



Prediction of Household Total Energy Expenditures Using Machine Learning Methods

Esma Kesriklioglu^{1,a,*}, Erkan Oktay^{1,b}

¹Department of Econometrics, Faculty of Economics and Administrative Sciences, Atatürk University, Erzurum, Türkiye

*Corresponding Author

ARTICLE INFO

Research Article

Received : 17.08.2022

Accepted : 31.08.2022

Keywords:

Machine Learning

Household Energy Expenditure

Deep Neural Networks

Regression

RapidMiner

ABSTRACT

Depending on the development rate and economic developments of the countries, the range of consumption habits is expanding uncontrollably. Along with the rising living standards in recent years, household energy consumption has been increasing and inevitably, it also causes a significant increase in energy demand. There is a growing concern about the energy use of households, which are a major energy user worldwide. Studies investigating the suitability of machine learning methods for predicting total household energy expenditure are insufficient. To fill this gap, this study presents a comparison of different machine learning methods for regression prediction of total household energy expenditure. This study is aimed to find the machine learning method that provides the best prediction performance. The Household Budget Survey 2019 data set obtained from the Turkish Statistical Institute (TURKSTAT) was used. Household consumption data of 11,521 households were analysed. Under the guidance of the literature review and expert opinion, variables directly or indirectly related to household energy expenditures were created. The prepared data were passed through data preprocessing, modelling, prediction, and evaluation stages using the open source RapidMiner software program. A regression-based machine learning approach was used to predict total household energy expenditure. In the modelling phase, DT, DFNN, KNN, GBT, and RF machine learning methods were used. As a result, DFNN method showed the best performance with the highest R^2 (0.99) and lowest RMSE (5.5). The results of the analysis show that more accurate results are obtained with the DFNN method in the prediction of total household energy expenditures.

Türk Akademik Sosyal Bilimler Araştırma Dergisi, 5(2): 110-118, 2022

Makine Öğrenmesi Yöntemleri Kullanılarak Hanehalkı Toplam Enerji Harcamaları Tahmini

MAKALE BİLGİSİ

ÖZ

Araştırma Makalesi

Geliş : 17.08.2022

Kabul : 31.08.2022

Anahtar Kelimeler:

Makine Öğrenmesi

Hanehalkı Enerji Harcaması

Derin Sinir Ağları

Regresyon

RapidMiner

Ülkelerin kalkınma hızı ve ekonomik gelişmelere bağlı olarak tüketim alışkanlıklarının yelpazesi kontrolsüz bir şekilde genişlemektedir. Son yıllarda yükselen yaşam standartları ile birlikte hanehalkı enerji tüketimi artmakta ve kaçınılmaz olarak, enerji talebinin de önemli ölçüde artmasına neden olmaktadır. Dünya çapında önemli bir enerji kullanıcısı olan hanehalklarının enerji kullanımı hakkında artan bir endişe bulunmaktadır. Hanehalkı toplam enerji harcamalarını tahmin etmek için makine öğrenmesi yöntemlerinin uygunluğunu araştıran çalışmalar yetersizdir. Bu boşluğu gidermek için bu çalışmada, hanehalkı toplam enerji harcamalarının regresyon tahmini için farklı makine öğrenmesi yöntemlerinin karşılaştırılması sunulmuştur. Bu çalışma, en iyi tahmin performansını sağlayan makine öğrenmesi yönteminin bulunmasını amaçlamıştır. Çalışmada, Türkiye İstatistik Kurumu (TÜİK)'ten elde edilen, Hanehalkı Bütçe Araştırması anketi 2019 veri seti kullanılmıştır. 11.521 hanenin hanehalkı tüketim verileri incelenmiştir. Yapılan literatür taraması ve uzman görüşü rehberliğinde hanehalkı enerji harcamaları ile doğrudan veya dolaylı olarak ilgili değişkenler oluşturulmuştur. Hazırlanan veriler açık kaynak erişimli RapidMiner yazılım programı ile veri ön işleme, modelleme, tahmin ve değerlendirme aşamalarından geçirilmiştir. Hanehalkı toplam enerji harcamalarını tahmin etmek için regresyona bağlı farklı makine öğrenmesi yaklaşımları kullanılmıştır. Modelleme aşamasında, DT, DFNN, KNN, GBT ve RF makine öğrenmesi modelleri kullanılmıştır. Sonuç olarak, en yüksek R^2 (0,99) ve en düşük RMSE (5,5) ile en iyi performansı, DFNN modeli göstermiştir. Analiz sonuçları, hanehalkı toplam enerji harcamaları tahmininde derin sinir ağları yöntemi ile daha doğru sonuçlar elde edildiğini göstermektedir.

^a esmakriklioglu@hotmail.com

^b <https://orcid.org/0000-0001-6916-8643> | erkanoktay@atauni.edu.tr

^c <https://orcid.org/0000-0002-1739-3184>



Giriş

Enerji, modern ekonomilerde tüm sektörlerin üzerine kurulduğu yapı taşı olarak hizmet eder; bu nedenle, tüm ekonomik faaliyetlerin temelini oluşturur (Atems ve Hotaling, 2018). Dünyanın her yerinde, ülkelerin kalkınma hızı muazzam bir şekilde artmakta ve bu artış kaçınılmaz olarak her alanda enerji talebinin de artmasına neden olmaktadır (Chou ve Tran, 2018). Ekonomik gelişmelere bağlı olarak tüketim alışkanlıkları kontrolsüz bir şekilde genişlemektedir. Bu kontrolsüz büyüme, enerji kullanımı konusunda birçok yönden dünyamızın varlığını tehdit eden bir etken olmasına yol açmaktadır (TSKB, 2020). Tüketiciler için etkili bir enerji politikası geliştirilmesi açısından, bazı hanelerin neden diğerlerinden daha fazla enerjiye ihtiyaç duyduğunu bilmek önemlidir (Vringer ve ark., 2007). Bu anlamda hane halkları gelecekteki enerji tasarrufu için büyük bir potansiyeli temsil etmektedir. Dolayısıyla, hanehalkı enerji tüketimini azaltmanın uygun yollarını bulmak, enerji verimliliği programlarını etkin bir şekilde planlamak ve yürütmek politika yapımcıların temel hedeflerindedir (Salari ve Javid, 2017). 2019 yılında Türkiye’de, toplam enerji tüketiminde en fazla payı; ilk sırada %24,7 ile konut ve hizmet sektörü, ikinci sırada %23,8 ile sanayi sektörü ve üçüncü sırada %23,3 ile çevrim ve enerji sektörü oluşturmaktadır. Bunları %19,2 ile ulaştırma sektörü ve %3,3 ile tarım-hayvancılık sektörü izlemektedir. Enerji dışı tüketimin payı ise %4,9’ dur (ETKB, 2021). Türkiye’deki enerji tüketiminin büyük bir kısmını hanehalklarının enerji tüketimi oluşturmaktadır. Günümüz toplumu çeşitli enerji kaynaklarına bağımlı olduğundan çoğu ülkede, yerli sektör enerji tüketiminde en büyük paya sahiptir. Durumu Türkiye açısından değerlendirdiğimizde şunları söyleyebiliriz; Türkiye gibi enerji açısından dışa bağımlı ülkelerde, cari açığı besleyen enerji tüketim değerleri büyük önem arz etmektedir. Ayrıca, yaşam standartlarının iyileştirilmesi ve değişen tüketici tutumlarının enerji tüketiminin çeşitlenmesine yol açtığı söylenebilir. Bu bağlamda hanehalklarının enerji kullanımı, hem yerel hem de küresel olarak enerji tüketiminde önem arz etmektedir.

Bugün, büyük miktarda ve yüksek boyutlu verilerin var olması nedeniyle hanehalkı toplam enerji tüketimini tahmin etmek zor olabilmektedir. Bu nedenle, verilen bir problemi çözmek için, makinenin öğrenme algoritmalarını kullanarak ve problem ortamından elde edilen verileri inceleyerek çözüme ulaşmak hedeflenmektedir. Böylece makine öğrenmesi, çok büyük boyutlu verileri analiz etmeyi sağlamaktadır. Ayrıca, bu verileri en doğru şekilde değerlendirerek farklı uygulamalar için kullanılabilir kalıpları ortaya çıkarmaktadır. Veri kümelerinde ilk bakışta anlamsız görünen gizli bilgileri ortaya çıkarmaya çalışan makine öğrenmesi yöntemleri ile analiz yapmak, hanehalkı toplam enerji tüketim modellerinin anlaşılmasına yardımcı olmaktadır. Uygun makine öğrenmesi teknikleri, enerji tüketimini iyileştirmek ve doğru stratejilere ulaşmak için bilinmeyen kalıpların keşfedilmesini sağlamaktadır. Temel olarak uygulanan makine öğrenmesi teknikleri, veri kalıplarının analizi sonucunda sınıflandırma, tahmin ve regresyon modellerinde kullanılmaktadır. Bu bağlamda, yapay zeka (Artificial Intelligence -AI), özellikle makine öğrenmesi (Machine Learning -ML) çağımızın en önemli

teknolojisidir. Hanehalkı enerji tüketimi profilini analiz etmek, çok sayıda parametrenin tahmin edilmesini içerir. Bu bağlamda makine öğrenmesi yaklaşımları, veri ön işleme (preprocessing) süreçleri, çapraz geçerlilik (cross validation) yöntemleri ile eğitime ve model karşılaştırma yoluyla tahminleri etkinleştirme avantajına sahiptir (Burnett ve Kiesling, 2022).

Çalışmamızda, hanehalklarının toplam enerji harcama profillerini analiz etmek için TÜİK’in düzenli olarak her yıl yayınladığı Hanehalkı Bütçe Anketi kullanılmıştır. Türkiye hanehalkı toplam enerji harcamalarını, ML alanındaki yöntemler ile kapsamlı bir şekilde inceleyen çalışmalar yetersizdir. Bu çalışma, hanelerin toplam enerji harcamalarının ML yöntemleri ile karşılaştırmalı olarak incelenmesi neticesinde önceki çalışmaları bir adım ileriye taşımıştır. Ayrıca “ısıtma ve mutfak harcama”, “aydınlatma harcama”, “yakıt harcama” değerlerinin toplanarak hanelerin “toplam enerji harcama” değerlerinin elde edilmesi, alana yeni bir bakış açısı sunmaktadır. Bunlara ek olarak, bu çalışma; hanehalkı enerji tüketimini kontrol etmeyi ve enerji hizmetlerinin kalitesini iyileştirmeyi amaçlayan işletmelere yol gösterici olacaktır. Dahası, sistemi planlama kararları ve düzenleyici politikalar için daha etkin kararlar alınmasını sağlayacaktır. Hanehalkı enerji harcamalarını bir bütün olarak ele alan bu çalışma büyüyen literatüre yeni bir bakış açısı kazandıracaktır.

Literatür

Hanehalkları, hanelerdeki faaliyetler için enerjiyi çok yönlü olarak kullanırlar. Hanelerde talep edilen enerji, elektrik kullanımından su kullanımına, yemek pişirmeden yerleşim yerlerinde aydınlatmaya kadar çeşitli amaçlar için kullanılmaktadır. Bu kullanım hem doğrudan elektrik ve doğalgaz olarak düşünülebildiği gibi aynı zamanda dolaylı olarak üretim, ulaşım ve diğer mal ve hizmetlerin ticaretini içerir. Bunlarla birlikte, enerjiyi hangi demografik özelliklere ve hangi sosyo-ekonomik alt yapıya sahip hanelerin kullandığı önem taşımaktadır. Hanehalkının enerji kullanımını etkileyen bazı temel değişkenler bulunmaktadır. Bunlar; demografik özellikler, ekonomik faktörler, bireysel faktörler, kültür, yapı, yaşam tarzı, teknoloji, iklim ve politikadır. Hanehalkı enerji harcama profillerini makine öğrenmesi yöntemleri ile araştıran çalışmalar aşağıda özetlenmiştir.

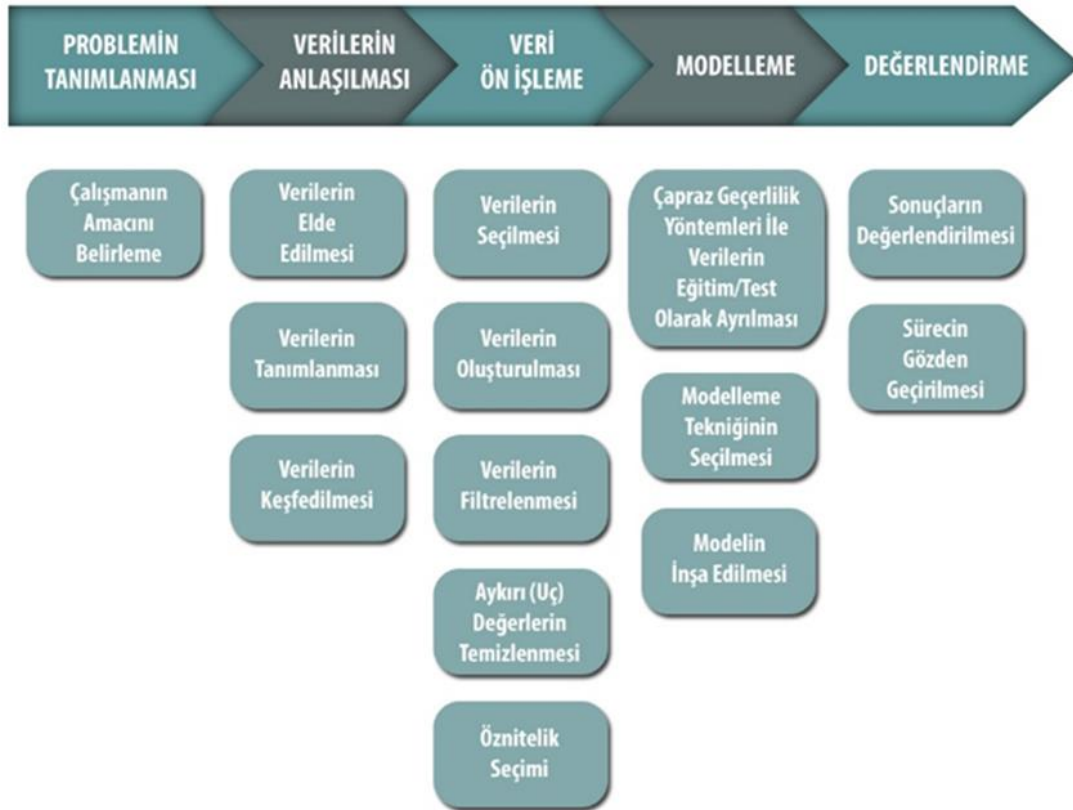
Elektrik enerjisi tüketiminin tahminini yapan bir çalışmada; üç modelleme tekniği karşılaştırılmış, karar ağacı (Decision Tree- DT) ve sinir ağları (Neural Networks - NN) modelleri regresyon analizinden daha iyi alternatifler olarak bulunmuştur (Tso ve Yau, 2007). Denetimli makine öğrenmesi tekniklerini kullanan bir çalışmada; bir hanenin elektrik tüketiminden hareketle hanenin belirli “özelliklerini” otomatik olarak tahmin etmek için bir sistem önerilmiştir (Beckel ve ark., 2014). DT ve yapay sinir ağları (Artificial Neural Networks – ANN) gibi yapay zeka yöntemlerini kullanarak bina elektrik enerjisini tahmin eden çalışmada; tek tahmin yönteminin yanı sıra, iki tahmin yönteminin birleşiminin daha doğru sonuçlar verdiği belirtilmiştir (Ahmad ve ark., 2014). Konutlarda elektrik tüketimini tahmin etmek için

hibrit bir model yaklaşımı öneren çalışmada; beş farklı makine öğrenme algoritması kullanılarak gerçek zamanlı ölçülen veriler ile hibrit model karşılaştırılmıştır. Önerilen hibrit modelin tahmin doğruluğunu %6–10 oranında iyileştirdiği vurgulanmıştır (Dong ve ark., 2016). Enerji tüketimini tahmin etmek için dokuz makine öğrenmesi tekniğini analiz eden bir çalışmada; İki veya daha fazla tahmin tekniğini birleştiren hibrit modelin çeşitli kombinasyonlarının, enerji tahmininde daha etkili olduğu bulunmuştur (Deb ve ark., 2017). Bina enerji tüketiminin tahmini için rastgele orman (Random Forest - RF) ve ANN modellerinin karşılaştırıldığı bir çalışmada; ANN modeli, RF modelinden marjinal olarak daha iyi performans göstermiştir. Ayrıca iki modelin de karşılaştırılabilir tahmin gücüne sahip olduğu ve bina enerji uygulamalarında neredeyse eşit derecede uygulanabilir olduğu belirtilmiştir (Ahmad ve ark., 2017). Ticari binalarda enerji tüketimini tahmin etmek için makine öğrenmesi yöntemlerini kullanan çalışmada; gradyan artırmalı regresyon modellerinin diğer yöntemlerden daha iyi performans gösterdiği bulunmuştur (Robinson ve ark., 2017). Makine öğrenme tekniklerini kullanarak hanehalklarının kullanım kalıplarına bağlı olarak enerji tüketimlerini tahmin eden çalışmada; hibrit modelin, tekli ve toplu modellerden daha doğru sonuçlar verdiği vurgulanmıştır (Chou ve Tran, 2018). Elektrik enerjisi tüketim tahmini için hibrit bir yapay sinir ağı kullanan çalışmada; önerilen modelin tahmin doğruluğu açısından daha önceki yöntemlerden daha üstün performansa ulaştığı sunulmuştur (Li ve ark., 2018). Ev aletleri enerji tüketimini tahmin eden bir çalışmada; beş makine öğrenmesi tahmin modeli kullanılmıştır. Derin

öğrenme yöntemi olan uzun-kısa süreli bellek ağı (Long Short-Term Memory Network - LSTM) en iyi performansı göstermiştir (Wang ve ark., 2020). Enerji tüketimini tahmin etmek için bina enerji tüketiminin beş yıllık veri kümesini kullanan bir çalışmada; önerilen RF tabanlı bir tahmin modeli, iyi bir tahmin doğruluğu ortaya koymuştur (Pham ve ark., 2020). Binalarda enerji tasarrufu sağlamak için uyarlanabilecek faktörleri araştıran bir çalışmada kümeleme, karar ağacı ve birliktelik kurallarının kullanıldığı bir yaklaşım önerilmiştir (Nazeriye ve ark., 2021). Akıllı sayaç verilerini kullanarak hane profillerini belirlemek için yarı denetimli öğrenme yaklaşımı öneren çalışmada; önerilen yaklaşımın yalnızca sınırlı etiketli veri mevcut olduğu zaman denetimli öğrenme yöntemlerinden daha iyi performans gösterdiği belirlenmiştir (Wang ve ark., 2022). Veri boyutunun model performansı üzerindeki etkisinin araştırıldığı çalışmada; yıllık enerji tüketimini tahmin etmek için derin öğrenme, topluluk öğrenme ve makine öğrenme modelleri karşılaştırılmıştır. 9 modelin geliştirildiği çalışmada derin öğrenme modelinin diğer modellerden daha verimli olduğu açıklanmıştır (Olu-Ajayi ve ark., 2022).

Makine öğrenmesi teknikleri kullanan çalışmaların literatür taraması sonuçlarına göre, hanehalkı enerji harcamalarını, makine öğrenmesi alanındaki yöntemler ile kapsamlı bir şekilde inceleyen çalışmalar hala keşfedilmesi gereken bir çalışma alanıdır. Bu çalışma, hanehalklarının toplam enerji harcamalarının farklı makine öğrenmesi yöntemleri ile karşılaştırmalı olarak incelenmesi neticesinde literatüre titiz bir katkı sağlayacaktır.

Bu çalışmanın makine öğrenmesi algoritması aşamaları, Şekil 1'de sunulmuştur.



Şekil 1. Makine Öğrenmesi Algoritması Aşamaları
Figure 1. Machine Learning Algorithm Stages

Çizelge 1. Değişkenlerin Listesi ve Tanımı

Table 1. List and Definition of Variables

Değişkenler	Kategoriler
Toplam Enerji Harcama	Aydınlatma Harcama + Isınma ve Mutfak Harcama + Yakıt Harcama (Araç)
Aydınlatma Harcama Isınma ve Mutfak Harcama Yakıt Harcama (Araç) Yıllık Kullanılabilir Gelir Elektrik Harcaması Hanehalkı Harcama Değeri	TL
Konutun Tipi	1 Müstakil Konut 2 Apartman Dairesi
Mülkiyet Durumu	1 Ev Sahibi 2 Kiracı 3 Lojman 4 Ev Sahibi Değil ama Kira Ödemiyor
Binanın İnşa Tarihi	1 1961-1980 2 1981-2000 3 2001 ve sonrası 4 1960 ve öncesi
Oda Sayısı Konutun Alanı İkinci Konut Durumu Doğal Gaz Sahipliği Tarla Bahçe Sera Sahipliği	(Salon dahil) m ² 1 Var 2 Yok
Cinsiyet	1 Erkek 2 Kadın
Yaş	Bitirdiği Yaş
Eğitim Durumu	1 Bir Okul Bitirmesi-İlkokul 2 Ortaokul 3 Lise 4 Yüksek Okul, Lisans-Lisansüstü
Medeni Durum	1 Hiç Evlenmedi 2 Evli 3 Boşandı-Eşi öldü
Çalışma Durumu	1 Çalışıyor 2 Çalışmıyor
Meslek	1 Yöneticiler 2 Profesyonel Meslek Mensupları 3 Teknisyenler, Teknikerler ve Yardımcı Profesyonel Meslek Mensupları 4 Büro Hizmetlerinde Çalışan Elemanlar 5 Hizmet ve Satış Elemanları 6 Nitelikli Tarım, Ormancılık ve Su Ürünleri Çalışanları 7 Sanatkarlar ve İlgili İşlerde Çalışanlar 8 Tesis ve Makine Operatörleri ve Montajcıları 9 Nitelik Gerektirmeyen İşlerde Çalışanlar 10 Çalışmıyor
Hanede Okuyan Sayısı Hanedeki Çalışan Sayısı Hanedeki Çocuk Sayısı Hanedeki Yaşlı Sayısı Hanehalkı Büyüklüğü Otomobil Sayısı Hanedeki Elektrikli Alet Sayısı	Pozitif Tam Sayı
Günlük Alışveriş Hizmetlerine Ulaşım Zorluğu Toplu Ulaşım Hizmetlerine Ulaşım Zorluğu Bankacılık Hizmetlerine Ulaşım Zorluğu Sağlık Hizmetlerine Ulaşım Zorluğu Posta Hizmetlerine Ulaşım Zorluğu Zorunlu Eğitim Hizmetlerine Ulaşım Zorluğu	1 Çok Kolay 2 Kolay 3 Orta 4 Zor 5 Çok Zor

*Toplam enerji harcama (bağımlı değişken), diğer değişkenler (bağımsız değişken)

Yöntem

Veri ve Değişkenler

Bu çalışmada, her yıl TÜİK tarafından yayınlanan Hanehalkı Bütçe Araştırması Anketi 2019 veri seti kullanılmıştır. Hanehalkı Bütçe Araştırması Anketi nde ana örnekleme çerçevesi olarak Ulusal Adres Veri Tabanı ve örnekleme yöntemi olarak; tabakalı iki aşamalı küme örnekleme yöntemi kullanılmaktadır (TÜİK, 2016). Hanehalkı Bütçe Araştırması Anketi, hanehalklarına ait önemli birçok bilginin toplanmasında ve düzenlenmesinde kullanılmaktadır. Her bir yıla ait ortalama 10.000 kayıt içeren bu veri seti, ML yöntemlerinin uygulanması için yeterli sayıdadır. Anket verilerinin tüketim kısmını içeren yaklaşık 300 farklı tüketim kalemi mevcuttur. Bu kalemlerden her biri ayrı ayrı ya da bir araya getirilerek ML yöntemleri ile incelenebilmektedir.

Bu çalışmada kullanılan demografik, sosyoekonomik değişkenler ve konut özellikleri değişkenleri Çizelge 1'te gösterilmiştir.

Veri Ön İşleme

Veriye uygulanan veri ön işleme, öznitelik seçimi, modelleme, tahmin ve performans değerlendirme aşamaları için RapidMiner (RapidMiner Studio 9.10.10) açık kaynak kodlu yazılım programı kullanılmıştır. 2019 yılı bütçe anketindeki 11.521 hanenin veri seti incelenmiştir. Veri setinde, 2019 yılında toplanan hanehalkı tüketim özelliklerine ilişkin bilgileri içeren üç farklı veri dosyası bulunduğu görülmüştür. Bunlar; fert karakteristikleri (cinsiyet, eğitim durumu, medeni durum, ...), hane karakteristikleri (konut tipi, konutun inşa tarihi, oda sayısı, ...), tüketim karakteristikleri (tüketilen mal ve hizmetlere ilişkin değerler) dir. Bu veri seti kullanılarak değişkenler, alan bilgisi ve uzman görüşü rehberliğinde yapılan literatür taraması sonucunda çalışmanın amacına uygun olarak birleştirilmiştir. Halihazırda veri setinde var olan değişkenlerin yanı sıra yeni değişkenler de oluşturulmuştur. Modelleme işlemine hazır olması için veriler bazı temizleme, yapılandırma, filtreleme işlemlerinden geçirilmiştir. Bu anlamda ilk olarak “ısınma ve mutfak harcama”, “aydınlatma harcama”, “yakıt harcama” değerleri toplanarak hanelerin toplam enerji harcama değerleri elde edilmiştir. Bazı değişkenler çalışmamızdaki amaç için anlamlı bir etkiye sahip değildir. Bu nedenle analiz için anlamsız olan toplam enerji harcama değerleri, ısınma ve mutfak harcama değerleri ve aydınlatma harcama değerleri içerisinde sifıra eşit olan haneler sırasıyla veri setinden çıkarılmıştır. Ayrıca, otomobil sahibi olup yakıt harcama değeri sifıra eşit olan haneler analizimiz için anlamsız olup veri setinden çıkarılmıştır. Devam eden süreçte aykırı (uç) değerleri bulmak için yoğunluğa dayalı bir aykırı değer hesaplama yöntemi olan lokal aykırı faktör (LOF- Local Outlier Faktor) yöntemi kullanılmıştır. LOF, bir nesnenin kendisini çevreleyen komşuluğa göre ne kadar izole olduğunun derecesi olarak sayısal bir ölçekte aykırılık sağlayarak yerel aykırı değerleri tespit için geliştirilmiş bir yöntemdir (Breunig ve ark., 2011). Değerlendirme sonucunda aykırı değerler veri setinden çıkarılmıştır. Yapılan uygulamalar sonucunda analizimiz için hazır halde temiz bir veri seti elde edilmiştir. “Toplam enerji harcama” değişkeni “bağımlı değişken” olarak

belirlenmiştir. Literatür taraması sonucunda oluşturulan Çizelge 1 deki diğer değişkenler ise “bağımsız değişkenler” olarak tespit edilmiştir.

Öznitelik seçme (feature selection) yöntemleri kullanılarak belirlediğimiz 33 bağımsız değişken içerisinde en uygun değişkenler seçilmiştir. Öznitelik seçiminin odak noktası; girdi verilerini doğru bir şekilde karakterize edebilen, gürültü ve alakasız değişkenlerin etkilerini azaltırken yüksek tahmin sonuçları sağlayan bir değişken alt kümesi bulmaktır. Literatürde öznitelik seçimi için çeşitli yöntemler kullanılmıştır (Arunadevi ve Nithya, 2016; Chandrashekar ve Sahin, 2014; Ghasemi ve ark., 2020; Lee ve ark., 2011). Bu çalışmada, İleri Seçim (Forward Selection), Geriye Eleme (Backward Elimination), Optimize Edilmiş Seçim –Evrimsel (Optimize Selection- Evoluntary) öznitelik seçim yöntemleri denenmiştir. Model performans sonucuna en fazla katkıyı sağlayan Optimize Edilmiş Seçim (Evrimsel) yöntemi kullanılarak analize devam edilmiştir. Bu evrimsel yöntemde bir genetik algoritma kullanılmaktadır. Genetik algoritma, doğal evrim sürecini taklit eden bir buluşsal arama algoritmasıdır. Öznitelik seçimi ise, verilen örnek kümesi içinde sınıflandırma analizi için en uygun öznitelikleri seçen bir yöntemdir (Panthong ve Srivihok, 2015). Çalışmamızda öznitelik seçiminden sonra tahmin performansına daha fazla katkı sağlayan 33 değişkenden 18 tanesi seçilmiştir.

Modelleme

Bu çalışmada, hanehalkı toplam enerji harcamalarını tahmin etmek için regresyona bağlı makine öğrenimi yaklaşımları kullanılmıştır. Modelleme aşamasında DT, DFNN, KNN, GBT ve RF makine öğrenmesi modelleri kullanılmıştır. Yöntemlerin eğitim ve test verisi olarak ayrılması işleminde, 10-katlı çapraz geçerlilik yöntemi uygulanmıştır. 10-katlı çapraz geçerlilik yönteminde veri seti 10 parçaya ayrılır. Bu parçalardan her biri sırayla, test veri seti ve geri kalan 9 parça eğitim veri seti olarak kullanılır. Bu işlem 10 kez tekrar edilir. Doğruluk değerlerinin ortalaması alınarak modelin doğruluk oranı belirlenir.

Karar Ağacı (Decision Tree -DT)

Karar ağacı algoritması ilk olarak 1963 yılında Morgan ve Sonquist (1963) tarafından geliştirilmiştir. Karar ağaçları regresyon ve sınıflandırma için kullanılan genel amaçlı bir tahmin yöntemidir. Karar ağaçlarının temel özelliği, yapraklar veya düğümler olarak adlandırılan alt kümeler oluşturarak hedef veri alanını tahmin etmektir. Bu bağlamda DT, çok aşamalı ve ağaç benzeri bir yapıya dayanır (Zhang ve ark., 2005). Ağacın dallarında tahmin elde edilmiyorsa bu dalda düğüm oluşur. Buna karşın dallarda tahmin elde ediliyorsa dalın sonunda yaprak oluşur. Bu yaprak, veri üzerinden elde edilen tahminlerden biridir (Ayık ve ark., 2007). Her alt küme için aynı işlem bir yaprak oluşana kadar ardışık olarak yürütülür. Burada temel amaç, karmaşık bir kararı birkaç basit karara dönüştürmektir ve yorumlanması daha kolay bir çözüme ulaşmaktır (Zhang ve ark., 2005). Karar ağaçları, güçlü, esnek ve kullanımı kolay yöntemlerdir. Eksik verilerin varlığında, eksik verileri yeni modellere dahil etmenin birden çok yolunu sunmaktadır. Bu nedenle, yüksek kaliteli sonuçların üretilmesini sağlar (De Ville, 2013).

Derin Sinir Ağları (Deep Neural Networks)

Derin sinir ağı, giriş ve çıkış katmanları arasında birden çok gizli katmana sahip bir yapay sinir ağı algoritmasıdır. Yapay sinir ağları, kabaca insan beynindeki nöronların nasıl çalıştığından ilham alan bir makine öğrenmesi disiplini. Sinir Ağları, insan sinir hücrelerine benzer prensipler ile çalışır. Bunlar, çeşitli temel değişkenler arasındaki ilişkiyi yakalayan ve verileri insan beyninin yaptığı gibi işleyen bir dizi algoritmadır. Karmaşık sinir ağlarını bazı alanlarda başarılı bir şekilde eğitmek, çok fazla veri ve işlem kapasitesi gerektirir. Bu nedenle farklı uygulamalar için çeşitli sinir ağları türleri vardır (Jain vd., 1996). Çalışmamızda, derin ileri beslemeli sinir ağı (deep feed-forward artificial neural network, DFNN) kullanılmıştır. Kullanılan derin sinir ağı algoritması, geri yayılım kullanılarak stokastik gradyan inişi ile eğitilmiş çok katmanlı bir ileri beslemeli yapay sinir ağına dayanır. İleri Beslemeli Sinir Ağı, düğümlerin dairesel olarak bağlandığı yapay bir sinir ağıdır. Bazı rotaların döngülediği ileri beslemeli bir sinir ağı, tekrarlayan sinir ağı (Recurrent Neural Network- RNN)'nin tam tersidir. İleri beslemeli modelde, girdi yalnızca bir yönde işlendiğinden, temel sinir ağı mimarisinin bir türüdür. Bu tür sinir ağı, tüm bilgileri yalnızca ilettiği için çok katmanlı sinir ağı olarak da bilinir. Sinir ağı, delta kuralı olarak bilinen bir özelliği kullanarak düğümlerinin çıktılarını istenen değerlerle karşılaştırabilir ve ağı daha doğru çıktı değerleri oluşturmak için eğitim yoluyla ağırlıklarını değiştirmesine izin verir. Bu eğitim ve öğrenme prosedürü, gradyan inişle sonuçlanır. Çok katmanlı algılayıcılarda ağırlıkları güncelleme tekniği hemen hemen aynıdır, ancak süreç geri yayılım olarak adlandırılır. Bu gibi durumlarda, son katman tarafından sağlanan çıkış değerleri, ağ içindeki her bir gizli katmanı değiştirmek için kullanılır.

K-en yakın komşu (K-Nearest Neighbor – KNN)

K-en yakın komşu modeli, Cover ve Hart (1967) tarafından sınıflama ve regresyon problemlerine çözüm olarak öne sürülmüştür. Amaç girdi ve çıktı değişkenleri arasındaki ilişkiyi modelleyen, KNN regresyon fonksiyonu olarak da bilinen fonksiyonu tahmin etmektir. KNN algoritması eğitim veri seti olarak bilinen geçmiş verileri toplar ve toplanan bu verileri yeni test verilerini tahmin etmek için kullanır. Sonra eğitim veri setinin k-en yakın komşuluk verileri elde edilir. Bu verilere dayanarak her yeni veri için tahmin yapılır. KNN algoritması yalnızca bir tane kullanmak yerine en yakın eğitim örneklerinin önceden tanımlanmış bir sayısını kullanarak bu prosedürü genişletir (Olu-Ajayı, 2017). Yeni veri daha sonra, k en yakın komşu arasından en çok benzediği sınıfa atanır (Ayık ve ark., 2007; Poloczec ve ark., 2014). KNN, sınıflandırma ve regresyon için kullanılan en yaygın ve anlaşılır makine öğrenme yöntemlerinden biridir. Bununla birlikte, bu teknik birçok uygulama için çok etkilidir.

Gradyan Artırmalı Ağaç (Gradient Boosted Tree - GBT)

Gradyan artırmalı ağaç yaklaşımı, 1999 yılında Friedman (2001) tarafından geliştirilmiştir. GBT, "Gradient descent" ve "boosting" kelimelerinin bir araya gelmesi ile oluşan bir karar ağacı algoritmasıdır. GBT, aynı zamanda güçlü tahmine dayalı modellerin oluşturulabileceği bir topluluk öğrenme tekniğidir. GBT algoritmaları, bilinen bir dizi bireysel regresyon ağacının birleştirilmesiyle elde edilir. Zayıf olarak adlandırılan ağaçlardan daha sağlam tahmin modelleri oluşturur

(Gilbert ve ark., 2020). Bu yöntem, modeli diğer artırma yöntemleri gibi aşamalar halinde meydana getirir ancak bunları keyfi bir türevlenebilir kayıp fonksiyonu optimize ederek genelleştirir (Flores ve Keith, 2019). Gradyan iniş yoluyla kayıp fonksiyonları azaltırken karar ağaçlarını birleştirerek tahmin hatalarını azaltır. Regresyon ve sınıflandırma problemlerinde kullanılmaktadır (Olu-Ajayı ve ark., 2022).

Rastgele Orman (Random Forest - RF)

Breiman (2001) tarafından geliştirilen Rastgele Orman modeli, farklı regresyon ağaçlarından oluşan tahmin edicilerin bir kombinasyonudur. RF, ağaç yapılı tahmincilerin gelişmiş bir sürümü olarak düşünülebilir (Akar ve Güngör, 2012). Tahmin için bir dizi ağaç kullanmanın ardındaki mantık, birden fazla ağacın tahminini birleştirerek her ağacın kararsızlık sorununu azaltmaktır (Wang ve ark., 2018). Veri setinden farklı alt kümeler oluşturulmakta ve sonra bu alt kümeler, karar ağacı topluluğundaki her bir karar ağacı tarafından eğitilmektedir. Her bir düğümde rastgele olarak seçilen değişkenler arasından en iyisi kullanılarak her bir düğüm dallara ayrılmaktadır. Her düğümü bölmek için rastgele bir özellik seçimi kullanılarak ağaçlar geliştirilir. Geliştirilen ağaçlar daha fazla bölünme mümkün olmayana kadar büyütülür ve budanmaz (Archer ve Kimes, 2008). Güçlü bir tahmin modeli olan RF, sınıflandırma ve regresyon tahminlerinde yüksek performans göstermektedir.

Model Değerlendirme

Farklı modeller veya algoritmalar kullanan çalışmaların karşılaştırılması için model değerlendirme kriterleri son derece önemlidir. Makine öğrenmesi yöntemleri kullanan farklı literatürlerin tahmin kriterlerinin karşılaştırıldığı çalışmalarda (Liu ve ark., 2019; Olu-Ajayı ve ark., 2022) çeşitli değerlendirme kriterleri kullanıldığı tespit edilmiştir. Bahsi geçen çalışmaların incelenmesi sonucunda; çalışmamızdaki modellerin performansını değerlendirmek için ortalama kare hata (Mean Square Error- MSE), kök ortalama kare hata (Root Mean Square Error - RMSE) ve R-kare (Predictive Coefficient Of Determination - R²) performans kriterleri kullanılmıştır.

MSE; bir veri kümesindeki tahmin edilen değerler ile gerçek değerler arasındaki ortalama kare farkını ifade eden bir kriterdir. MSE ne kadar düşükse, model bir veri kümesine o kadar iyi uymaktadır. RMSE; bir veri kümesindeki tahmin edilen değerler ile gerçek değerler arasındaki ortalama kare farkının karekökünü ifade eden bir kriterdir. RMSE ne kadar düşükse, model bir veri kümesine o kadar iyi uymaktadır. Bir modelin bir veri kümesine ne kadar uyduğunu değerlendirirken, yanıt değişkeniyle aynı birimlerde ölçüldüğü için RMSE daha sık kullanılır. Tersine, MSE, yanıt değişkeninin kare birimleri olarak ölçülür. RMSE, bir makine öğrenmesi modelinin tahmin değeri ile gerçek değerleri arasındaki farkların ölçüsü olarak kullanılan bir kriterdir. Gözlenen değer ile gerçek değer arasındaki sapmayı ölçmek için kullanılır. Değerlendirme yapılırken en düşük RMSE değerine sahip model "en iyi" model olarak seçilir. Çünkü bu model, veri setinden gerçek değerlere en yakın tahminler yapan modeldir. R², bağımsız değişken tarafından açıklanabilen bağımlı değişkendeki varyans oranını belirleyen bir regresyon modeli kriteridir.

RMSE, MSE ve R²'nin hesaplama formülleri,

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

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

$$R^2 = 1 - \frac{\sum_{i=1}^m (\hat{y}_i - y_i)^2}{\sum_{i=1}^m (y_i - \bar{y}_i)^2}$$

dir. Burada y_i , i zaman noktalarındaki verilerin gerçek değeri, \hat{y}_i zaman noktalarındaki verilerin tahmin edilen değeridir ve \bar{y}_i örneklemin ortalama değeridir, m ise örneklemin toplam sayısıdır.

Bulgular

Modelleme aşamasında DT, DFNN, KNN, GBT, RF olmak üzere beş makine öğrenmesi modeli kullanılmıştır. Bu modeller içerisinde en iyi tahmin modelini bulmak için RMSE ve R² performans değerlendirme sonuçları tablo 2'de sunulmuştur. Performans değerlendirme kriterlerine göre, RMSE'si düşük ve R²'si yüksek olan model daha iyi performansa sahip olan modeldir. Bu bağlamda, RMSE'si sıfıra yakın ve R²'si 1 e yakın olan model en iyi tahmin modeli olarak ortaya çıkar. En yüksek R²'ye sahip iki model olduğu görülmektedir. Bunlar R² değerleri 0,99 olan DFNN ve KNN modelleridir. Bu iki yöntem arasında DFNN'nin RMSE değeri 5,5 ve KNN'nin RMSE değeri 29,8 olarak bulunmuştur. RMSE'si düşük olan yöntem DFNN olduğu için en iyi performansı DFNN modeli gösterir. KNN ise en iyi ikinci modeldir. DFNN ve KNN modellerinden daha düşük performansa sahip olan yöntem, R² değeri 0,97 ve RMSE değeri 57,7 olan DT yöntemidir. R² değeri 0,95 ve RMSE değeri 85,4 olan RF, DT ile yakın performans değerlerine sahiptir. En düşük R² değeri 0,76 ve en yüksek RMSE değeri 171,9'a sahip olan GBT modeli en düşük performans gösteren yöntem olarak bulunmuştur. Genel olarak, hanehalkı toplam enerji harcamalarını tahmin için beş makine öğrenmesi modeli karşılaştırılmıştır. Çizelge 2'deki modeller arasında, DFNN tahmin modelinde gerçek değere çok yakın değerler elde edildiği görülmektedir.

Çizelge 2. Modellerin Performans Sonuçları
Table 2. Performance Results of Models

Model	R ²	RMSE
DFNN	0,99*	5,5*
DT	0,97	57,7
RF	0,95	85,4
KNN	0,99	29,8
GBT	0,76	171,9

* değerler, en iyi performansı temsil eder.

Tartışma ve Sonuç

Dünyadaki ekonomik faaliyetlerin artması, teknolojinin gelişmesi ve değişen yaşam koşulları neticesinde hanehalklarının enerji tüketimi de artmaktadır. Hanehalkları, toplam kullanım dağılımı göz önüne alındığında, enerji kullanımı konusunda önemli bir payı

temsil etmektedir. Hanehalkı enerji kullanımı alanında tahmin elde etmek için hangi yöntemin en iyi veya uygun olduğu, hangi model veya algoritmanın önerilmesi gerektiği hala keşfedilmesi gereken bir araştırma konusudur. Bugüne kadar performans değerlendirmesi için istatistiksel yöntemler ile yapılan çalışmalar, birçok alanda sıklıkla kullanılmaktadır. Buna karşılık, hanehalklarının enerji harcama profillerini anlamaya çalışan makine öğrenmesi yöntemleri ile yapılan çalışmalar gelişmeye devam eden bir literatüre sahiptir.

Bu çalışmada; TÜİK Hanehalkı Bütçe Araştırması Anketi 2019 yılı verileri incelenmiştir. Ankette yer alan tüketim kalemlerinden hanehalkı enerji harcamaları ile ilgili olan değişkenler uzman görüşü ve literatür eşliğinde oluşturulmuştur. Veriler ön işleme aşamalarından geçirilmiş olup filtrelenen bazı özellikler ile veri seti temizlenmiştir. LOF aykırı değer bulma yöntemi ile uç değerler veri setinden çıkarılmıştır. Optimize Edilmiş Seçim (Evrimsel) öznitelik seçme yöntemi ile en uygun öznitelikler elde edilmiştir. 10-katlı çapraz geçerlilik yaklaşımı ile veri, eğitim ve test veri setine ayrılmıştır. Sonrasında farklı makine öğrenmesi yöntemlerine dayalı modeller kurulmuştur. Hanehalkı toplam enerji harcamasını tahmin etmek için DT, DFNN, KNN, GBT v RF olmak üzere beş makine öğrenmesi yöntemi karşılaştırılmıştır. DFNN, hanehalkı toplam enerji harcama tahmini için diğer modellerden daha iyi sonuçlar vermiştir. Son yıllarda yapılan çalışmalar (Olu-Ajayı ve ark., 2022; Wang ve ark., 2020) bu sonucu desteklemektedir. Bununla birlikte KNN model sonucunun, ikinci derecede iyi bir performans sağlaması Olu-Ajayı (2017) çalışmasını desteklemektedir. DT, RF modelleri birbirlerine yakın düzeyde performansa sahip olarak bulunmuştur. GBT ise en düşük performansı sağlamıştır. Makine öğrenimi yöntemlerinin kıyaslandığı çalışmada (Robinson ve ark., 2017), GBT diğer yöntemlerden daha iyi performans göstererek çalışmamızdan farklı bir sonuç elde edilmiştir. Dört makine öğrenmesi modeli ile regresyon tahmini yapan çalışmada (Singh ve ark., 2021), performans sırası yüksekten düşüğe doğru GBT, RF, KNN, DT şeklinde olup çalışmamızın tam tersi şekilde bir sonuç elde edilmiştir.

Sonuç olarak, hanehalkı toplam enerji harcamasını tahmin etmek için farklı alanlarda iyi sonuçlar veren makine öğrenmesi modelleri incelenmiştir. Son yıllarda kullanımı genişleyen DFNN, GBT modelleri ile klasik modeller olan ve çok sayıda araştırmacı tarafından incelenen KNN, RF ve DT modellerinin karşılaştırma sonuçları sunulmuştur. DFNN modelinin birkaç popüler geleneksel makine öğrenmesi modelinden daha iyi performans gösterdiği tespit edilmiştir. Bunun nedeni, DFNN yönteminin daha derin bir ağ mimarisine sahip olmasından kaynaklanmaktadır. Çalışmanın veri ön işleme, öznitelik seçimi, modelleme, tahmin ve performans değerlendirme aşamaları açık kaynak erişimli RapidMiner yazılım programı ile gerçekleştirilmiştir. Tahminde kullanılan LOF aykırı değer bulma, Optimize Edilmiş Seçim (Evrimsel), 10-katlı çapraz geçerlilik bulma algoritmaları farklı algoritma seçenekleri ile tekrar ele alınabilir. Bunlara ilaveten, gelecek çalışmalarda geliştirilmiş farklı derin sinir ağı modelleri kullanılabilir. Hanehalkı veri setindeki değişkenler, veri setinin kalitesini artırmak için farklı yollarla seçilebilir ve oluşturulabilir. Ayrıca istenen ayrıntı düzeyine ulaşmak için topluluk öğrenmesi modelleri dikkate alınarak tahminler yapılabilir.

Kaynaklar

- Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H. ve Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. *Renewable and Sustainable Energy Reviews*, 33, 102–109. doi:10.1016/j.rser.2014.01.069
- Ahmad, M. W., Mourshed, M. ve Rezgüi, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, 147, 77–89. doi:10.1016/j.enbuild.2017.04.038
- Aizenberg, I. N., Aizenberg, N. N. ve Vandewalle, J. (2000). Multiple-Valued Threshold Logic and Multi-Valued Neurons. *Multi-Valued and Universal Binary Neurons* içinde (ss. 25–80). Boston, MA: Springer US. doi:10.1007/978-1-4757-3115-6_2
- Akar, Ö. ve Güngör, O. (2012). Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması. *Journal of Geodesy and Geoinformation*, 1(2), 139–146. doi:10.9733/jgg.241212.1t
- Archer, K. J. ve Kimes, R. v. (2008). Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4), 2249–2260. doi:10.1016/j.csda.2007.08.015
- Arunadevi, J. ve Nithya, M. J. (2016). Comparison of feature selection strategies for classification using rapid miner. *International Journal of Innovative Research in Computer and Communication Engineering*, 4(7), 13556-13563.
- Atems, B. ve Hotaling, C. (2018). The effect of renewable and nonrenewable electricity generation on economic growth. *Energy Policy*, 112, 111–118. doi:10.1016/j.enpol.2017.10.015
- Ayık, Y. Z., Özdemir, A. ve Yavuz, U. (2007). Lise Türü Ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 441-454.
- Beckel, C., Sadamori, L., Staake, T. ve Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy*, 78, 397–410. doi:10.1016/j.energy.2014.10.025
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. doi:10.1023/A:1010933404324
- Breunig, M. M., Kriegel, H.-P., Ng, R. T. ve Sander, J. (2000). LOF. *Proceedings of the 2000 ACM SIGMOD international conference on Management of data - SIGMOD '00* içinde (ss. 93–104). New York, New York, USA: ACM Press. doi:10.1145/342009.335388
- Burnett, J. W. ve Kiesling, L. L. (2022). How do machines predict energy use? Comparing machine learning approaches for modeling household energy demand in the United States. *Energy Research & Social Science*, 91, 102715. doi:10.1016/j.erss.2022.102715
- Chandrashekar, G. ve Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1), 16–28. doi:10.1016/j.compeleceng.2013.11.024
- Chou, J.-S. ve Tran, D.-S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. *Energy*, 165, 709–726. doi:10.1016/j.energy.2018.09.144
- Cover, T. ve Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. doi:10.1109/TIT.1967.1053964
- Enerji ve Tabii Kaynaklar Bakanlığı, <https://enerji.gov.tr/enerji-isleri-genel-mudurlugu-denge-tablolari>, 2021.
- Deb, C., Zhang, F., Yang, J., Lee, S. E. ve Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*, 74, 902–924. doi:10.1016/j.rser.2017.02.085
- Deng, L. (2014). Deep Learning: Methods and Applications. *Foundations and Trends® in Signal Processing*, 7(3–4), 197–387. doi:10.1561/20000000039
- De Ville, B. (2013). Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(6), 448-455.
- Dong, B., Li, Z., Rahman, S. M. M. ve Vega, R. (2016). A hybrid model approach for forecasting future residential electricity consumption. *Energy and Buildings*, 117, 341–351. doi:10.1016/j.enbuild.2015.09.033
- Enerji Görünümü; TSKB; 2020. <https://www.tskb.com.tr>.
- Flores, V. ve Keith, B. (2019). Gradient Boosted Trees Predictive Models for Surface Roughness in High-Speed Milling in the Steel and Aluminum Metalworking Industry. *Complexity*, 2019, 1–15. doi:10.1155/2019/1536716
- Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29(5), 1189–1232. <http://www.jstor.org/stable/2699986>
- Ghasemi, F., Neysiani, B. S. ve Nematbakhsh, N. (2020). Feature Selection in Pre-Diagnosis Heart Coronary Artery Disease Detection: A heuristic approach for feature selection based on Information Gain Ratio and Gini Index. *2020 6th International Conference on Web Research (ICWR)* içinde (ss. 27–32). IEEE. doi:10.1109/ICWR49608.2020.9122285
- Gilbert, C., Browell, J. ve McMillan, D. (2020). Leveraging Turbine-Level Data for Improved Probabilistic Wind Power Forecasting. *IEEE Transactions on Sustainable Energy*, 11(3), 1152–1160. doi:10.1109/TSTE.2019.2920085
- Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., ... & Webster, D. R. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22), 2402-2410.
- Ivakhnenko, A. G. ve Lapa, V. G. (1966). Cybernetic predicting devices. *Purdue Univ Lafayette Ind School Of Electrical Engineering*.
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). “Artificial neural networks: A tutorial”. *Computer*, 29(3), 31-44.
- Lee, S., Schowe, B. ve Sivakumar, V. (2011). *Feature selection for high-dimensional data with RapidMiner. Technical Report*.
- Li, K., Xie, X., Xue, W., Dai, X., Chen, X. ve Yang, X. (2018). A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. *Energy and Buildings*, 174, 323–334. doi:10.1016/j.enbuild.2018.06.017
- Liu, Z., Wu, D., Liu, Y., Han, Z., Lun, L., Gao, J., ... ve Cao, G. (2019). Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction. *Energy Exploration & Exploitation*, 37(4), 1426-1451.
- Morgan, J. N. ve Sonquist, J. A. (1963). Problems in the Analysis of Survey Data, and a Proposal. *Journal of the American Statistical Association*, 58(302), 415–434. doi:10.1080/01621459.1963.10500855
- Nazeriye, M., Haeri, A., Haghghat, F. ve Panchabikesan, K. (2021). Understanding the influence of building characteristics on enhancing energy efficiency in residential buildings: A data mining based study. *Journal of Building Engineering*, 43, 103069. doi:10.1016/j.job.2021.103069
- Olu-Ajayi, R. (2017). An investigation into the suitability of k-Nearest Neighbour (k-NN) for software effort estimation. *International Journal of Advanced Computer Science and Applications*, 8(6).
- Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F. ve Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering*, 45, 103406. doi:10.1016/j.job.2021.103406
- Panthong, R. ve Srivihok, A. (2015). Wrapper Feature Subset Selection for Dimension Reduction Based on Ensemble Learning Algorithm. *Procedia Computer Science* içinde (C. 72). doi:10.1016/j.procs.2015.12.117

- Pham, A.-D., Ngo, N.-T., Ha Truong, T. T., Huynh, N.-T. ve Truong, N.-S. (2020). Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *Journal of Cleaner Production*, 260, 121082. doi:10.1016/j.jclepro.2020.121082
- Poloczek, J., Treiber, N. A. ve Kramer, O. (2014). KNN Regression as Geo-Imputation Method for Spatio-Temporal Wind Data (ss. 185–193). doi:10.1007/978-3-319-07995-0_19
- Rapidminer, <https://rapidminer.com/>, erişim tarihi; mayıs 2022.
- Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. A. ve Pendyala, R. M. (2017). Machine learning approaches for estimating commercial building energy consumption. *Applied Energy*, 208, 889–904. doi:10.1016/j.apenergy.2017.09.060
- Salari, M. ve Javid, R. J. (2017). Modeling household energy expenditure in the United States. *Renewable and Sustainable Energy Reviews*, 69, 822–832. doi:10.1016/j.rser.2016.11.183
- Singh, U., Rizwan, M., Alaraj, M. ve Alsaidan, I. (2021). A Machine Learning-Based Gradient Boosting Regression Approach for Wind Power Production Forecasting: A Step towards Smart Grid Environments. *Energies*, 14(16), 5196. doi:10.3390/en14165196
- Tso, G. K. F. ve Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9), 1761–1768. doi:10.1016/j.energy.2006.11.010
- Vringer, K., Aalbers, T. ve Blok, K. (2007). Household energy requirement and value patterns. *Energy Policy*, 35(1). doi:10.1016/j.enpol.2005.12.025
- Wang, F., Lu, X., Chang, X., Cao, X., Yan, S., Li, K., ... Catalão, J. P. S. (2022). Household profile identification for behavioral demand response: A semi-supervised learning approach using smart meter data. *Energy*, 238, 121728. doi:10.1016/j.energy.2021.121728
- Wang, W., Ji, T., Sun, J., Xiang, L., Xie, T. ve Xie, W. (2020). Prediction model of household appliance energy consumption based on machine learning Prediction of Sea Clutter Based on Recurrent Neural Network Prediction model of household appliance energy consumption based on machine learning. *Journal of Physics: Conference Series*, 1453, 12064. doi:10.1088/1742-6596/1453/1/012064
- Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S. ve Ahrentzen, S. (2018). Random Forest based hourly building energy prediction. *Energy and Buildings*, 171, 11–25. doi:10.1016/j.enbuild.2018.04.008
- Zhang, J., Li, X., Jiang, W., Wang, Y., Li, C., Wang, Q. ve Rao, S. (2005). A Novel Ensemble Decision Tree Approach for Mining Genes Coding Ion Channels for Cardiopathy Subtype (ss. 852–860). doi:10.1007/11540007_106.