endüstri Mühendisliği Bölümü, İstanbul, Türkiye, aslihansagiroglu@arel.edu.tr

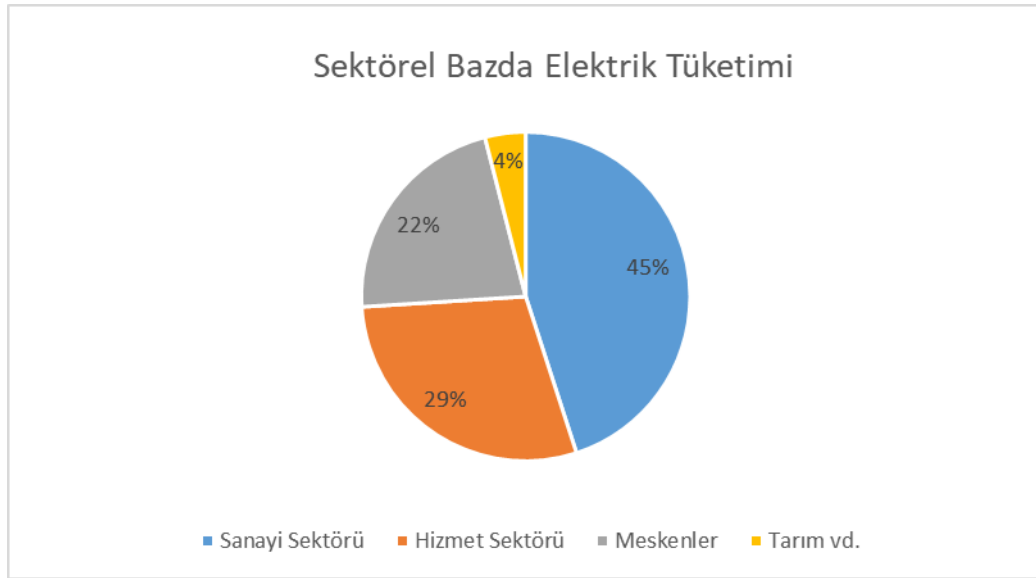
²Yıldız Teknik Üniversitesi, Endüstri Mühendisliği Bölümü, İstanbul, Türkiye, ataskin@yildiz.edu.tr

*Sorumlu Yazar/Corresponding Author

1. Giriş

Günümüz dünyasında tüketilen enerjinin büyük bir bölümü hâlâ fosil yakıtlardan elde edilmektedir. Bu durumun sebep olduğu çevresel kirlilik sebebiyle ve kaynakların etkin kullanımı gereğince günümüzde enerji verimliliği kavramı hem ulusal hem de uluslararası mecralarda üzerinde oldukça çalışılan bir konu haline gelmiştir.

Ülkemizde Enerji ve Tabii Kaynaklar Bakanlığının yaptığı tanıma göre enerji verimliliği, binalarda yaşam standardı ve hizmet kalitesinin, endüstriyel işletmelerde ise üretim kalitesi ve miktarının düşüşüne yol açmadan, birim veya ürün miktarı başına enerji tüketiminin azaltılması olarak tanımlanmıştır (Enerji Verimliliği - T.C. Enerji ve Tabii Kaynaklar Bakanlığı, t.y.). Çalışma ve Sosyal Güvenlik bakanlığının 2019 yılına kadar paylaştığı elektrik enerjisinin sektörlere göre kullanım verileri şekil 1’de görülmektedir. Bu verilere göre 2019 yılında Türkiye’de, elektrik enerjisi tüketiminde en fazla payı sanayi sektörü (45%) ve hizmet sektörü (29%) almış, bunları meskenler (22%) ile tarım ve diğer sektörler (4%) takip etmiştir T.C. Çevre, Şehircilik ve İklim Değişikliği Bakanlığı, t.y.).



Şekil 1. Sektörel Bazda Elektrik Tüketimi Dağılım Grafiği

Bu kapsamda binalarda enerji verimliliği, yeşil binalar ve çevre dostu binalar kavramları gün geçtikçe literatürde daha geniş yer bulmaya başlamıştır (Doğan & Seçme, 2018; Erdede vd., 2014; Güzelkokar, 2019). Bu kavramların bulunduğu ortak nokta enerjinin yapılarda etkin kullanımı yani verimlilik odaklı olmasıdır. Çevre ve Şehircilik Bakanlığının hazırlamış olduğu Binalarda Enerji Performans Yönetmeliği (BEP)’ne göre çevreyi koruma ve etkin enerji kullanımı hedefleri doğrultusunda, bina enerji performansını iyileştirmeye yönelik mimari tasarımlar, ısı yalıtım esasları

ve sızdırmazlık, havalandırma ve iklimlendirmenin tasarımları gibi konuların yanında ısıtma ve soğutma sistemleri tasarımı ve uygulama esasları konusunda birçok noktaya değinilmiş ve yönlendirmeler yapılmıştır(Binalarda Enerji Performansı Yönetmeliği, t.y.).

“5627 Sayılı Enerji Verimliliği Kanunu” ve bu kanun doğrultusunda çıkartılan BEP Yönetmeliği’ne göre, enerji kaynaklarının verimli kullanılmasını, enerji israfının önlenmesini ve çevrenin korunmasını sağlamak için, asgari olarak binanın enerji ihtiyacı ve enerji tüketim sınıflandırması, yalıtım özellikleri ve ısıtma ve/veya soğutma sistemlerinin verimi ile ilgili bilgileri içeren belge, enerji kimlik belgesidir (Binalarda Enerji Performansı Yönetmeliği, t.y.).



Şekil 2. Enerji Kimlik Belgesi(Şükür, t.y.)

Şekil 2’de görülen enerji kimlik belgesine göre her enerji tüketim performansı ayrı sınıflandırılır. Yani bir binanın ısıtma yükü ve soğutma yükü enerji sınıfları birbirinden farklı olabilmektedir. Isıtma enerji yükü performansı A sınıfı olan bir binanın soğutma enerji yükü veya aydınlatma enerji yükü performansı B sınıfı olabilir.

Sertifikalandırma öncesinde ve genel anlamda enerji planlaması yapılabilmesi için binalarda enerji tüketimlerinin önceden görülebilmesi oldukça önemlidir. Özellikle binaların tasarım aşamasında fiziksel özelliklerinin enerji verimliliğine uygun olarak tasarlanması, sonradan yapılacak çalışmalara oranla daha etkili olmaktadır. Bu anlamda enerji performans tahmini ve enerji tüketim tahmini çalışmaları oldukça önemli olup literatürde sıkça karşımıza çıkmaktadır (Abbasimehr vd., 2023; Khan & Osińska, 2023; Tsanas & Xifara, 2012). Bu çalışmada binalarda enerji performansı ölçümü için binaların ısıtma ve soğutma yükleri ele alınmıştır. Binaların fiziksel özellikleri giriş verisi olarak kullanılıp ısıtma ve soğutma yükleri üzerinden enerji performansı tahminlenmiştir. Tahminleme aşamasında birbirinden farklı modeller kullanılarak tahmin performansı en iyilenmeye çalışılmış, tahminleme öncesi ise kümeleme analizine yer verilmiştir. Kümeleme işlemi sırasında 3

farklı algoritma kullanılmış ve en iyi performansı gösteren algoritmanın kümeleme sonuçlarıyla analizlere devam edilmiştir.

Kullanılan kümeleme yöntemleri literatürde sıklıkla kullanılan ve karşılaştırılan K-ortalamlar, hiyerarşik kümeleme(aglomeratif) ve Gürültü İçeren Uygulamalar için Yoğunluk Tabanlı Mekânsal Kümeleme (DBSCAN) algoritmalarıdır.(L. K. Cheng vd., 2019; Dong vd., 2019; Garikapati vd., 2021; Guo vd., 2024; Jahan & Sanam, 2024; Zurini, 2013)

Veriler hem kümelenerek hem de kümeleme yapılmadan tahminleme işlemine tabi tutularak kümeleme analizinin verilerde oluşturduğu homojenliğin tahmin performansına etkisi analiz edilmiş, aynı zamanda birbirinden farklı 7 makine öğrenmesi algoritmasının performansı birbiriyle kıyaslanmıştır. Analizler sayısal veriler üzerinden gerçekleştirilmiş olup, elde edilen sonuçlar değerlendirilmiştir.

2. İlgili Çalışmalar

Literatürde binalarda enerji tüketim performansı tahminleri, regresyon modelleri ve enerji sınıflarının tahminlenmesi şeklinde ele alınmıştır. Bu konuda yapılan çalışmaların özeti bu bölümde sunulacaktır.

Ekici ve Aksoy, binalarda ısıtma ve soğutma yüklerinin tahmininde uyarlamalı ağ tabanlı bulanık çıkarım sistemi (ANFIS) yöntemini uygulamış ve tahmin performansının yüksekliğini belirterek yöntemi alternatif tahmin yöntemi olarak önermişlerdir (Bektas Ekici & Aksoy, 2011). Tsanas ve Xifara binalarda ısıtma ve soğutma yüklerinin tahminini lineer regresyon(LR) ve rastgele orman(RO) yöntemleriyle gerçekleştirmiş, istatistiksel olarak da yorumlamışlardır (Tsanas & Xifara, 2012). Aqlan vd. binalarda enerji verimliliğini etkileyen faktörlere dikkat çekmek istemiş, yöntem olarak yapay sinir ağları (YSA) ve K-ortalamlar algoritmalarını kullanmışlardır(Aqlan vd., 2014). Cheng ve Cao binaların enerji performansını tahmin etmek için evrimsel çok değişkenli uyarlamalı regresyon eğrilerinin kullanılmasını önermişlerdir (M.-Y. Cheng & Cao, 2014). Chou ve Bui binalarda enerji verimliliği verileri üzerinden destek vektör regresyonu (DVR), YSA, sınıflandırma ve regresyon ağacı, ki-kare otomatik etkileşim detektörü, genel doğrusal regresyon ve topluluk çıkarım modeli dahil olmak üzere çeşitli veri madenciliği teknikleri kullanarak tahmin performanslarını kıyaslamışlardır (Chou & Bui, 2014). Sönmez vd. yapay arı kolonisi tabanlı k-en yakın komşu(KEK), genetik algoritma tabanlı KEK, uyarlanabilir yapay genetik algoritma ve uyarlanabilir YSA içeren yapay arı kolonisi algoritması kullanmış ve klasik YSA ve KEK algoritmalarına göre daha iyi sonuçlar elde etmişlerdir (Sonmez vd., 2015). Duarte vd. karar ağacı(KA), çok katmanlı algılayıcılar (ÇKA), RO, DVR yöntemleriyle binalarda enerji tüketimi tahminlemesi yapmışlardır (Duarte vd., t.y.)Naji vd. 2016 binaların enerji tüketimini tahminlemek

için ekstrem öğrenme makinesi metodunu kullanmışlardır, sonuçları genetik algoritma ve yapay sinir ağlarına göre daha iyi bulmuşlardır(Naji, Keivani, vd., 2016). Naji vd. binalarda enerji verimliliği ölçümü için ANFIS kullanmışlardır (Naji, Shamshirband, vd., 2016). Yang vd. genetik algoritma ve YSA metotlarını hibrit olarak kullanmış ve sadece YSA algoritmasının kullanıldığı çalışmanın sonuçlarına göre daha düşük hata oranı elde etmişlerdir(Yang vd., 2016). Sholahudin vd. binalarda ısıtma ve soğutma yükleri üzerinden YSA ile tahmin işlemi yapmış ve elde edilen değerleri etkileyen değişkenleri irdelemişlerdir (Sholahudin vd., 2016). Nilashi vd. kümeleme ve temel bileşenler analiziyle birlikte ANFIS kullanarak binalarda enerji performans tahmini yapmış, sonuçların etkili olduğunu belirtmişlerdir (Nilashi vd., 2017). Al-Rakhami vd. ısıtma ve soğutma yükü tahmini için extra gradyan artırma (XGA) algoritması ile topluluk öğrenmesi yöntemini kullanmışlardır (Al-Rakhami vd., 2019). Peker vd. binalarda ısıtma ve soğutma yüklerinin tahmininde KEK, RO, LR ve DVR algoritmalarını kullanmıştır. Bu algoritmalarından RO algoritmasıyla en iyi sonucu elde etmiştir (Peker vd., 2017). Yücel ve Namlı çalışmalarında 127 binaya ait verilerle iki model oluşturmuş, ısıtma sınıfı modelinde çok katmanlı YSA, Bayes sınıflandırıcı, KEK sınıflandırıcısı ve C4.5 algoritması kullanmış; soğutma sınıfında ise yalnızca YSA modeli enerji sınıflandırması yapmışlardır (Namlı & Yücel, 2018). Kavaklıoğlu konutların ısıtma ve soğutma yüklerini modellemek için en küçük kareler yöntemi ve kısmi en küçük kareler yöntemini kullanmıştır (Kavaklioglu, 2018). Roy vd. çok değişkenli uyarlanabilir regresyon, ekstrem öğrenme makinesi ve iki yöntemin hibrit modelini kullanarak performanslarını klasik yöntemlerle kıyaslamışlardır (Sekhar Roy vd., 2018). Kumar vd. ekstrem öğrenme makinesinin varyantlarını içeren bir topluluk modeliyle binalarda enerji performansı tahmini yapmışlardır ve tatmin edici sonuçlar elde etmişlerdir (Kumar vd., 2018). Susac vd. tarafından 3659 bina üzerinden önce ki-kare testi ile enerji verimliliği açısından en önemli üç özellik çıkarılmış ve sonrasında bu özellikler üzerinden kümeleme yapılmıştır. Ardından YSA ile hem küme bazında hem total veride enerji performansları tahminlenmiş ve kümeleme analizinin tahmin performansına etkisi incelenmiştir (Zekić-Sušac vd., 2018). Gümüşçü vd. çalışmalarında binalarda ısıtma ve soğutma yükleri eşik değerler belirlenerek sınıflara atanmış ve KEK, DVR ve KA algoritmaları kullanarak bu algoritmaların performansları kıyaslamıştır(Gümüşçü vd., 2018). Khalil vd. binaların ısıtma ve soğutma yükleri üzerinden YSA yöntemiyle bir tahmin çalışması gerçekleştirmiş ve tahmin performansını %99,6 olarak ölçmüştür (Khalil vd., 2019). Roy vd. binalarda ısıtma ve soğutma yükü tahmini için derin sinir ağlarını kullanmışlar, sonuçları gradyan artırma(GA) ile karşılaştırmışlardır (Roy vd., 2020). Seyedzadeh makalesinde binalarda enerji yüklerini DVM, YSA, RO, gradyan artırılmış regresyon ağaçları, XGA ve Gauss süreçleri algoritmalarını kullanarak tahminlemiş ve sonuçlarını kıyaslamıştır (Seyedzadeh vd., 2019). Moradzadeh vd. binalarda ısıtma ve soğutma enerji yüklerinin tahmin edilmesi için YSA ve DVM yöntemlerini uygulamış ve bu yöntemlerin etkinliğini göstermişlerdir (Moradzadeh vd., 2020). Dilber

ve Özdemir binalarda ısıtma ve soğutma yükleri tahmini için doğrusal regresyon, KEK, DVR, KA, RO, uç gradyan artırma (UGA) ve YSA algoritmalarını kıyaslamış ve UGA algoritmasını en başarılı algoritma olarak bulmuşlardır (Dilber & Özdemir, 2022). Olu-Ajayi vd. çalışmalarında binalarda enerji sınıflarının tahmin edilmesi için RO, GA, Ekstra Ağaçlar (EA), KA, KEK, DVR, GP, ÇKA ve Ada artırma algoritmalarını kullanmış olup, GA algoritması tahmin performansı en iyi algoritma olarak seçilmiştir (Olu-Ajayi vd., 2022). Yıldız vd. çalışmalarında binalarda ısıtma ve soğutma enerji yüklerinin tahminlemede YSA yöntemini kullanmışlardır (Yıldız vd., 2021). Sajjad vd. ısıtma ve soğutma yüklerini aynı anda tahmin edebilmek için kapılı tekrarlayan ünite şeklinde ifade edilen GRU'ya dayalı bir tahmin modeli sunmuş ve yüksek doğruluk oranları elde etmişlerdir (Sajjad vd., 2020). Canbay ve Taş LR, SVR, RF, DT, KNN makine öğrenmesi yöntemlerinin tek çıktı ve çok çıktı regresyon analizinde kullandığı çalışmalarında probleme uygun parametreler ve katmanlar ile tasarlanan bir derin sinir ağı ile hem tek çıktı hem çok çıktı regresyon analizinde yüksek başarı elde edilmiştir (Canbay & Taş, 2022). Lu vd. binalarda ısıtma ve soğutma yükü tahmini için makine öğrenmesi modellerini bir arada kullanabilen AutoML tabanlı bir model önermişlerdir (Lu vd., 2023). Yu vd. ısıtma, soğutma ve elektrik yüklerinin tahminlemede uzun kısa süreli bellek (LSTM), XGA ve GRU yöntemlerini kullanmışlardır (Yu vd., 2023).

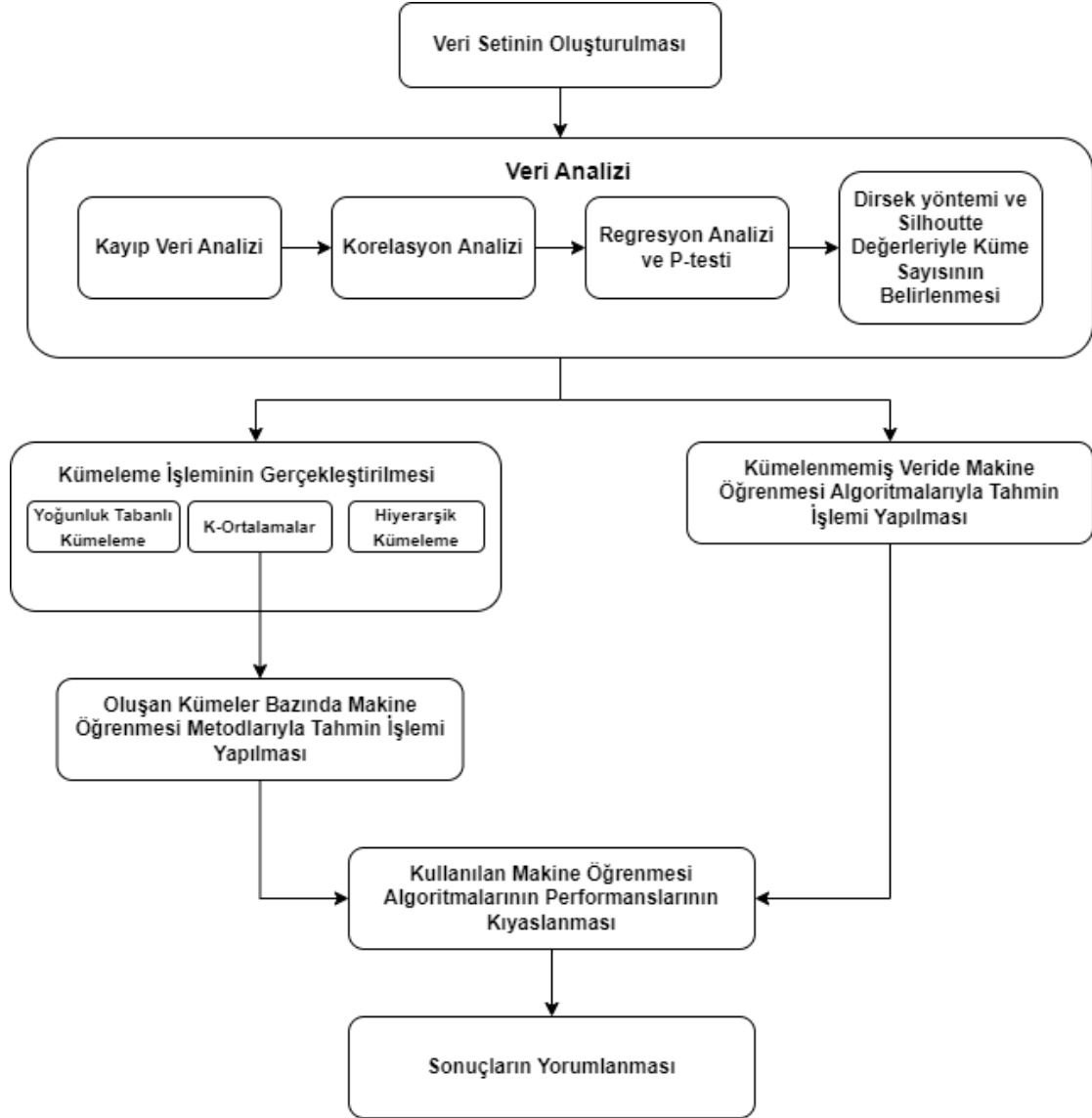
Genel olarak çalışmalarda birden fazla makine öğrenmesi algoritması kullanılarak karşılaştırma yapılmış olup, kullanılan veri setlerinin farklılığına göre farklı algoritmaların en iyi tahmin performansı sergilediği görülmüştür. Makine öğrenmesi algoritmaları istatistiksel analizlerle de desteklenmiştir. Bu çalışmada p testi ve korelasyon analizi gibi istatistiksel analizler yapılmış, kümeleme analizi tahminleme işlevini destekleyici analiz olarak kullanılmıştır. Optimum küme sayısı iki farklı yöntemle elde edilmiş ve sonuçları karşılaştırılmıştır.

3. Materyal ve Metot

Makine öğrenmesi, yapay zekânın bir alt dalı olarak literatürde karşımıza çıkmaktadır. Makine öğrenmesi modelleri, üç türe ayrılabilir: i) denetimli tabanlı makine öğrenmesi modelleri; ii) denetimsiz tabanlı makine öğrenmesi modelleri; iii) topluluk öğrenimine dayalı makine öğrenmesi modelleri. Denetimli tabanlı makine öğrenmesi modelleri, regresyon, KA, parametrik olmayan ve parametrik sınıflandırma ve destek vektör makinesi (DVM) gibi modelleri ifade eder (Ahmad vd., 2018). Denetimsiz tabanlı öğrenme modelleri ise kümeleme analizlerini içerir. Bu çalışmada denetimli ve denetimsiz makine öğrenmesi algoritmaları kullanılmıştır.

Binalarda enerji verimliliği ölçümleri için ısıtma ve soğutma yüklerinin ayrı ayrı tahmin edilmesi amacıyla 768 adet binaya ait veriler kullanılmıştır. Bu veriler önce homojenliği sağlamak ve homojenliğin tahmin performansına olan etkisini incelemek için kümeleme analizine tabi tutulmuş,

sonrasında ise küme bazında makine öğrenmesi algoritmalarıyla tahmin işlemi gerçekleştirilmiştir. Kümeleme işlemi sırasında 3 farklı kümeleme algoritması kullanılmıştır. Tahmin işlemi aynı zamanda kümelenmemiş veriye de uygulanmış ve tahmin performansları birbiriyle karşılaştırılmıştır. Tahminleme işlemi için de ayrıca birden fazla algoritma kullanılarak performansları kıyaslanmıştır. Çalışmanın akış şeması şekil 3'te sunulmaktadır.



Şekil 3. Önerilen metodoloji akış şeması

3.1. Veri Analizi

Öznitelik seçimi (feature selection), bir veri setinde bulunan n adet özellik arasından o veri setini en iyi temsil edecek k adet özelliğin seçilmesidir (Forman, 2003). Veri analizi işlemlerinde, önce kayıp veri analizi yapılmış sonrasında öznitelik seçimi için korelasyon analizi yapılmıştır. Korelasyon analizi iki değişken arasındaki pozitif ya da negatif yönlü ilişkiyi ifade eder. Korelasyon değerleri “1” değerine ne kadar yakınsa değişkenler arasındaki ilişki o derece kuvvetli, “0” değerine

ne kadar yakınsa değişkenler arası ilişki o derece zayıftır. Değerin negatif olması sadece ilişkinin ters yönlü olduğunu ifade eder. Spearman sıra korelasyon katsayısı iki değişken arasındaki gücün ölçümü olarak Pearson korelasyon katsayısına alternatif parametrik olmayan bir sıra istatistiğidir. Çalışmada korelasyon ilişkisi incelenirken Spearman korelasyon katsayısı kullanılmıştır.

Değişkenler arasındaki ilişkiyi açıklamak için kullanılan matematiksel model regresyon modeli olarak adlandırılır (Birkes & Dodge, 2011). Sonraki adımda 8 adet girdi değişkeni ve 2 adet çıktı değişkeni arasındaki ilişkiyi irdelemek için regresyon modeli (1) denklemine göre kurulmuştur;

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad i=1,2,\dots,n \quad (1)$$

Model üzerinden p değerleri hesaplanmış ve değişkenlerin modeldeki anlamlılığı yorumlanmıştır. Önemli olan bu değerlerin 0,05 (anlamlılık düzeyi kabulü) değerinden küçük olması (ve ne derecede küçük olduğu) ile ilgili yapılan yorumdur (Mehmet; Subaşı, 2005).

3.2. Kümeleme Analizi

Kümeleme işlemi, verilerin homojen olarak dağılmasını sağlamak amacıyla yapılmaktadır. Kümeleme analizi için kullanılan 3 algoritmanın özellikleri bu kısımda anlatılacaktır.

K-Means algoritması, enerji literatüründe yaygın olarak kullanılan ve yüksek başarı oranı rapor edilen bir yöntem olduğundan bu çalışmada tercih edilmiştir (Miraftabzadeh vd., 2023). Ayrıca karşılaştırmalı bir analiz gerçekleştirebilmek amacıyla, yine literatürde etkinliği sıklıkla vurgulanan DBSCAN (Aurangzeb, 2024; Economopoulou vd., 2024; Wang vd., t.y.) ve aglomeratif kümeleme algoritmaları da değerlendirmeye alınmıştır (AlMahamid & Grolinger, 2022; Khadka vd., 2024).

3.2.1. Gürültü İçeren Uygulamalar için Yoğunluk Tabanlı Mekânsal Kümeleme

Literatürde sıklıkla DBSCAN olarak geçen algoritma, yoğunluk tabanlı bir kümeleme algoritmasıdır ve özellikle farklı şekillerdeki kümeleri belirlemek ve gürültüyü ayırt etmek için tasarlanmıştır. Algoritma, her veri noktasının belirlenen bir yarıçap (Eps) içindeki komşularının sayısını (MinPts) kontrol eder ve belirli bir yoğunluk eşiğini geçen noktaları kümelere dahil eder.

Çalışma prensibi olarak ise; her nokta için belirli bir Eps yarıçapı içinde kaç komşusu olduğuna bakılır. Eğer bir noktanın komşu sayısı MinPts değerinden büyükse, bu nokta çekirdek nokta olarak kabul edilir. Bir çekirdek noktadan başlayarak, ona komşu noktalar bulunur ve bu noktalar da kümeye eklenir. Eğer bir nokta bir küme içindeki çekirdek noktaya doğrudan veya dolaylı olarak bağlı ise

kümeye dahil edilir. Gürültü noktaları (yoğunluğu düşük olanlar) kümelere dahil edilmez ve gürültü olarak işaretlenir (Sander vd., 1998).

3.2.2. Hiyerarşik Kümeleme (Aglomeratif Algoritma)

Aglomeratif hiyerarşik kümeleme, her bir veri noktasını başlangıçta ayrı bir küme olarak ele alan ve ardından benzer kümeleri iteratif olarak birleştirerek hiyerarşik bir yapı oluşturan bir kümeleme yöntemidir. Bu süreç, bütün veri noktaları tek bir kümede birleşene veya belirlenen bir durma kriterine ulaşana kadar devam eder. Hangi seviyede birleştirmelerin duracağına kullanıcı karar verir. Sonuçlar, dendrogram adı verilen ağaç yapılarıyla görselleştirilir. Bu yöntemin klasik uygulamalarında, iki kümenin benzerliği küme içi mesafeler veya bağlantı ölçütleri ile belirlenir. Tek bağlantı, tam bağlantı, ortalama bağlantı ve Ward yöntemi gibi farklı benzerlik ölçütleri, birleşme kararlarını etkiler. (Bakkelund, 2022; Fernández & Gómez, 2008)

3.2.3. K-Ortalamlar ile Kümeleme

Kümeleme analizinde kullanılan hiyerarşik olmayan algoritmalar içerisinde en sık kullanılan algoritmalarından biri K-ortalamlar algoritmasıdır.

K-ortalamlar algoritması, kümeleme hatasına göre en uygun çözümleri bulur. Birçok kümeleme uygulamasında kullanılmış olan hızlı yinelemeli bir algoritmadır. Kümeleme hatasını en aza indirmek için küme merkezlerinin başlangıçta rasgele konumlara yerleştirilmesiyle başlayan ve küme merkezlerinin her adımda hareket ettirilmesiyle ilerleyen nokta tabanlı bir kümeleme yöntemidir (Likas vd., 2003).

K-ortalamlar algoritmasının uygulanabilmesi için bir küme sayısı belirtilmelidir. Bunun için kullanılan çeşitli yöntemler mevcut olmakla birlikte, dirsek (elbow) yöntemi öne çıkmaktadır. Bu yöntemde küme içi kareler toplamı (KİKT) değeri hesaplanır. Bu değer, her bir verinin küme merkezine olan uzaklığının karelerinin toplamı alınarak hesaplanır. KİKT'deki değişim miktarının azaldığı nokta dirsek noktası olarak adlandırılır ve optimum k değeri olarak kullanılır (Coşkun vd., 2021).

Başka bir optimum küme sayısı belirleme yöntemi olarak silhoutte skoru kullanılır. Silhouette skoru, kümeleme geçerliliğini değerlendirmek için kullanılan bir araçtır. Açıkça birbirinde ayrılmış kümeler için uygun bir oran ölçeği verisi kullanarak en uygun küme sayısını seçmek için oluşturulan bir değerdir (Rousseeuw, 1987).

3.3. Tahminleme

Tahminleme işlemi mevcut veri setindeki giriş ve hedef değerlerine bağlı olarak gelecekteki bir olayın veya durumun sonuç değerini tahmin etmeyi amaçlar. Bir makine öğrenmesi algoritmasında tahminleme modeli ise eğitim verileri kullanarak örüntüleri yakalar ve bu örüntülere dayanarak gelecekteki gözlemleri tahmin eder.

Kullanılan algoritmalar konuyla ilgili literatürde sıkça kullanılan ve başarı oranları yüksek olan LR, KEK, KA, RO, GA, XGA algoritmalarıdır (Bektas Ekici & Aksoy, 2011; DiLber & Özdemir, 2022; Duarte vd., t.y.; Gümüşçü vd., 2018; Olu-Ajayi vd., 2022; Peker vd., 2017; Sonmez vd., 2015; Tsanas & Xifara, 2012; Yu vd., 2023). Bu algoritmalara ek olarak başarısının denenmesi ve aşırı öğrenmeye karşı olan başarısı sebebiyle CatBoost (CB) algoritması karşılaştırma algoritmaları arasına eklenmiştir.(Zhang vd., 2021)

Bu çalışmada kullanılan algoritmaların açıklamaları aşağıdaki gibidir:

LR: LR bir veya birden fazla (bağımsız) değişken ile bir (bağımlı) değişken arasındaki bağlantıyı modellemek için kullanılan istatistiksel bir yöntemdir (Cook, 2015).

GA: GA algoritmasında temel mantık şu şekildedir; gelişmiş tahmin doğruluğuna sahip "güçlü bir öğrenen" model için "zayıf öğrenenler" birkaç basit modeli yinelemeli olarak bir araya getirir. GA yöntemi özet olarak kayıp fonksiyonunu en aza indiren ek bir model bulmayı amaçlayan sayısal bir optimizasyon algoritması olarak tanımlanabilir (Yüce & Kabak, 2021).

XGA: XGA, GA algoritması üzerinden geliştirilmiş bir makine öğrenmesi modelidir (Chen & Guestrin, 2016). GA yönteminin sınırlamalarını ortadan kaldırmak için oluşturulan XGA yönteminin getirdiği yenilik model karmaşıklığını kontrol eden bir amaç fonksiyonu içermesidir. Bu fonksiyon, paralel hesaplamaları ve optimum hesaplama hızının korunmasını mümkün kılar (Ercan, 2022).

CB: CB algoritması eğitim sırasında sınıflandırma özellikleriyle başa çıkabilen güçlü bir makine öğrenmesi algoritmasıdır. CB, mevcut ağaç için yeni ayırma noktaları oluştururken doğruluğu iyileştirmek için kombinasyonları dikkate almak üzere açgözlü (greedy) stratejiler kullanır. Ek olarak, algoritma gradyan yanlılığının üstesinden gelebilir ve aşırı öğrenmeyi etkili bir şekilde önleyebilir (Zhang vd., 2021).

KEK: KEK yöntemi en yakın k adet komşunun ortalamasını temel alan bir algoritmadır. Örüntü tanımlamada bu yöntem parametrik olmayan, sınıflandırma ya da regresyon analizleri için kullanılır. Bu yöntemde eğitim kümesindeki verilere göre Öklid uzaklığı ya da seçilebilecek başka bir uzaklık hesaplama yöntemi ile bütün elemanlar arasındaki uzaklık hesaplanır (Altman, 1992).

KA: KA algoritması veri yapılarındaki ağaç yapısı ile benzerlik göstermektedir. KA, düğüm ve dallardan oluşan anlaşılması kolay, her bir dalın bir olasılık durumunu temsil ettiği ve prensip olarak

veriyi özyineleyerek alt gruplara dallanma yaparak bölen bir algoritmadır (Albayrak & Yılmaz, 2009).

RO: RO algoritması, temeli KA algoritmasına dayanan bir çeşit topluluk öğrenme algoritmasıdır. Birden fazla karar ağacı ile karar ormanı oluşturur. RO algoritmasında her düğüm, o düğümde rastgele seçilen bir tahmin alt kümesi arasından en iyisini seçerek bölünür (Breiman, 2001).

4. Uygulama

Bu çalışmada 768 binaya ait veri seti daha önce benzer çalışmalarda kullanılmış ve açık erişime sunulmuş bir veri setidir(Tsanas & Xifara, 2012). 768 satır veri içeren bu veri seti 8 bağımsız değişken (X1....X8) ve 2 bağımlı değişkenden (Y1, Y2) oluşmaktadır. Bu veriler 12 adet binanın farklı simülasyonlarından elde edilmiştir. Her bir gözlem ayrı bir bina içindir. Veri setindeki değişkenler ve tanımları sırasıyla tablo 1’de listelenmiştir.

Tablo 1: Değişken Açıklamaları

Değişken	Değişken Açıklama
X1	Nispi Yoğunluk
X2	Yüzey Alanı
X3	Duvar Alanı
X4	Çatı Alanı
X5	Toplam Yükseklik
X6	Yönlendirme
X7	Cam Alanı
X8	Cam Alanı Dağılımı
Y1	Isıtma Yüğü
Y2	Soğutma Yüğü

Veri setinin anlaşılması amacıyla örnek olarak 4 adet binaya ait veriler tablo 2’de görölmektedir.

Tablo 2: Veri seti örneğı

X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y1
0,79	637	343	147	7	2	0	0	28,52	37,73
0,76	661,5	416,5	122,5	7	2	0,1	2	33,16	33,91
0,69	735	294	220,5	3,5	4	0,25	1	12,73	15,48
0,9	563,5	318,5	122,5	7	2	0,4	1	36,47	39,28
0,66	759,5	318,5	220,5	3,5	3	0,4	1	15,23	18,03

Veri setindeki 768 bina verisi önce kayıp veri analizine tabi tutulmuştur. Veri setinde kayıp veri olmadığı görüldükten sonra veriler öznitelik seçimi ile irdelenmiş ve ilgili istatistiksel analizler yapılmıştır.

4.1. Veri Analizi İşlemleri

Modeli daha anlamlı kılmak için yapılan istatistiksel analizlerden biri korelasyon analizidir. Bu çalışmada korelasyon analizini gerçekleştirmek için Spearman korelasyon analizi, SPSS programı kullanılarak yapılmış ve tablo 3'teki sonuçlar elde edilmiştir.

Tablo 3: Korelasyon Değerleri

GİRİŞ DEĞİŞKENLERİ	Y1	Y2
X1	0,622	0,651
X2	-0,622	-0,651
X3	0,471	0,416
X4	-0,804	-0,803
X5	0,861	0,865
X6	-0,004	0,018
X7	0,323	0,289
X8	0,068	0,046

Tablo 3'ten görüldüğü gibi X6 ve X8 değişkenlerinin korelasyon değerleri diğer değişkenlere oranla önemli derecede düşük kalmaktadır. Korelasyon analizi sonrasında p testi değerlerine bakılarak korelasyon değerlerinin yanı sıra anlamlılık değerleri de değerlendirilmiş, sonuçlara göre model içerisinde anlamsız gözükten ve çıkış değerleriyle ilişki seviyeleri düşük olan değişkenler veri setinden çıkarılmıştır.

Sekiz girdi değişkeni ve iki çıktı değişkeniyle oluşturulan regresyon modeli üzerinden yapılan p -testi sonuçları tablo 4'te görülmektedir.

Tablo 4: P değerleri

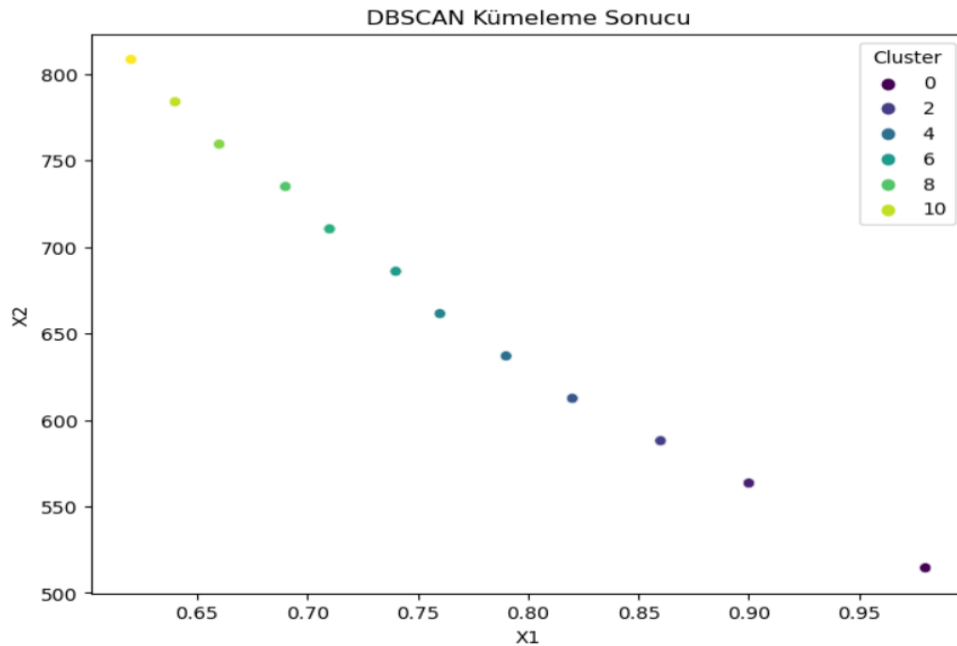
GİRİŞ DEĞİŞKENLERİ	Y1	Y2
X1	0,000	0,000
X2	0,000	0,000
X3	0,000	0,000
X4	0,000	0,000
X5	0,000	0,000
X6	0,805	0,240
X7	0,000	0,000
X8	0,040	0,594

Tablo 4’te görülen “0,000” değerleri binde birlik sayıya yuvarlandığı için “0” olarak görülmektedir. P değeri 0,05’ten ne kadar düşükse, değışken model içerisinde o derece anlamlıdır. İki analiz üzerinden yorumlama yapılacak olursa, X6 ve X8 değışkenleri hem çıkış değışkenleriyle olan düşük korelasyonları hem de *p* testi sonucu elde edilen anlamlılık değeri ortak paydada göz önüne alındığında, modelden çıkarılacak değışkenler olarak göze çarpmaktadır. Çalışmaya 6 giriş değışkeni ile devam edilmiştir.

4.2. Kümeleme Analizi ve Tahminleme

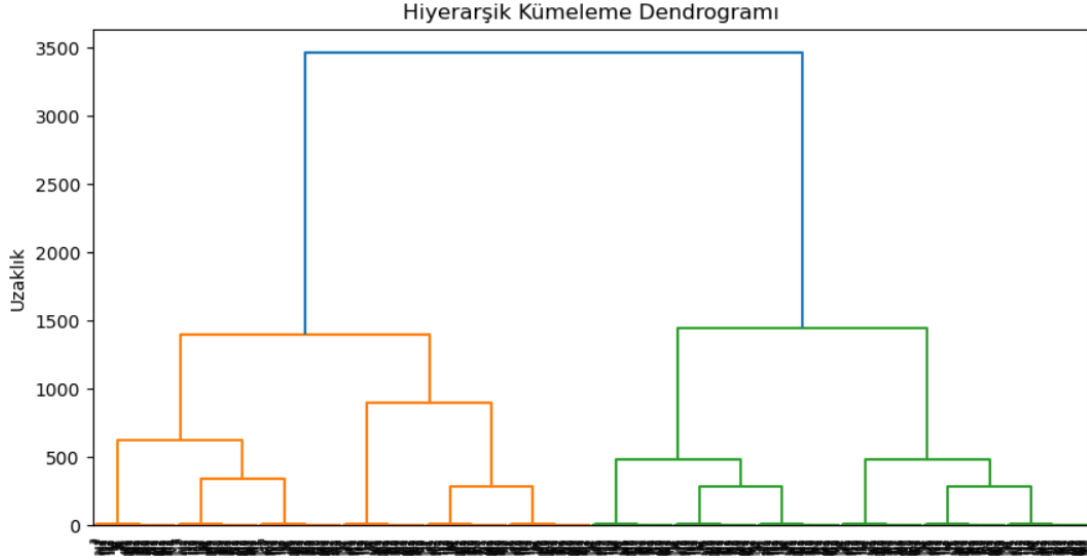
Kümeleme analizi için DBSCAN, hiyerarşik kümeleme ve K-ortalamlar algoritmaları olmak üzere 3 farklı algoritma kullanılmıştır. DBSCAN algoritması optimum küme sayısını kendi belirler. Bu kümeleme metodunda küme sayısının manuel belirlenmemesi sebebiyle küme sayıları değıştirilerek silhoutte değeriğine bakılmamıştır. Algoritma en doğru küme sayısını kendi belirlemektedir. DBSCAN algoritmasıyla yapılan kümeleme sonucu elde edilen kümeleme diyagramı şekil 4’te gösterilmiştir. Algoritma veriyi 12 kümeye ayırmış ve bu kümelemeyi farklı renklerle grafikte gösterilmiştir. Algoritmanın değışkenleri olan Eps değeri, $eps_value = 0.2 + (n_clusters * 0.1)$ formülüyle dinamik olarak kullanılmıştır.

MinPts değeri ise literatürde çok yüksek olmayan boyutlu verilerde genelde 4 veya 5 seçildiği için 5 olarak alınmıştır. (Ankerst vd., 1999)



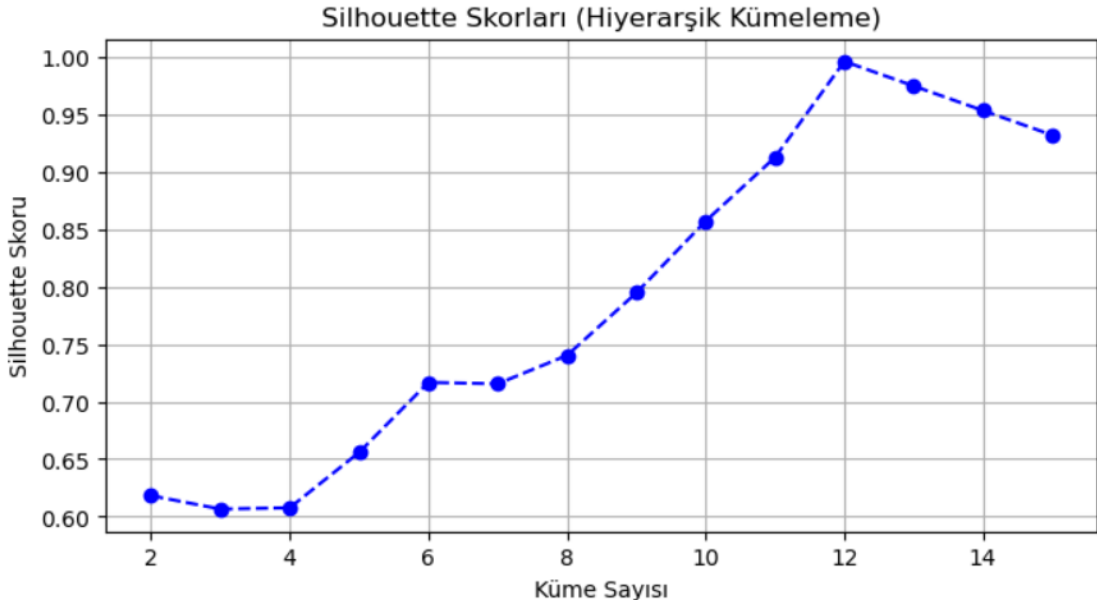
Şekil.4 DBSCAN Algoritması Kümeleme Diagramı

Hiyerarşik kümeleme metodu olarak ise Aglomeratif metod kullanılmıştır. Kümeleme sonucu oluşan dendrogram şekil 5’te görülmektedir. Dendrogram 12 kümeye ayrılıncaya kadar ilerlemiştir.



Şekil.5 Aglomeratif Kümeleme Diagramı

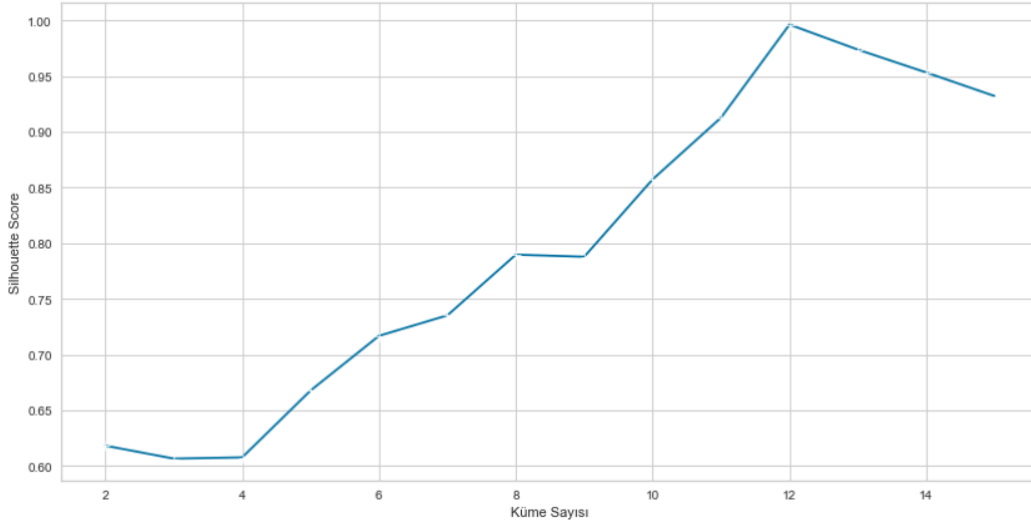
Aglomeratif metotta küme sayısı 2-15 arasında değişen sayılar denenmiş ve elde edilen silhoutte skorlarına göre şekil 6’daki grafik elde edilmiştir.



Şekil.6 Aglomeratif Kümeleme Silhoutte Değerleri

Hiyerarşik kümelemede oluğu gibi K-ortalamar metodunda da küme sayıları 2 ile 15 arasında seçilerek yapılan kümeleme analizinde silhoutte skorları elde edilmiş, bu değerler üzerinden uygun k

değeri seçilmiştir. Ayrıca dirsek metoduyla da optimum k sayısı belirlenmiş ve elde edilen bu 2 farklı k değeri kümeleme analizinde kullanılmıştır.



Şekil 7: Küme Sayısına Göre Silhouette Skor Değerleri

Şekil 6 ve şekil 7'den anlaşıldığı üzere silhouette skoru $k=12$ 'de en yüksek değere ulaşmış sonrasında düşüşe geçmiştir. Silhouette skorları dikkate alındığında kümeleme işleminde küme sayısı 12 olarak seçilir ve kümeleme işlemi bu değer üzerinden gerçekleştirilir. Küme sayısı 12 olarak seçildiğinde tablo 5'te örnek olarak hangi binaların hangi kümelere atandığı görülebilmektedir.

Tablo 5: Silhouette Skorlarına Göre Örnek Kümeleme Tablosu

X1	X2	X3	X4	X5	X7	Küme No
0,62	808,5	367,5	220,5	3,5	0	8
0,71	710,5	269,5	220,5	3,5	0,1	11
0,98	514,5	294	110,25	7	0,1	5
0,86	588	294	147	7	0,1	7
0,9	563,5	318,5	122,5	7	0	1

Kümeleme analizi sonucunda veriler her bir kümede 64 bina verisi barındıracak şekilde 12 kümeye ayrılmıştır. Ardından, 12 küme içerisinde ayrı ayrı LR, GA, XGA, CB, KNN, KA ve RO algoritmaları kullanılarak küme bazında binaların ısıtma ve soğutma yükleri tahmin edilmiştir. Sonuçların performansını kıyaslamak amacıyla açıklayıcılık katsayısı R^2 ve ortalama kare hatası MSE (mean squared error) ölçütleri kullanılmıştır.

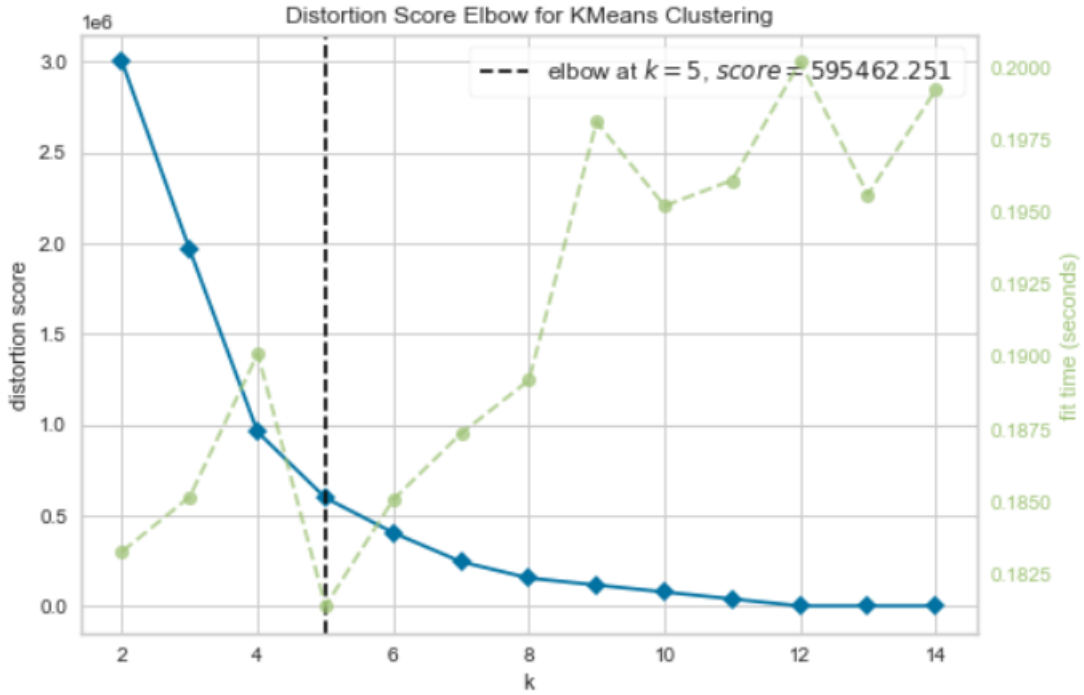
Tablo 6'da 12 kümenin her birinde yapılan tahminleme işlemlerinde makine öğrenmesi algoritmalarının performans değerleri görülebilmektedir. R^2 değeri 1'e ne kadar yakınsa tahmin performansı o derece başarılı, 0 değerine yaklaştıkça ve negatif değerler aldıkça o derece başarısız

olmaktadır. MSE deęerleri ise 0'a ne kadar yakınsa hata deęeri o kadar dūşük olur. Örneęin; 0 no'lu kümede en iyi tahmin performansı gösteren algoritmalar ısıtma yükünde 0,993 R^2 ve 0,028 MSE deęerleriyle GA ve KA ve CB algoritmaları; soęutma yükünde ise 0,913 R^2 ve 0,17 MSE deęeriyle GA, KA ve CB algoritmalarıdır.

Tablo 6: 12 Küme Bazında Makine Öğrenmesi Algoritmalarının Tahmin Performansı Değerleri

Algo- ritma	Enerji Yükü	0		1		2		3		4		5		6		7		8		9		10		11	
		R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE
LR	Isıtma Yükü	0,861	0,613	0,839	2,806	0,914	1,743	0,874	0,651	0,788	2,810	0,912	0,097	0,910	0,500	0,858	1,765	0,822	0,861	0,863	1,560	0,893	0,536	0,831	0,811
	Sogutma Yükü	0,836	0,322	0,584	5,030	0,934	0,609	0,729	0,613	0,221	13,991	0,973	0,393	0,922	0,233	0,524	6,125	0,909	0,197	0,676	2,441	0,933	0,233	0,729	0,683
GA	Isıtma Yükü	0,993	0,028	0,963	0,644	0,984	0,311	0,979	0,109	0,960	0,520	0,997	0,054	0,993	0,035	0,975	0,305	0,924	0,365	0,967	0,374	0,996	0,017	0,953	0,222
	Sogutma Yükü	0,913	0,170	0,528	5,720	0,970	0,270	0,801	0,450	0,053	17,013	0,991	0,121	0,965	0,105	0,370	8,103	0,956	0,095	0,652	2,625	0,935	0,224	0,807	0,486
XGA	Isıtma Yükü	0,942	0,252	0,875	2,179	0,897	2,088	0,899	0,522	0,804	2,593	0,952	1,137	0,960	0,220	0,873	1,572	0,934	0,316	0,879	1,369	0,948	0,260	0,916	0,403
	Sogutma Yükü	0,877	0,240	0,552	5,426	0,781	2,022	0,664	0,759	0,099	16,182	0,936	0,926	0,895	0,316	0,370	8,099	0,742	0,562	0,630	2,791	0,878	0,424	0,756	0,616
CB	Isıtma Yükü	0,993	0,028	0,963	0,644	0,984	0,311	0,979	0,109	0,960	0,520	0,997	0,054	0,993	0,035	0,975	0,305	0,924	0,365	0,967	0,374	0,996	0,017	0,953	0,222
	Sogutma Yükü	0,913	0,170	0,528	5,720	0,970	0,270	0,801	0,450	0,053	17,013	0,991	0,121	0,965	0,105	0,370	8,103	0,956	0,095	0,652	2,625	0,935	0,224	0,807	0,486
KNN	Isıtma Yükü	0,941	0,259	0,924	1,319	0,952	0,964	0,939	0,314	0,879	1,605	0,958	1,001	0,954	0,255	0,940	0,743	0,869	0,630	0,921	0,895	0,952	0,242	0,901	0,475
	Sogutma Yükü	0,816	0,361	0,533	5,655	0,944	0,515	0,750	0,566	0,010	18,147	0,970	0,429	0,925	0,224	0,320	8,749	0,909	0,196	0,645	2,680	0,925	0,259	0,714	0,720
KA	Isıtma Yükü	0,993	0,028	0,963	0,644	0,984	0,311	0,979	0,109	0,960	0,520	0,997	0,054	0,993	0,035	0,975	0,305	0,924	0,365	0,967	0,374	0,996	0,017	0,953	0,222
	Sogutma Yükü	0,913	0,170	0,528	5,720	0,970	0,270	0,801	0,450	0,053	17,013	0,991	0,121	0,965	0,105	0,370	8,103	0,956	0,095	0,652	2,625	0,935	0,224	0,807	0,486
RO	Isıtma Yükü	0,991	0,039	0,963	0,645	0,985	0,301	0,979	0,104	0,955	0,586	0,996	0,075	0,993	0,036	0,975	0,301	0,926	0,357	0,967	0,368	0,996	0,016	0,952	0,227
	Sogutma Yükü	0,910	0,177	0,539	5,582	0,971	0,267	0,801	0,450	0,070	16,651	0,991	0,123	0,963	0,111	0,372	8,080	0,954	0,099	0,659	2,576	0,936	0,220	0,805	0,491

En uygun k değeriyle kümeleme yapabilmek için silhoutte skorları ve dirsek metodu olmak üzere iki farklı yöntem kullanılmaktadır. Bu iki yöntemin uygulanması sonucunda 2 farklı k değeriyle kümeler oluşturulup, bu kümelerde gerçekleştirilen tahminleme performansları sonucu kıyaslanabilecektir. Şekil 8'deki grafiğe bakıldığında dirsek yöntemi kullanıldığında 5 küme sayısı (grafiğin kırılım noktası) en uygun k değeri olarak görülmektedir. DBSCAN algoritması küme sayısını kendi belirlediği için elbow metodunda çıkan değer hiyerarşik kümeleme metodu ve K-ortalamlar metoduyla denenmiştir. Bu denemeler sonucunda $k=5$ değeri kullanılarak DBSCAN algoritmasıyla kümeleme yapıldığında elde edilen silhoutte skoru 0,655 iken K-ortalamlar ile kümelendiğinde 0,667 değeri elde edilmiştir. Değerler birbirine yakın olsa da K-ortalamlar daha iyi sonuç verdiği için kümeleme işlemi K-ortalamlar ile yapılmıştır.



Şekil 8: Dirsek yöntemi Grafiği

Bu aşamadan sonra kümeleme analizi işlemine geçilmiştir. $K=5$ küme sayısı ve K-ortalamlar yöntemi ile yapılan kümeleme sonucunda oluşan kümelerdeki bina sayıları tablo 7'de görülmektedir.

Tablo 7:5 Kümeye Ait Bina Sayıları

Küme No	Bina Sayısı
0	192
1	128
2	192
3	192
4	64

Kümeleme yapılan binalara ait 6 değişkenle birlikte binaların ait oldukları kümelere ait veriler örnek olarak 5 bina için tablo 8’de gösterilmektedir.

Tablo 8: Dirsek Yöntemine Göre Örnek Kümeleme Tablosu

X1	X2	X3	X4	X5	X7	Küme No
0,98	514,5	294	110,25	7	0	1
0,69	735	294	220,5	3,5	0,1	3
0,82	612,5	318,5	147	7	0,25	2
0,71	710,5	269,5	220,5	3,5	0,25	3
0,76	661,5	416,5	122,5	7	0,4	4

Bu aşamadan sonra beş kümenin her birinde ayrı ayrı regresyon tabanlı makine öğrenmesi algoritmalarıyla binaların Y1 ve Y2 ısıtma ve soğutma yükleri, sırasıyla çıkış değerleri olarak alınarak tahmin işlemleri yapılmış, kullanılan algoritmaların performansları karşılaştırılmıştır. Tablo 9’den görüleceği üzere oluşan 5 küme için de ayrı ayrı LR, GA, XGA, CB, KNN, KA ve RO algoritmalarının tahmin performansları R^2 ve MSE değerleri üzerinden karşılaştırılmıştır. Örnek vermek gerekirse 1 no’lu kümede ısıtma yükünde 0,993 R^2 ve 0,117 MSE değerleriyle en iyi performans gösteren algoritmalar GA, KA ve CB algoritmalarıdır. Soğutma yükünde ise 0,868 R^2 ve 1,888 MSE değerleriyle LR algoritmasıdır.

Tablo 9: 5 Küme Bazında Makine Öğrenmesi Algoritmalarının Tahmin Performansı Değerleri

Algoritmalar	Enerji Yüğü	0		1		2		3		4	
		R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE	R ²	MSE
LR	Isıtma Yüğü	0,358	3,398	0,924	1,324	0,978	1,123	0,813	0,783	0,914	1,743
	Soğutma Yüğü	-0,170	6,425	0,868	1,888	0,790	8,856	0,911	0,199	0,934	0,609
GA	Isıtma Yüğü	0,976	0,122	0,993	0,117	0,990	0,486	0,939	0,255	0,984	0,311
	Soğutma Yüğü	0,972	0,149	0,845	2,213	0,795	8,623	0,948	0,116	0,970	0,270
XGA	Isıtma Yüğü	0,909	0,478	0,934	1,158	0,949	2,610	0,872	0,534	0,897	2,088
	Soğutma Yüğü	0,895	0,574	0,824	2,517	0,818	7,645	0,788	0,479	0,781	2,022
CB	Isıtma Yüğü	0,977	0,116	0,993	0,117	0,990	0,490	0,939	0,255	0,984	0,311
	Soğutma Yüğü	0,973	0,147	0,845	2,213	0,795	8,624	0,948	0,116	0,970	0,270
KNN	Isıtma Yüğü	0,936	0,334	0,972	0,483	0,983	0,876	0,901	0,413	0,952	0,964
	Soğutma Yüğü	0,967	0,176	0,834	2,378	0,748	10,637	0,926	0,165	0,944	0,515
KA	Isıtma Yüğü	0,977	0,116	0,993	0,117	0,990	0,493	0,939	0,255	0,984	0,311
	Soğutma Yüğü	0,973	0,147	0,845	2,213	0,795	8,624	0,948	0,116	0,970	0,270

Silhouette skorları ve dirsek yöntemine göre kümelenen bina verilerinden elde edilen ısıtma ve soğutma yüklerinin tahmin edilmesinin yanı sıra kümelenmemiş veride aynı algoritmalarla aynı işlemler tekrarlanmıştır. Algoritmalarından elde edilen R^2 ve MSE değerleri tablo 10’da görülmektedir.

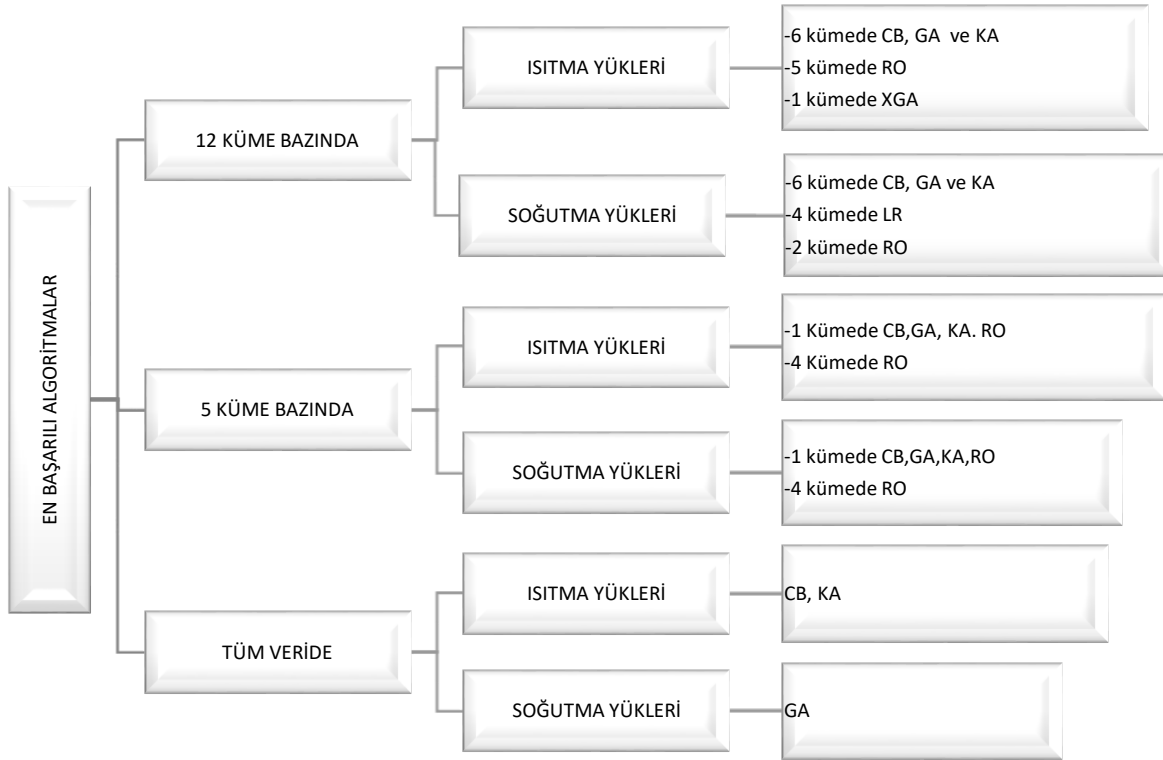
Tablo 10: Kümelenmemiş Veride Makine Öğrenmesi Algoritmalarının Tahmin Performans Değerleri

Algoritma	Enerji Yükleri	R^2	MSE
LR	Isıtma Yüğü	0,911	9,221
	Soğutma Yüğü	0,892	9,931
GA	Isıtma Yüğü	0,997	0,270
	Soğutma Yüğü	0,967	3,044
XGA	Isıtma Yüğü	0,990	0,951
	Soğutma Yüğü	0,956	4,001
CB	Isıtma Yüğü	0,997	0,251
	Soğutma Yüğü	0,967	3,055
KNN	Isıtma Yüğü	0,992	0,826
	Soğutma Yüğü	0,951	4,474
KA	Isıtma Yüğü	0,997	0,251
	Soğutma Yüğü	0,967	3,055
RO	Isıtma Yüğü	0,997	0,252
	Soğutma Yüğü	0,967	3,056

Tablo 10’da sunulan sonuçlara göre algoritmaların R^2 değerlerinin 1’e oldukça yakın oldukları yani oldukça iyi bir performans gösterdikleri görülmektedir. Isıtma yüğü tahmininde en iyi performans gösteren algoritmalar 0,997 R^2 ve 0,251 MSE değerleriyle CB ve KA algoritmalarıdır. Soğutma yüğü tahmininde ise 0,967 R^2 ve 3,044 MSE değerleriyle GA algoritmasıdır. Tüm makine öğrenmesi algoritmalarının R^2 değerleri 0,85 ve üzeri olduğu göz önüne alındığında iyi performans gösterdikleri görülebilir.

5. Bulgular ve Tartışma

Çalışmada yapılan kümeleme analizinde, küme sayısı olarak hem dirsek yöntemine göre seçilen optimum küme sayısı hem de silhouette skorlarına göre seçilen optimum küme sayısı kullanılarak kümeleme yapılmıştır. Makine öğrenmesi algoritmalarının kümelenmemiş veride ve kümelenen verilerde gerçekleştirdiği tahmin performansları şekil 6’da diyagram olarak sunulmuştur. CB, GA ve KA algoritmalarının en başarılı çıktığı analizlerde RO algoritması binde birlik derecesindeki farklar ile daha düşük bir başarı yakalamıştır, yani tahmin performansı CB, GA ve KA algoritmaları ile hemen hemen aynıdır.



Şekil 6.En iyi performans sergileyen algoritmalar

Dirsek yöntemine göre yapılan kümeleme analizi sonucunda elde edilen kümelerde gerçekleştirilen tahminleme sonuçlarının R^2 ve MSE değerleri, silhoutte skorlarına göre elde edilen kümelerde yapılan tahminleme sonuçlarının R^2 ve MSE değerlerinden daha iyi elde edilmiştir. Bunun yanında kümelenen verilerde tüm algoritmalar için elde edilen R^2 ve MSE değerlerinin kümelenmemiş veride elde edilen R^2 ve MSE değerlerinden daha iyi olduğunu söylemek mümkün değildir. Hatta kümelenmemiş verideki R^2 ve MSE değerleri tablo 6, tablo 9 ve tablo 10 karşılaştırması yapıldığında kümelenmiş verilerden elde edilenlere göre daha yüksek R^2 ve daha düşük MSE değerleri elde edilmiştir.

Çalışmada tahminleme öncesi kümeleme yapılmasının sebebi kümelemenin oluşturduğu homojen yapılar sebebiyle tahmin performansının artmasının beklentisidir. Örnek olarak incelersek LR algoritmasının kümelenmemiş veride uygulanmasıyla ısıtma ve soğutma yükleri tahmininde elde edilen R^2 ve MSE değerleri sırasıyla 0,911, 9,221 ve 0,892 , 9,931'dir. Veriler k=5 değeriyle kümelendiğinde 4 no'lu kümedeki ısıtma ve soğutma yüklerinin tahmin sonuçlarına ait R^2 ve MSE değerleri sırasıyla 0,914 , 1,743 ve 0,934 , 0,609'dur. Veriler k=12 ile kümelendiğinde ise 5 no'lu kümede ısıtma ve soğutma yüklerinin tahmin sonuçlarına ait R^2 ve MSE değerleri sırasıyla 0,912 , 0,097 ve 0,973 , 0,393'tür. Bu algoritma için bazı kümelenmiş verilerden elde edilen tahmin performansının kümelenmemiş verilerden daha iyi olduğu görülebilmektedir. Tüm algoritmaları göz

önünde bulundurduğumuzda ise, kümelemenin tahmin performansına belirgin bir olumlu katkısı olduğu konusunda genelleme yapmak mümkün olmamaktadır. Bunun sebebi veri setinin küçülmesi sebebiyle algoritmaların öğrenmelerinin zorlaşması olabilir. Tahminleme öncesi kümeleme, modelin karmaşıklığının azaltılması ve öğrenme hızının artırılması için çok daha büyük bir veri setinde gerçekleştirildiği takdirde daha olumlu sonuçlar alınabileceği beklenebilir.

6. Sonuçlar ve Öneriler

Enerji verimliliği günümüz dünyasında tüketimin de artmasıyla çok önemli bir noktaya gelmiş, enerji verimliliği ile ilgili alınan önlemler enerji tüketiminin azaltılmasına, kaynakların verimli kullanılmasına ve buna bağlı olarak karbon emisyonlarının düşürülmesine olanak sağlamıştır. Enerji tüketim ve enerji performans tahmini çalışmaları da bu önlemler kapsamında yapılan çalışmalara öngörü sağlayan ve yönlendiren çalışmalar olarak karşımıza çıkmaktadır. Bu makalede binaların enerji performansı üzerinden bu çalışmalar ileriye taşınmış olup aynı zamanda enerji tüketim değerlerine hangi girdi değerlerinin (binaların fiziksel özellikleri) daha çok etki ettiği de araştırılmıştır. Böylece bir binanın tasarım aşamasında enerji performansı önceden görülerek buna göre yapılacak tasarımlar yönlendirilebilecektir. Bu durum enerji kimlik belgesinin edinilebilmesi aşamasında da oldukça kolaylık sağlayacaktır. Yöntem olarak çalışmanın getirdiği yenilik ise kümeleme analizinin, makine öğrenmesi algoritmalarının enerji yükleri tahmin performansına etkisinin araştırılması ve tahminin tek algoritma ile değil, yedi adet makine öğrenmesi algoritması kıyaslanarak elde edilmiş olmasıdır. Kümeleme işlemi yapılırken üç farklı kümeleme metodunun kullanılmasının yanında iki farklı optimum k sayısı dikkate alınması da yöntemin literatürden farkını ortaya koymaktadır. Geleceğe yönelik çalışmalarda enerji verimliliği tahmininde derin öğrenme algoritmaları kullanılabilir. Eğer konuyla ilgili bir zaman serisi verisi elde edilebilirse zaman bağımlılıklarını yakalamada başarılı olduğu için Uzun Kısa Süreli Bellek algoritması derin öğrenme algoritması olarak kullanılabilir. Eğer termal görüntüler veya bina cephe analizleri gibi görsel veriler elde edilebilirse Konvolüsyonel Sinir Ağları algoritması kullanılabilir. Ayrıca aynı verilerle regresyon tabanlı modeller yerine sınıflandırma işlemi yapılabilir ve sınıflandırma işlemi yapılmadan önce verilerin kümelenmesinin, sınıflandırma performansına etkisi araştırılabilir. Bir binanın enerji kimlik belgesinde belirtilen sınıflardan hangisine ait olduğu bilgisi elde edilir ve ilgili veri seti oluşturulabilirse önceden bu sınıflar öngörülebilir.

Kaynaklar

- Abbasimehr, H., Paki, R., & Bahrini, A. (2023). A novel XGBoost-based featurization approach to forecast renewable energy consumption with deep learning models. *Sustainable Computing: Informatics and Systems*, 38, 100863. <https://doi.org/10.1016/j.suscom.2023.100863>
- Ahmad, T., Chen, H., Huang, R., Yabin, G., Wang, J., Shair, J., Azeem Akram, H. M., Hassnain Mohsan, S. A., & Kazim, M. (2018). Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. *Energy*, 158, 17-32. <https://doi.org/10.1016/j.energy.2018.05.169>
- Albayrak, Y. doç dr ali S., & Yilmaz, Ö. gör şebnem K. (2009). VERİ MADENCİLİĞİ: KARAR AĞACI ALGORİTMALARI VE İMKB VERİLERİ ÜZERİNE BİR UYGULAMA. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1), Article 1.
- AlMahamid, F., & Grolinger, K. (2022). Agglomerative Hierarchical Clustering with Dynamic Time Warping for Household Load Curve Clustering. *2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 241-247. <https://doi.org/10.1109/CCECE49351.2022.9918481>
- Al-Rakhami, M., Gumaei, A., Alsanad, A., Alamri, A., & Hassan, M. M. (2019). An Ensemble Learning Approach for Accurate Energy Load Prediction in Residential Buildings. *IEEE Access*, 7, 48328-48338. <https://doi.org/10.1109/ACCESS.2019.2909470>
- An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression*. (2024).
- Aqlan, F., Ahmed, A., Srihari, K., & Khasawneh, M. (2014). Integrating Artificial Neural Networks and Cluster Analysis to Assess Energy Efficiency of Buildings. *İçinde IIE Annual Conference and Expo 2014*.
- Aurangzeb, K. (2024). DBSCAN-based energy users clustering for performance enhancement of deep learning model. *Journal of Intelligent & Fuzzy Systems*, 46(3), 5555-5573. <https://doi.org/10.3233/JIFS-235873>
- Bakkellund, D. (2022). Order preserving hierarchical agglomerative clustering. *Machine Learning*, 111(5), 1851-1901. <https://doi.org/10.1007/s10994-021-06125-0>
- Bektas Ekici, B., & Aksoy, U. T. (2011). Prediction of building energy needs in early stage of design by using ANFIS. *Expert Systems with Applications*, 38(5), 5352-5358. <https://doi.org/10.1016/j.eswa.2010.10.021>
- BİNALARDA ENERJİ PERFORMANSI YÖNETMELİĞİ*. (t.y.).
- Birkess, D., & Dodge, D. Y. (2011). *Alternative Methods of Regression*. John Wiley & Sons.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- Canbay, P., & Taş, H. (2022). Yapıların Isıtma ve Soğutma Yükünün Yapay Zeka ile Tahmini. *International Journal of Pure and Applied Sciences*, 8(2), 478-489. <https://doi.org/10.29132/ijpas.1166227>
- cevreselgostergeler.csb.gov.tr. (t.y.). - Çevresel Göstergeler. Geliş tarihi 18 Mart 2024, gönderen <https://cevreselgostergeler.csb.gov.tr/sektorlere-gore-nihai-enerji-tuketimi-i-85804>
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794. <https://doi.org/10.1145/2939672.2939785>
- Cheng, L. K., Selamat, A., Zabil, M. H. M., Selamat, M. H., Alias, R. A., Puteh, F., Mohamed, F., & Krejcar, O. (2019). Comparing the Accuracy of Hierarchical Agglomerative and K-means Clustering on Mobile Augmented Reality Usability Metrics. *2019 IEEE Conference on Big Data and Analytics (ICBDA)*, 34-40. <https://doi.org/10.1109/ICBDA47563.2019.8987044>
- Cheng, M.-Y., & Cao, M.-T. (2014). Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines. *Applied Soft Computing*, 22, 178-188. <https://doi.org/10.1016/j.asoc.2014.05.015>
- Chou, J.-S., & Bui, D.-K. (2014). Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings*, 82, 437-446. <https://doi.org/10.1016/j.enbuild.2014.07.036>
- Cook, R. D. (2015). Linear Hypothesis: Regression (Graphics). İçinde J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (Second Edition, ss. 157-161). Elsevier. <https://doi.org/10.1016/B978-0-08-097086-8.42141-0>
- Coşkun, E., Gündoğan, E., Kaya, M., & Alhajj, R. (2021). Veri madenciliği yöntemleri kullanarak yoğun bakım ünitesindeki hastaların sınıflandırması. *Computer Science*. <https://doi.org/10.53070/bbd.990718>

- DiLber, B., & Özdemir, A. F. (2022). Estimating Heating and Cooling Loads of Energy Efficient Buildings with Tidy Models. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, 1091-1106. <https://doi.org/10.29130/dubited.939933>
- Doğan, M., & Seçme, D. (2018). *Çevre Dostu Binalar ve Yeşil Bina Sertifika Sistemleri*.
- Dong, G., Jin, Y., Wang, S., Li, W., Tao, Z., & Guo, S. (2019). DB-Kmeans: An Intrusion Detection Algorithm Based on DBSCAN and K-means. *2019 20th Asia-Pacific Network Operations and Management Symposium (APNOMS)*, 1-4. <https://doi.org/10.23919/APNOMS.2019.8892910>
- Duarte, G. R., Capriles, P., & Lemonge, A. (t.y.). *Prediction of energy load of buildings using machine learning methods*.
- Economopoulou, E.-E., Vonitsanos, G., Mylonas, P., & Kanavos, A. (2024). Efficient Energy Disaggregation Using DBSCAN: A Novel Approach for Enhanced Energy Management. İçinde I. Maglogiannis, L. Iliadis, I. Karydis, A. Papaleonidas, & I. Chochliouros (Ed.), *Artificial Intelligence Applications and Innovations. AIAI 2024 IFIP WG 12.5 International Workshops* (ss. 236-249). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-63227-3_16
- Enerji Verimliliği—T.C. Enerji ve Tabii Kaynaklar Bakanlığı. (t.y.). Geliş tarihi 18 Mart 2024, gönderen <https://enerji.gov.tr/enerji-verimliliği>
- Ercan, U. (2022). Havayolu Taşımacılığında Müşteri Memnuniyetinin Makine Öğrenmesi Yöntemleri İle Belirlenmesi. *Alanya Akademik Bakış*. <https://doi.org/10.29023/alanyaakademik.1095574>
- Erdede, S. B., Erdede, B., & Bektaş, S. (2014). *SÜRDÜRÜLEBİLİR YEŞİL BİNALAR VE SERTİFİKA SİSTEMLERİNİN DEĞERLENDİRİLMESİ*.
- Fernández, A., & Gómez, S. (2008). Solving Non-Uniqueness in Agglomerative Hierarchical Clustering Using Multidendrograms. *Journal of Classification*, 25(1), 43-65. <https://doi.org/10.1007/s00357-008-9004-x>
- Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification [J]. *Journal of Machine Learning Research - JMLR*, 3.
- Garikapati, P., Balamurugan, K., Latchoumi, T. P., & Malkapuram, R. (2021). A Cluster-Profile Comparative Study on Machining AlSi7/63% of SiC Hybrid Composite Using Agglomerative Hierarchical Clustering and K-Means. *Silicon*, 13(4), 961-972. <https://doi.org/10.1007/s12633-020-00447-9>
- Guo, H., Liu, X., & Zhang, Q. (2024). Identifying daily water consumption patterns based on K-means Clustering, Agglomerative Hierarchical Clustering, and Spectral Clustering algorithms. *AQUA-WATER INFRASTRUCTURE ECOSYSTEMS AND SOCIETY*, 73(5), 870-887. <https://doi.org/10.2166/aqua.2024.294>
- Gümüşçü, A., Tenekeci, M. E., Beşli, N., İlhan, M. A., Aslan, E., & Abamor, S. (2018). Yapılarda Isıtma ve Soğutma Yüğü Sınıfların Makine Öğrenme Yöntemleri ile Belirlenmesi. *Harran Üniversitesi Mühendislik Dergisi*, 3(2), Article 2.
- Güzelkokar, O. (2019). *Mevcut Yapıların Sürdürülebilir Yeşil Binalara Dönüştürülmesi*.
- Jahan, I., & Sanam, T. F. (2024). A comprehensive framework for customer retention in E-commerce using machine learning based on churn prediction, customer segmentation, and recommendation. *Electronic Commerce Research*. <https://doi.org/10.1007/s10660-024-09936-0>
- Kavaklioglu, K. (2018). Robust modeling of heating and cooling loads using partial least squares towards efficient residential building design. *Journal of Building Engineering*, 18, 467-475. <https://doi.org/10.1016/j.jobe.2018.04.018>
- Khadka, S., Wagle, A., Dhakal, B., Gautam, R., Nepal, T., Shrestha, A., & Gonzalez-Longatt, F. (2024). Agglomerative Hierarchical Clustering Methodology to Restore Power System considering Reactive Power Balance and Stability Factor Analysis. *International Transactions on Electrical Energy Systems*, 2024(1), 8856625. <https://doi.org/10.1155/2024/8856625>
- Khalil, A. J., Barhoom, A. M., Abu-Nasser, B. S., Musleh, M. M., & Abu-Naser, S. S. (2019). *Energy Efficiency Prediction using Artificial Neural Network*. 3(9).
- Khan, A. M., & Osińska, M. (2023). Comparing forecasting accuracy of selected grey and time series models based on energy consumption in Brazil and India. *Expert Systems with Applications*, 212, 118840. <https://doi.org/10.1016/j.eswa.2022.118840>
- Kumar, S., Pal, S. K., & Singh, R. P. (2018). Intra ELM variants ensemble based model to predict energy performance in residential buildings. *Sustainable Energy, Grids and Networks*, 16, 177-187. <https://doi.org/10.1016/j.segan.2018.07.001>
- Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, 36(2), 451-461. [https://doi.org/10.1016/S0031-3203\(02\)00060-2](https://doi.org/10.1016/S0031-3203(02)00060-2)

- Lu, C., Li, S., Reddy Penaka, S., & Olofsson, T. (2023). Automated machine learning-based framework of heating and cooling load prediction for quick residential building design. *Energy*, 274, 127334. <https://doi.org/10.1016/j.energy.2023.127334>
- Mehmet; Subaşı, M. (2005). Bilimsel Çalışmalarda P-Değerinin Rapor Edilmesi (P<0.01?, P0.05?). *Tarım Bilimleri Dergisi*, 11(4), 1. https://doi.org/10.1501/Tarimbil_0000000565
- Miraftabzadeh, S. M., Colombo, C. G., Longo, M., & Foiadelli, F. (2023). K-Means and Alternative Clustering Methods in Modern Power Systems. *IEEE Access*, 11, 119596-119633. <https://doi.org/10.1109/ACCESS.2023.3327640>
- Moradzadeh, A., Mansour-Saatloo, A., Mohammadi-Ivatloo, B., & Anvari-Moghaddam, A. (2020). Performance Evaluation of Two Machine Learning Techniques in Heating and Cooling Loads Forecasting of Residential Buildings. *Applied Sciences*, 10(11), 3829. <https://doi.org/10.3390/app10113829>
- Naji, S., Keivani, A., Shamshirband, S., Alengaram, U. J., Jumaat, M. Z., Mansor, Z., & Lee, M. (2016). Estimating building energy consumption using extreme learning machine method. *Energy*, 97, 506-516. <https://doi.org/10.1016/j.energy.2015.11.037>
- Naji, S., Shamshirband, S., Bassar, H., Keivani, A., Alengaram, U. J., Jumaat, M. Z., & Petković, D. (2016). Application of adaptive neuro-fuzzy methodology for estimating building energy consumption. *Renewable and Sustainable Energy Reviews*, 53, 1520-1528. <https://doi.org/10.1016/j.rser.2015.09.062>
- Namli, E., & Yücel, M. (2018). YAPAY ZEKÂ MODELLERİ İLE BETONARME YAPILARA AİT ENERJİ PERFORMANS SINIFLARININ TAHMİNİ. *Uludağ University Journal of The Faculty of Engineering*, 22(3), 325-346. <https://doi.org/10.17482/uumfd.332320>
- Nilashi, M., Dalvi-Esfahani, M., Ibrahim, O., Bagherifard, K., Mardani, A., & Zakuan, N. (2017). A soft computing method for the prediction of energy performance of residential buildings. *Measurement*, 109, 268-280. <https://doi.org/10.1016/j.measurement.2017.05.048>
- Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Machine learning for energy performance prediction at the design stage of buildings. *Energy for Sustainable Development*, 66, 12-25. <https://doi.org/10.1016/j.esd.2021.11.002>
- Peker, M., Özkaraca, O., & Kesimal, B. (2017). Enerji Tasarruflu Bina Tasarımı İçin Isıtma ve Soğutma Yüklerini Regresyon Tabanlı Makine Öğrenmesi Algoritmaları İle Modelleme. *Bilişim Teknolojileri Dergisi*, 443-449. <https://doi.org/10.17671/gazibtd.310154>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Roy, S. S., Samui, P., Nagtode, I., Jain, H., Shivaramakrishnan, V., & Mohammadi-ivatloo, B. (2020). Forecasting heating and cooling loads of buildings: A comparative performance analysis. *Journal of Ambient Intelligence and Humanized Computing*, 11(3), 1253-1264. <https://doi.org/10.1007/s12652-019-01317-y>
- Sajjad, M., Khan, S. U., Khan, N., Haq, I. U., Ullah, A., Lee, M. Y., & Baik, S. W. (2020). Towards Efficient Building Designing: Heating and Cooling Load Prediction via Multi-Output Model. *Sensors*, 20(22), 6419. <https://doi.org/10.3390/s20226419>
- Sander, J., Ester, M., Kriegel, H.-P., & Xu, X. (1998). Density-Based Clustering in Spatial Databases: The Algorithm GDBSCAN and Its Applications. *Data Mining and Knowledge Discovery*, 2(2), 169-194. <https://doi.org/10.1023/A:1009745219419>
- Sekhar Roy, S., Roy, R., & Balas, V. E. (2018). Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. *Renewable and Sustainable Energy Reviews*, 82, 4256-4268. <https://doi.org/10.1016/j.rser.2017.05.249>
- Seyedzadeh, S., Pour Rahimian, F., Rastogi, P., & Glesk, I. (2019). Tuning machine learning models for prediction of building energy loads. *Sustainable Cities and Society*, 47, 101484. <https://doi.org/10.1016/j.scs.2019.101484>
- Sholahudin, Alam, A. G., Baek, C. I., & Han, H. (2016). Prediction and Analysis of Building Energy Efficiency Using Artificial Neural Network and Design of Experiments. *Applied Mechanics and Materials*, 819, 541-545. <https://doi.org/10.4028/www.scientific.net/AMM.819.541>
- Sonmez, Y., Guvenc, U., Kahraman, H. T., & Yilmaz, C. (2015). A comparative study on novel machine learning algorithms for estimation of energy performance of residential buildings. *2015 3rd International Istanbul Smart Grid Congress and Fair (ICSG)*, 1-7. <https://doi.org/10.1109/SGCF.2015.7354915>
- Şükür, F. Z. (t.y.). *YILDIZ TEKNİK ÜNİVERSİTESİ İKLİM DEĞİŞİKLİĞİ EYLEM PLANI*.

- Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49, 560-567. <https://doi.org/10.1016/j.enbuild.2012.03.003>
- Wang, D., Lu, X., & Rinaldo, A. (t.y.). *DBSCAN: Optimal Rates For Density-Based Cluster Estimation*.
- Yang, F., Gao, P., & Lu, Y. (2016). Evolving Resilient Back-Propagation Algorithm for Energy Efficiency Problem. *MATEC Web of Conferences*, 77, 06016. <https://doi.org/10.1051/mateconf/20167706016>
- Yildiz, M. E., Beyhan, F., & Uçar, M. K. (2021). Enerji Verimli Bina Tasarımı için Yapay Sinir Ağları ile Isıtma Soğutma Yükünün Belirlenmesi. *Academic Perspective Procedia*, 4(1), 91-100. <https://doi.org/10.33793/acperpro.04.01.17>
- Yu, H., Zhong, F., Du, Y., Xie, X., Wang, Y., Zhang, X., & Huang, S. (2023). Short-term cooling and heating loads forecasting of building district energy system based on data-driven models. *Energy and Buildings*, 298, 113513. <https://doi.org/10.1016/j.enbuild.2023.113513>
- Yüce, T., & Kabak, M. (t.y.). *Makine Öğrenmesi Algoritmaları ile Detay Üretim Alanları İçin İş Merkezi Kırılımında Üretim Süresi Tahminleme*.
- Zekić-Sušac, M., Scitovski, R., & Has, A. (2018). Cluster analysis and artificial neural networks in predicting energy efficiency of public buildings as a cost-saving approach. *Croatian Review of Economic, Business and Social Statistics*, 4(2), 57-66. <https://doi.org/10.2478/crebss-2018-0013>
- Zhang, C., Chen, X., Wang, S., Hu, J., Wang, C., & Liu, X. (2021). Using CatBoost algorithm to identify middle-aged and elderly depression, national health and nutrition examination survey 2011–2018. *Psychiatry Research*, 306, 114261. <https://doi.org/10.1016/j.psychres.2021.114261>
- Zurini, M. (2013). Comparative Analysis of K-Means and DbSCAN Algorithms. İçinde C. Boja, L. Batagan, M. Doinea, C. Ciurea, P. Pocatilu, A. Ion, R. Magos, L. Cotfas, A. Velicanu, C. Amancei, M. Andreica, & A. Zamfiroiu (Ed.), *INTERNATIONAL CONFERENCE ON INFORMATICS IN ECONOMY* (ss. 646-651). Bucharest Univ Economic Studies-Ase.