



Araştırma Makalesi

**Journal of Innovative Engineering
and Natural Science**

(Yenilikçi Mühendislik ve Doğa Bilimleri Dergisi)

<https://dergipark.org.tr/en/pub/jiens>

Yüksek performanslı betonun basınç dayanımının farklı makine öğrenimi algoritmaları ile tahmin edilmesi

Muhammet Gökhan Altun^a, Ahmet Hakan Altun^b

^aÇanakkale Onsekiz Mart Üniversitesi, Mühendislik Fakültesi, İnşaat Mühendisliği, Çanakkale, 17100, Türkiye.

^bBursa Uludağ Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Bursa, 16059, Türkiye.

MAKELE BİLGİSİ

Makale Geçmişi:

Geliş 24 Eylül 2024

Düzeltilme 20 Aralık 2024

Kabul 20 Ocak 2025

Çevrimiçi mevcut

Anahtar Kelimeler:

Beton

Basınç dayanımı

Makine öğrenimi

ÖZET

Betonun basınç dayanımı, beton bileşenlerinin miktarları ve özellikleri, yaşı, ortam koşulları, deneysel koşullar gibi birçok faktörden etkilenmektedir. Betonun en önemli özelliği olan basınç dayanımının belirlenmesi amacıyla makine öğrenimi algoritmaları alternatif bir yöntem olarak kullanılmaktadır. Bu çalışmada, yüksek performanslı betonun basınç dayanımını tahmin etmek amacıyla 1030 satırlık açık veri seti üzerinde altı farklı makine öğrenimi modeli kullanılmıştır. Ayrıca mevcut veri setine türetilen yeni öznitelikler ilave edilerek betonun basınç dayanımını tahmin etme süreçlerindeki etkileri incelenmiştir. Bu bağlamda yeni özniteliklerin algoritmaların performansına olan katkısı değerlendirilmiş ve hangi algoritmaların en iyi sonuçları verdiği analiz edilmiştir. Elde edilen sonuçlara göre doğru tahmin etme yeteneği ve süre açısından en iyi sonucu XGBoost ve LightGBM algoritmaları göstermiştir. Buna ilaveten, veri setine iki yeni öznitelik daha eklenmesi kullanılan makine öğrenimi algoritmalarının doğru tahmin etme yeteneğini arttırmıştır.

Prediction of high performance concrete compressive strength with different machine learning algorithms

ARTICLE INFO

Article history:

Received 24 Sep 2024

Received in revised form 20 Dec 2024

Accepted 20 Jan 2025

Available online

Keywords:

Concrete

Compressive strength

Machine learning

ABSTRACT

Compressive strength of concrete is influenced by various factors including the amount and properties of concrete components, age, environmental conditions and experimental conditions. Machine learning algorithms are emerging as an alternative method for determining the compressive strength of concrete which is one of its most critical properties. In this study six different machine learning models were employed to predict the compressive strength of high-performance concrete using an open dataset of 1030 samples. Additionally the impact of incorporating newly derived features into the existing dataset on the prediction process was examined. The contribution of these new features to the performance of the algorithms was evaluated and the algorithms yielding the best results were analyzed. According to the results XGBoost and LightGBM demonstrated the best performance in terms of prediction accuracy and computational efficiency. Moreover, adding two new features to the dataset improved the predictive accuracy of the employed machine learning algorithms.

I. GİRİŞ

Dünya’da ve ülkemizde en çok kullanılan yapı malzemesi olan beton, temelde çimento, su, iri ve ince agregadan oluşmaktadır. Binalar, köprüler, barajlar, yollar, altyapı sistemleri başta olmak üzere hemen hemen her yapıda kullanılabilir. Betonun diğer yapı malzemelerine kıyasla bu kadar popüler olması, sahip olduğu birçok avantajdan kaynaklanmaktadır. Mukavemetinin yüksek olması, yüksek sıcaklığa karşı dayanıklı olması, zararlı

çevresel etkilere karşı dayanıklı olması ve ekonomik olması en önemli avantajları arasındadır [1, 2]. Öte yandan betonun tasarımı değiştirilerek veya farklı malzemeler ilave edilerek betonun birçok özelliği geliştirilebilmekte veya ihtiyaca yönelik özel betonlar üretilebilmektedir. Tek tip iri agregata tane çapı kullanılarak ve boşluk oranı yüksek tutularak geçirimsiz beton üretilebilmekte [3], çökme değeri sıfır olacak şekilde tasarım yapılarak silindir ile sıkıştırılabilen beton üretilebilmektedir [4]. Su azaltıcı katkının kullanımı ile sabit su/bağlayıcı oranında daha yüksek işlenebilirlik sağlanarak yüksek dayanımlı beton [5] ve kendiliğinden yerleşen beton [6] üretilebilmektedir. Öte yandan sürdürülebilir beton üretmek amacıyla uçucu kül, silis dumanı, cüruf gibi alternatif malzemeler çimento ile ikame edilmektedir. Söz konusu malzemeler betonun dayanımı, dayanıklılığı ve işlenebilirliği gibi bazı özelliklerini geliştirmektedir. Küresel CO₂ emisyonunun %5-7'sini oluşturan çimento endüstrisinde sağlanacak her bir tasarruf ekonomik avantajın yanında çevresel açıdan da önemli katkı sağlamaktadır [1, 7-9]. Ayrıca çeşitli lifler kullanılarak betonun dayanımı, tokluğu, darbe direnci gibi birçok mekanik özellikleri iyileştirilebilmektedir [10, 11].

Oldukça geniş kullanım alanına sahip beton teknolojisi gerek yeni malzemelerin geliştirilmesi gerekse farklı tekniklerin uygulanması ile her geçen gün gelişmeye ve genişlemeye devam etmektedir. Beton gibi karmaşık bir malzemenin özelliklerini belirlemek, bileşenleri ve çevresel koşullar başta olmak üzere tüm parametreler ve bunlar arasındaki etkileşimleri anlamak amacıyla doğrusal ve doğrusal olmayan regresyon, genetik programlama gibi farklı matematiksel modelleme teknikleri ile yüksek performans sergileyen farklı yapay zeka modelleri geliştirilmiştir. Doğrusal olmayan yapay zeka modelleri, bir sistemdeki girdi ve çıktı değişkenleri arasındaki karmaşık ilişkileri düşük hata oranlarında belirleyebilmektedir [12, 13].

Yeh [14] tarafından yapılan çalışmada, uçucu kül ve yüksek fırın cürufu içeren yüksek dayanımlı betonların basınç dayanımları regresyon analizi ve yapay sinir ağları (ANN) ile modellenmiştir. ANN ile yapılan modelin çok daha kesin sonuçlar verdiği belirlenmiştir. Çalışma kapsamında hazırlanan veriseti UCI Machine Learning Repository' de açık veri olarak yayınlanmaktadır. Aynı veriseti kullanılarak Nguyen-Sy ve diğ. [15] tarafından yapılan çalışmada, basınç dayanımını tahmin etmek amacıyla ANN haricinde destek vektör makineleri (SVM) ve aşırı gradyan artırma (XGB) makine öğrenimi modelleri kullanılmıştır. Elde edilen sonuçlara göre, XGB metodunda eğitirken geçen sürenin daha az, doğru tahminlerin ise daha yüksek olduğu tespit edilmiştir. Yine aynı veriseti kullanılarak yapılan bir diğer çalışmada, Nguyen ve diğ. [2] SVM, XGB, çok katmanlı algılayıcılar (MLP) ve gradyan artırma regresörü (GBR) modellerini kullanmış olup GBR ve XGB modellerinin daha iyi performans sergilediğini ifade etmişlerdir. Yörübulut ve diğ. [16] tarafından yapılan çalışmada, tahribatlı (karot) ve tahribatsız yöntemler (beton test çekici ve ultrasonik test cihazı) aracılığıyla veriler temin edilmiştir. Tahribatsız yöntemle temin edilen veriler kullanılarak ANN ve regresyon analizi ile betonun basınç dayanımı tahmin edilmiş ve karot numunelerinin basınç dayanımları ile oldukça benzer sonuçlar verdiği tespit edilmiştir.

Topçu ve diğ. [17] tarafından yapılan çalışmada, çelik gömülü uçucu kül içeren beton numunelere korozyon akımları verilerek veri kaydedici yardımıyla beş dakikalık aralıklarla ölçülmüştür. Korozyon akımları ANN kullanılarak modellenmiş, elde edilen sonuçlar ortalama karekök hata, ortalama mutlak yüzde hata ve korelasyon katsayısı kriterleri açısından karşılaştırılmıştır. Sonuç olarak ANN kullanılarak doğru modelleme sonuçlarının elde edilebileceği ifade edilmiştir. Boğa ve diğ. [18] tarafından yapılan çalışmada ise, öğütülmüş yüksek fırın cürufu ve kalsiyum nitrit bazlı korozyon önleyici kullanımının betonun mekanik ve dayanıklılık özellikleri üzerindeki etkileri incelenmiştir. Yapılan deneyler neticesinde elde edilen veriler ANN ve uygulamalı sinirsel bulanık çıkarım

sistemi (ANFIS) yöntemleri kullanılarak eğitilmiştir. Yöntemler, her parametre için diğer 42 numune ile test edilmiştir. Üretilen betonlarda kür sürelerinin uzatılması ve hava kürü yerine su kürü uygulanması betonun mekanik özelliklerinde ve klorür iyonu geçirgenliğinde önemli iyileşmeler sağlamıştır. Ayrıca deneysel verilerin ANN ve ANFIS modelleri aracılığıyla oldukça yakın bir ölçüde tahmin edilebileceği belirlenmiştir. Ofuyatan ve diğ. [19] tarafından yapılan çalışmada, silis dumanı ve polietilen tereftalat (PET) plastik atık kullanılarak üretilen kendiliğinden yerleşen betonların mekanik özellikleri tepki yüzeyi metodolojisi (RSM) ve ANN (Levenberge-Marquardt) modelleri ile tahmin edilmiştir. Hammoudi ve diğ. [20] tarafından yapılan çalışmada ise geri kazanılmış iri agrega içeren betonların 7, 28 ve 56 günlük basınç dayanımları ANN ve RSM modelleri kullanılarak tahmin edilmiştir. Her iki çalışmada da ANN modelinin RSM modeline kıyasla daha doğru tahminlerde bulunduğu bildirilmiştir.

Açıkgenç ve diğ. [21] tarafından yapılan çalışmada, çelik lifli betonun karışım oranlarını tahmin etmek amacıyla literatürden toplanan veriler kullanılarak ANN modeli geliştirilmiştir. Model altı girdi ve beş çıktıdan oluşmaktadır. Dayanım, kıvam sınıfı, maksimum agrega tane çapı, çelik lif kullanım oranı, çelik lif uzunluğu ve çelik lif çapı girdileri oluştururken, çimento dozajı, su miktarı, iri agrega miktarı, ince agrega miktarı ve kimyasal katkı miktarı ise çıktıları oluşturmaktadır. Sonuç olarak, ANN'nin çelik lifli betonların karışım oranlarını tahmin etmede güçlü bir potansiyele sahip olduğu, deneme karışımları ve zaman kaybı olmadan, istenen işlenebilirlik ve mekanik özelliklere sahip bir tasarımın mümkün olduğu ifade edilmiştir. Hossain ve diğ. [22] tarafından yapılan çalışmada, ultra yüksek çelik lifli betonun basınç dayanımı ANN ile tahmin edilmiştir. Veri seti %70 eğitim, %15 doğrulama ve %15 test kümesi olmak üzere üçe ayrılmıştır. Elde edilen sonuçlara göre bu yöntemin basınç dayanımı tahmininde kullanılabileceği ifade edilmiştir. Köksal ve diğ. [23] tarafından yapılan çalışmada ise çelik lifli betonda bulanık mantık çıkarım sistemi kullanılmış olup tahmin yeteneği yüksek bu yöntemin alternatif olarak kullanılabileceği belirtilmiştir.

Özetle, geleneksel veya özel betonlarda kullanılan malzemeler ile betonun dayanımı ve dayanıklılığı arasındaki ilişkileri belirlemek amacıyla çeşitli istatistiksel yöntemler ve yapay zeka modelleri kullanılmaktadır. Bu bağlamda yapılan çalışmalarda laboratuvar koşullarında üretilen, literatürden derlenen veya açık veri olarak yayımlanan verisetleri kullanılmaktadır. Doğru tahmin yeteneği en yüksek ve hata oranı en düşük modeli belirlemek amacıyla çeşitli yapay zeka ve makine öğrenimi modelleri kullanılmaktadır. Bu çalışmada, Yeh [14] tarafından yayınlanmış veri seti kullanılmıştır. Aynı veri setinin kullanıldığı çalışmalardan farklı olarak hem çeşitli algoritmalar kullanılmış hem de veri setindeki özniteliklere ilave yeni öznitelikler türetilerek tahminlemedeki etkisi belirlenmiş ve literatürdeki diğer çalışmalar ile kıyaslanmıştır. Çalışma kapsamında, rastgele orman (RF), destek vektör regresörü (SVR), XGBoost, adaptif arttırma (AdaBoost), kategorik arttırma (CatBoost) ve hafif gradyan arttırma (LightGBM) olmak üzere altı farklı makine öğrenimi modeli kullanılmıştır.

II. METODOLOJİ

Çalışma kapsamında eğitim ve değerlendirme süreçleri Google Colab platformu üzerinde gerçekleştirilmiştir. Colab çeşitli makine öğrenimi ve veri analizi görevleri için sunucu tabanlı bir ortam sağlayarak kullanıcıların merkezi işlemci birimi (CPU), grafik işlemci birimi (GPU) ve tensör işlemci birimi (TPU) kaynaklarına erişimini sağlamaktadır. Bu çalışmada Google Colab'in sağladığı sanal makinelerde kullanılan CPU tercih edilmiştir. Programlama dili olarak Python kullanılmıştır. Veri analizi iki farklı senaryo için gerçekleştirilmiştir. İlk

senaryoda Yeh [14] tarafından yayımlanan 8 girdi ve 1 çıktıdan oluşan veri seti kullanılmıştır. İkinci seneryoda iki farklı öznitelik daha ilave edilerek girdi sayısı 10'a çıkartılmıştır. Her iki durum için de 6 farklı makine öğrenimi algoritması kullanılarak tahminleme yapılmıştır.

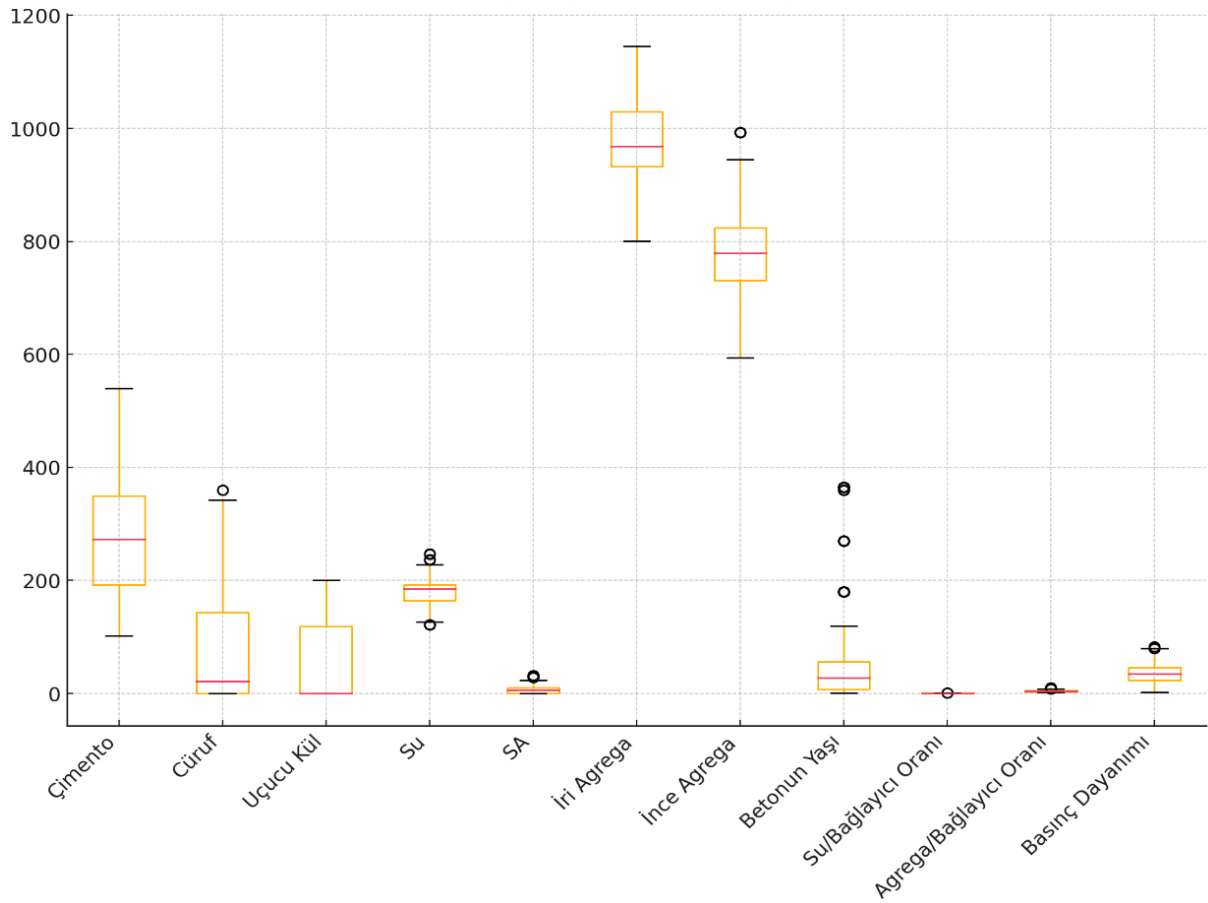
Çalışmada veriyi anlamlandırmak amacıyla ilk olarak istatistiksel veri analizi yapılmıştır. Ardından veriler standartlaştırılarak eğitim ve test kümesi olarak ikiye ayrılmıştır. Eğitim kümesi olarak belirlenen toplam verinin %80'i makine öğrenimi algoritmaları kullanılarak eğitilmiştir. Eğitilen modellerle test kümesi (%20) verileri tahmin edilmiştir. Makine öğrenimi modellerini, verileri eğitmek için kullanmadan önce ön işleme yapılmıştır. Öznitelikler tek tip bir ölçekte olmadığından, eğitim sürecinde bir veya birkaç özelliğin etkisinin fazla olmasını önlemek için standartlaştırma işlemi uygulanmıştır. Bu bağlamda, eğitimden önce öznitelikler ortalama 0 ve standart sapma 1 olacak şekilde standartlaştırılmıştır. Bu işlem verilerin her bir özelliğinin kendi ortalamasından sapmalarını ve standart sapma cinsinden değişimlerini dikkate alarak daha dengeli ve etkili bir eğitim süreci sağlamak amacıyla gerçekleştirilmiştir. Modeller normalleştirilmiş veriler kullanılarak eğitildikten sonra, çıktı özelliklerinin tahmine dayalı sonuçları, test aşamasında orijinal ölçüğe geri döndürülmüştür [2, 13].

2.1 Veri Seti ve İstatistiksel Analiz

Bu çalışmada, Yeh [14] tarafından yayınlanmış, UCI Machine Learning Repository'den alınan veri seti kullanılmıştır. Veri seti betonun basınç dayanımını tahmin etmek amacıyla çeşitli öznitelikleri içeren toplam 1030 satırdan oluşmaktadır. Her bir satır beton karışımının bileşenlerini ve betonun yaşını temsil eden sekiz temel özniteliği içermektedir. Bu öznitelikler çimento miktarı, cüruf miktarı, uçucu kül miktarı, su miktarı, süper akışkanlaştırıcı (SA) miktarı, iri agrega miktarı, ince agrega miktarı ve betonun yaşıdır. Beton karışım hesapları yapılırken içerisinde kullanılan bileşenlerin miktarlarına ilaveten su/bağlayıcı oranı ve agrega/bağlayıcı oranı da doğrudan dikkate alınan parametrelerdendir. Bundan dolayı veri setine mevcut öznitelikler kullanılarak hesaplanan ve betonun dayanımı ile doğrudan ilişkili olan su/bağlayıcı oranı ve agrega/bağlayıcı oranı olmak üzere iki yeni öznitelik daha eklenmiştir. Su/bağlayıcı oranı beton basınç dayanımı ile ters orantılı olup su miktarı arttıkça betonun dayanımı düşmektedir. Buna ilaveten agrega, bağlayıcı gibi beton bileşenlerinin miktarları ve birbirlerine oranları da betonun dayanımını önemli ölçüde etkileyen parametrelerdendir [24-26]. Su/bağlayıcı oranı, beton karışımındaki su miktarının bağlayıcı malzemelere (çimento, cüruf ve uçucu kül) oranını, agrega/bağlayıcı oranı ise, toplam agreganın (iri ve ince agrega) bağlayıcı malzemelere oranını ifade etmektedir. Bu ilave öznitelikler beton karışımlarının performansını daha hassas bir şekilde değerlendirmeyi mümkün kılmakta ve basınç dayanımını etkileyen faktörlere dair daha derinlemesine analizler yapma olanağı sağlamaktadır.

Veri setini eğitmeden önce kullanılacak verileri daha iyi anlamak ve analiz etmek amacıyla istatistiksel analiz yapılmıştır. Verilere ait istatistiksel dağılım ve aykırı değerler Şekil 1'de, verilerin minimum, maksimum, ortalama, standart sapma ve varyans değerleri ise Tablo 1'de verilmiştir. Tablo 1'den görüldüğü üzere bileşenler ağırlıkça kullanılmış olup veri setinde çimento miktarı 102-540 kg/m³, cüruf miktarı 0-359 kg/m³, uçucu kül miktarı 0-200 kg/m³, su miktarı 122-247 kg/m³, süper akışkanlaştırıcı miktarı 0-32 kg/m³, iri agrega miktarı 801-1145 kg/m³, ince agrega miktarı 594-993 kg/m³, betonun yaşı 1-365 gün, su/bağlayıcı oranı 0,24-0,90 ve agrega/bağlayıcı oranı 2,38-9,85 aralığındadır. Söz konusu bileşenler kullanılarak hazırlanan karışımların basınç dayanımları 2-83 MPa aralığındadır.

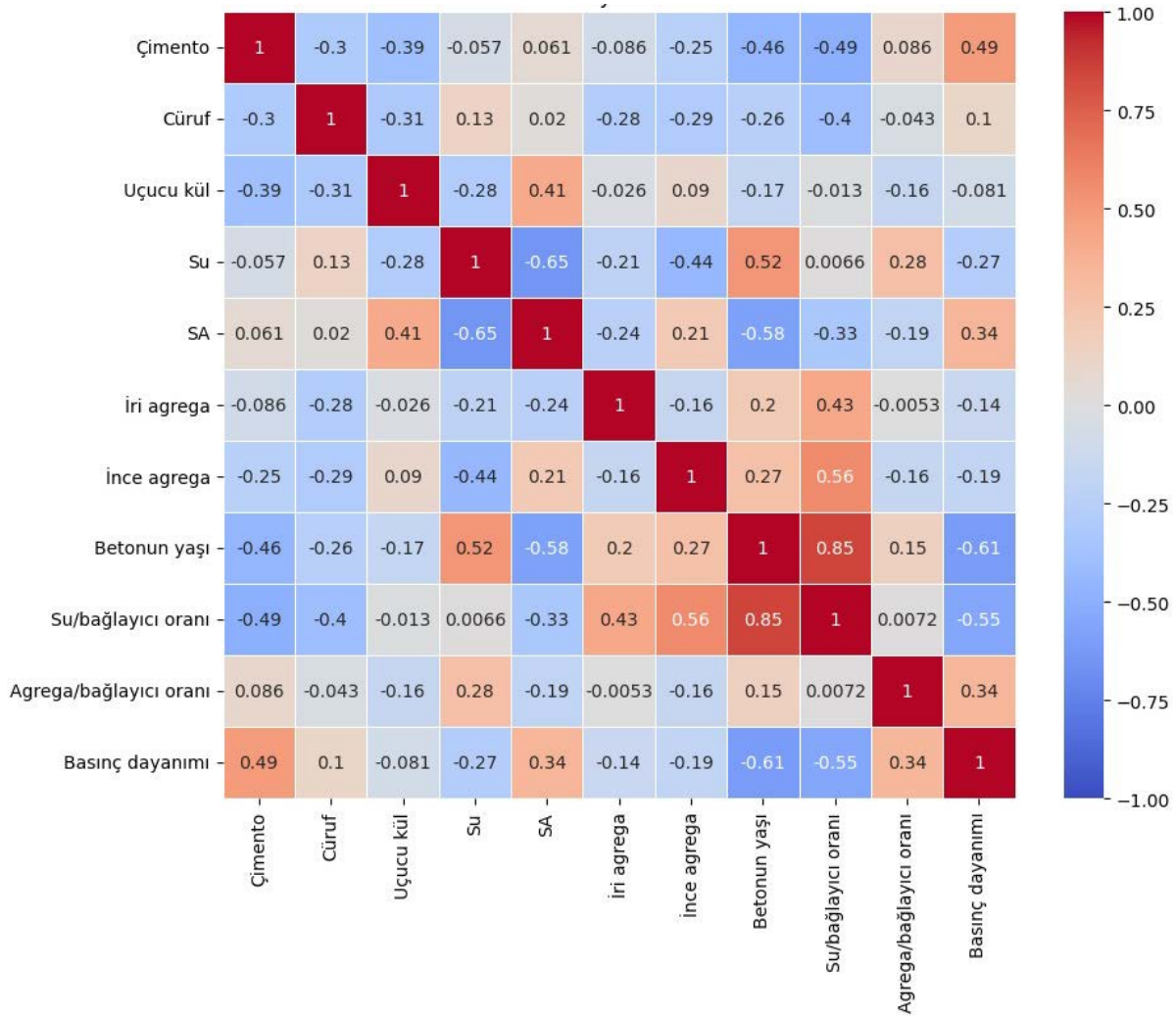
Veri setindeki tüm girdi ve çıktılar arasındaki ilişkileri daha iyi anlamak amacıyla özelliklerin korelasyonları analiz edilmektedir. Böylece bir özelliikle diğerleri arasındaki ilişki gözlemlenebilmekte ve kullanılacak tahmin modeli için fayda sağlanabilmektedir. Korelasyon katsayısı 0'a yaklaştıkça özellik çiftleri arasındaki ilişki zayıflamakta, -1 veya 1'e yaklaştığında ise güçlenmektedir. Diğer bir ifadeyle 0 değeri ilişkinin olmadığını, 1 değeri mükemmel derecede bir korelasyon olduğunu, -1 ise mükemmel derecede negatif bir korelasyon olduğunu göstermektedir [15, 27]. Örneğin su/bağlayıcı oranı ile betonun basınç dayanımı arasında ters orantılı bir ilişki söz konusudur. Karışımdaki su miktarı arttıkça betonun basınç dayanımı düşmektedir [24-26]. Betonun basınç dayanımı ile çimento miktarı arasında ise paralel bir ilişki söz konusudur. Karışımdaki çimento miktarı arttıkça sabit işlenebilirlikte betonun basınç dayanımı artmaktadır [24-26]. Veri setine ait korelasyon matrisi (ısı haritası) Şekil 2'de gösterilmiştir.



Şekil 1. Verilerin istatistiksel dağılımı ve aykırı değerler

Tablo 1. Veri setine ait istatistiksel detaylar

Öznitelik	Tanım	Min	Mak	Ortalama	Standart sapma	Varyans
Çimento (kg/m ³)	Girdi	102	540	279	104	10888
Cüruf (kg/m ³)	Girdi	0	359	72	86	7425
Uçucu kül (kg/m ³)	Girdi	0	200	56	64	4123
Su (kg/m ³)	Girdi	122	247	182	21	455
SA (kg/m ³)	Girdi	0	32	6	6	35
İri agregat (kg/m ³)	Girdi	801	1145	974	78	6019
İnce agregat (kg/m ³)	Girdi	594	993	773	80	6454
Betonun yaşı (gün)	Girdi	1	365	46	64	4062
Su/bağlayıcı oranı	Girdi	0,24	0,90	0,47	0,13	0,02
Agregat/bağlayıcı oranı	Girdi	2,38	9,85	4,56	1,23	1,52
Basınç dayanımı (MPa)	Çıktı	2	83	35	16	264



Şekil 2. Korelasyon matrisi (ısı haritası)

2.2 Makine Öğrenimi Algoritmaları

Bu çalışmada, betonun basınç dayanımını tahmin etmek amacıyla çeşitli makine öğrenimi algoritmaları kullanılmıştır. Sırasıyla her bir algoritmanın temel prensipleri ve uygulama detayları bu bölümde özetlenmiştir. SVR, veriyi en iyi şekilde modellemek için bir hiper düzlem bulmayı amaçlar. SVR hata terimlerini belirli bir eşğin altında tutarak modelin genelleme kapasitesini artırır. Bu çalışmada SVR'nin C ve ϵ hiperparametreleri kullanılarak çeşitli kombinasyonları test edilmiştir. Kernel fonksiyonu olarak lineer, polinomial ve Radial Basis Function (RBF) kernel seçenekleri değerlendirilmiştir [28, 29]. RF, birçok karar ağacının bir araya gelerek tahminlerde bulunmasını sağlayan bir topluluk yöntemidir. Her bir ağaç, eğitim veri setinin rastgele bir alt kümesi üzerinde eğitim alır ve final tahmini ağaçların tahminlerinin ortalaması veya çoğunluk oyuyla belirlenir. Bu yöntem modelin aşırı uyum riskini azaltır ve genellikle yüksek performans sağlar. Bu çalışmada farklı ağaç sayısı ve maksimum derinlik parametreleri ile modelin performansı değerlendirilmiştir [30]. XGBoost, yüksek performanslı bir gradyan artırma yöntemidir. XGBoost karar ağaçlarını ardışık olarak inşa ederken her bir ağacın önceki ağaçların hatalarını düzeltmesini sağlar. Bu yöntem hem doğruluk hem de hesaplama verimliliği açısından güçlüdür. Modelde öğrenme oranı, maksimum derinlik ve ağaç sayısı gibi hiperparametrelerin farklı kombinasyonları test edilmiştir [31, 32]. AdaBoost, zayıf öğrenicileri bir araya getirerek güçlü bir tahminci oluşturur. AdaBoost her bir zayıf öğrenicinin önceki öğrenicilerin hatalarını düzeltmeye odaklanmasını sağlar. Bu

çalışmada AdaBoost'un temel sınıflayıcı olarak karar ağaçları kullanarak model performansı değerlendirilmiştir. Ayrıca öğrenme oranı ve ağaç sayısı gibi hiperparametrelerin etkileri incelenmiştir [33, 34]. CatBoost, kategorik verilerle başa çıkma yeteneği ile bilinen bir gradyan artırma algoritmasıdır. CatBoost verileri kategorik özelliklerle en iyi şekilde işlemek için özel bir kodlama yöntemine sahiptir ve aşırı uyumu önlemek için çeşitli teknikler kullanır. CatBoost temel olarak sınıflandırma görevleri için geliştirilmiş olsa da gradyan artırma tabanlı yapısı sayesinde regresyon problemlerinde de etkili bir şekilde kullanılabilir. Bu çalışmada CatBoost'un öğrenme oranı, maksimum derinlik ve ağaç sayısı gibi hiperparametreleri ile modelin performansı gözlemlenmiştir [35, 36]. LightGBM, büyük veri setleriyle verimli ve hızlı çalışabilen bir gradyan artırma algoritmasıdır. LightGBM ağaçları histogram tabanlı bir yaklaşım kullanarak inşa etmekte ve bu sayede hem eğitim süresini kısaltmakta hem de bellek kullanımını optimize etmektedir [37, 38].

2.3 Değerlendirme Metrikleri

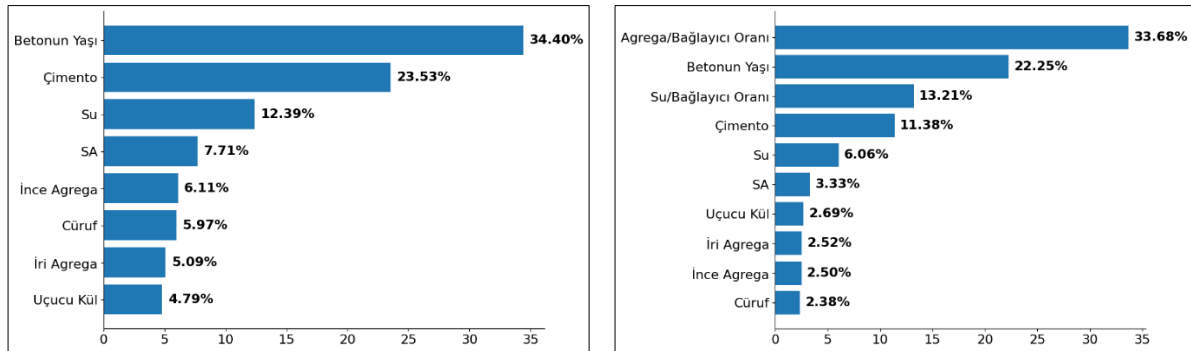
Çalışmada kullanılan makine öğrenimi algoritmalarının performanslarını değerlendirmek amacıyla doğrusal korelasyon katsayısı (R2), ortalama kare hata (MSE), ortalama karekök hatası (RMSE), ortalama mutlak hata (MAE) ve ortalama mutlak yüzde hatası (MAPE) metrikleri dikkate alınmıştır.

R2 değeri, modelin açıklama gücünü ölçen bir metriktir ve modelin hedef değişkenin varyansını ne kadar iyi açıkladığını belirtmektedir. Diğer bir ifade ile modelin veriye ne kadar iyi uyum sağladığının bir ölçüsüdür. Model performansını değerlendirmek için kullanılmaktadır. MSE değeri, tahmin edilen değerler ile gerçek değerler arasındaki farkların karelerinin ortalamasıdır. Modelin tahminlerindeki ortalama hata karelerini ve hataların büyüklüğünü ölçmek için kullanılmaktadır. RMSE değeri, tahmin edilen değerler ile gerçek değerler arasındaki farkların karelerinin ortalamasının kareköküdür. Tahminlerin ne kadar iyi olduğunu değerlendirmek için yaygın olarak kullanmakta ve modelin ortalama hata boyutunu standart bir birimde sunmaktadır. MAE değeri, tahmin edilen değerler ile gerçek değerler arasındaki mutlak farkların ortalamasıdır. Modelin tahminlerindeki ortalama mutlak hatayı ölçmekte ve her hata miktarını eşit şekilde değerlendirmektedir. MAPE değeri, tahmin edilen değerler ile gerçek değerler arasındaki mutlak yüzdelik farkların ortalamasıdır. R2 değerinin 1'e yakın olması modelin doğru tahmin yeteneğinin yüksek olduğu ifade etmektedir. Buna ilaveten MSE, RMSE, MAE ve MAPE hata değerlerinin düşük olması da daha iyi tahmin eden modele ulaşıldığını göstermektedir. Diğer bir ifade ile R2 ne kadar 1'e yakınsa ve MSE, RMSE, MAE ve MAPE hata oranları ne kadar düşüğe kullanılan modelin tahmini o kadar iyidir [13, 39, 40].

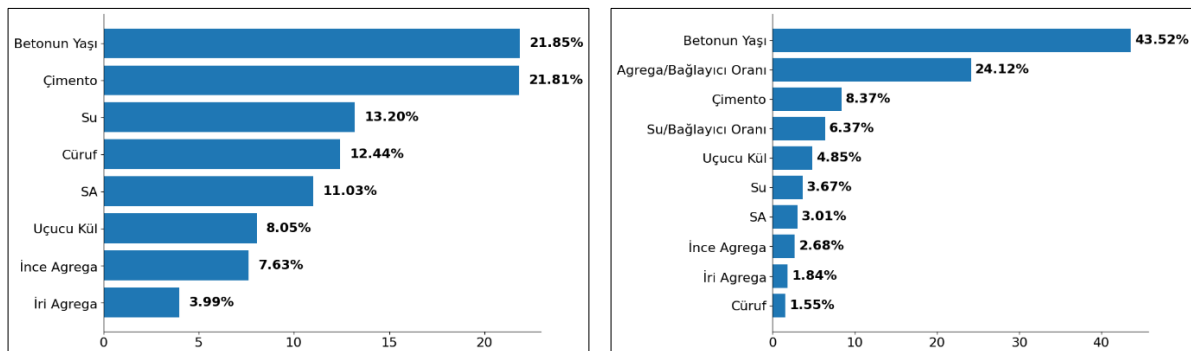
III. ANALİZ SONUÇLARI

Yüksek performanslı betonun basınç dayanımı, beton bileşenlerinden ve betonun yaşından oluşan veri seti kullanılarak farklı makine öğrenimi modelleri (RF, SVR, XGBoost, CatBoost, AdaBoost ve LightGBM) yardımıyla tahmin edilmiştir. Şekil 3-7'da sırasıyla her iki senaryo için de RF, XGBoost, CatBoost, AdaBoost ve LightGBM modellerine ait öznitelik önem grafikleri verilmiştir. Bu grafikler bir öznitelik modelin tahminleri üzerindeki etkisini göstermektedir. Öznitelik öneminin belirlenmesi, ağaç tabanlı yöntemlerde her bir düğümdeki bölünme iyileştirmelerinin ağırlıklı toplamını alarak ve bu toplamları tüm ağaçlar üzerinde ortalayarak yapılır. Bölünme iyileştirmesi, ağaç tabanlı yöntemlerde bir düğümde yapılan bölünmenin modelin tahmin performansına olan katkısını ölçen bir metriktir. Bu süreç her bir düğümdeki bölünme iyileştirmelerinin hesaplanmasını ve bu

iyileştirmelerin özelliklere göre toplanmasını içermektedir. Özellikle her bir düğümdeki bölünme iyileştirmesi düğümdeki safsızlık azalmasının ağırlıklı toplamı olarak hesaplanmaktadır. Bu iyileştirmeler tüm ağaçlar üzerinde ortalanarak her bir özelliğin önem skoru elde edilmektedir. Bu yöntem özneliklerin gerçek tahmin gücünü daha doğru bir şekilde yansıtmakta ve yanlışlığı azaltmaktadır [41]. Mevcut veri setinin kullanıldığı ilk senaryoda RF, XGBoost, CatBoost ve AdaBoost modellerinde ilk sıralarda betonun yaşı ve çimento yer alırken, LightGBM modelinde ise iri agrega ve çimento ilk sıralarda yer almıştır. Mevcut veri setine iki yeni özneliğin eklendiği ikinci senaryoda ise eklenen yeni öznelikler tüm modellerde üst sıralarda yer almıştır. Agregabağlayıcı oranı RF ve CatBoost modellerinde ilk sırada, XGBoost ve AdaBoost modellerinde ikinci sırada, LightGBM modelinde ise üçüncü sırada yer almıştır. Genel olarak grafiklerden su/bağlayıcı oranı ve agrega/bağlayıcı oranının tüm modellerde önemli bir role sahip olduğu anlaşılmaktadır. SVR modelinde ise doğrudan öznelik önemi skorlarını desteklemediği için grafiği çizdirilememiştir. Özneliklerin önemini dolaylı yoldan değerlendirmek amacıyla Recursive Feature Elimination (RFE) yöntemi kullanılmıştır. RFE, genellikle lineer kernel ile çalışan modellerde daha etkili sonuçlar verir. SVR modelinin kernel fonksiyonu GridSearchCV ile optimize edilirken en iyi sonuç rbf kernel ile elde edilmiştir. Bu nedenle, bu yöntem kullanılamamıştır [42].



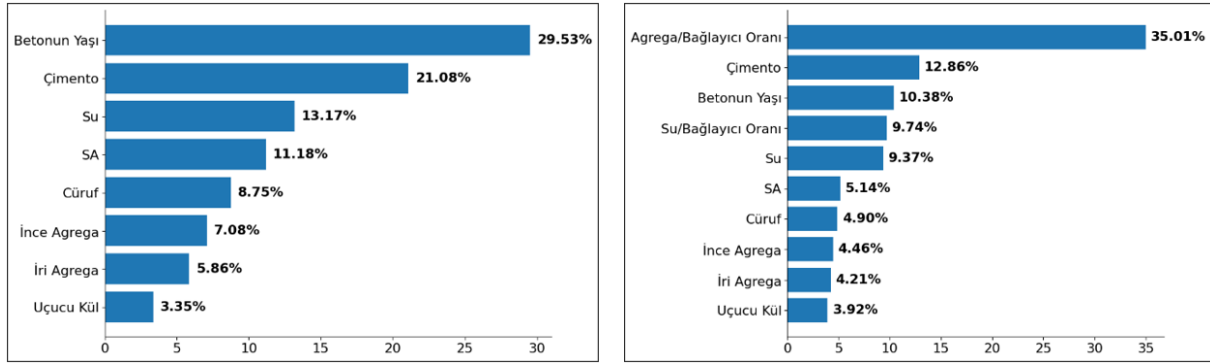
Şekil 3. RF modeline ait öznelik önem grafikleri



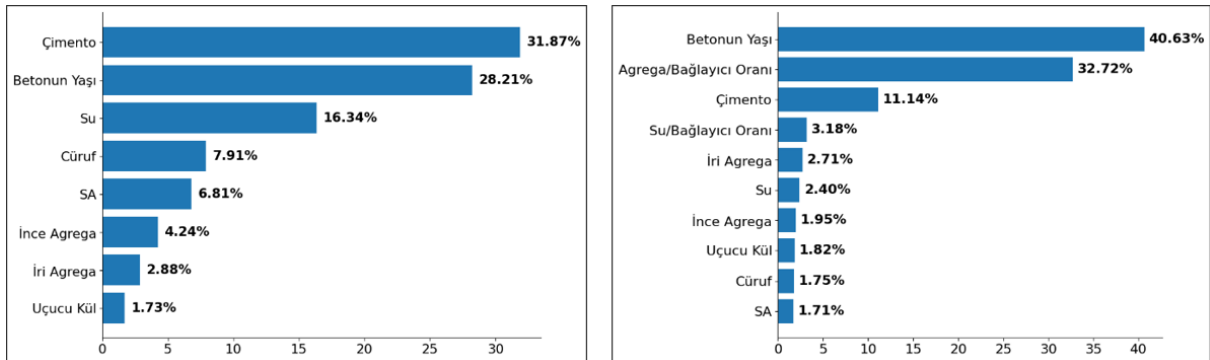
Şekil 4. XGBoost modeline ait öznelik önem grafikleri

Her iki senaryo üzerinde altı farklı makine öğrenimi algoritmasının regresyon sonuçları Şekil 8-13'te gösterilmiştir. Şekillerdeki her bir grafik ilgili algoritmanın tahmin ettiği değerlerin gerçek değerlere göre dağılımını göstermektedir. Grafiklere bakıldığında tüm algoritmaların tahmin değerleri gerçek değerlere yakın bir

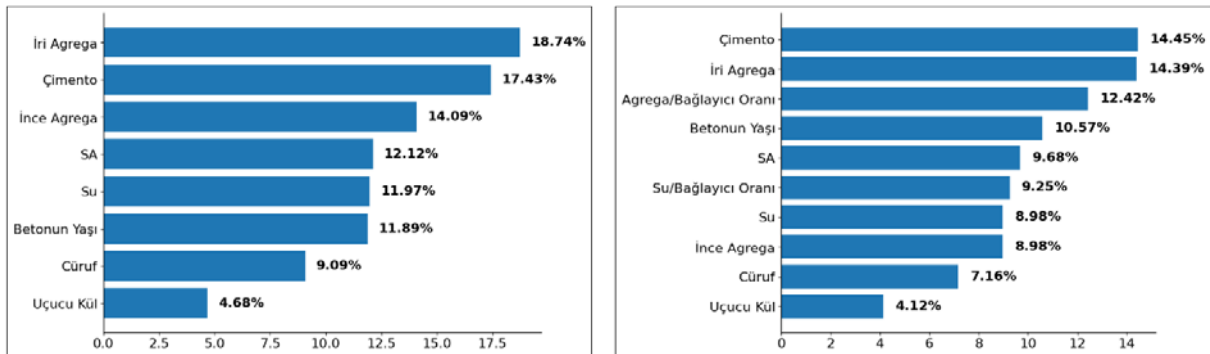
dağılım göstermektedir. Ancak XGBoost ve LightGBM algoritmaları diğer algoritmalara göre daha sıkı bir dağılım sergileyerek gerçek değerlere daha yakın tahminler üretmiştir. Özellikle bu iki algoritmada veri noktalarının büyük bir kısmı $x=y$ doğrusuna oldukça yakın bir konumda yer almaktadır. İkinci senaryolardaki dağılımlar incelendiğinde tüm modellerde ilk senaryoya göre aykırı değerlerde ufakta olsa azalmalar gözlemlenmiştir. Diğer bir ifadeyle yeni özneliklerin eklenmesi ile daha iyi sonuçlar elde edilmiştir. Bu durum bu algoritmaların modelin karmaşık yapısını daha iyi yakaladığını ve dolayısıyla daha iyi genelleme performansı sergilediğini göstermektedir.



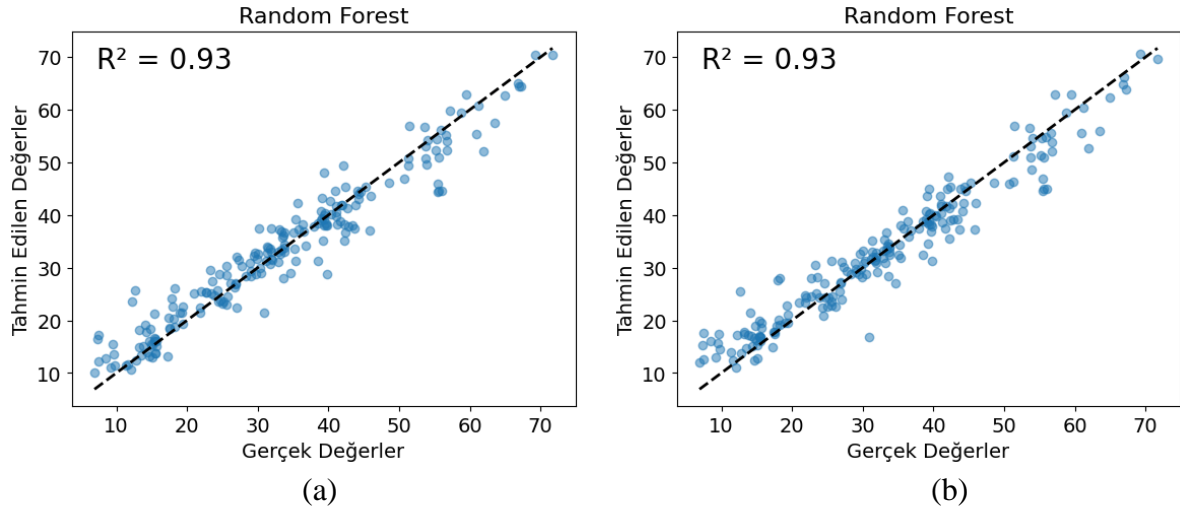
Şekil 5. CatBoost modeline ait öznelik önem grafikleri



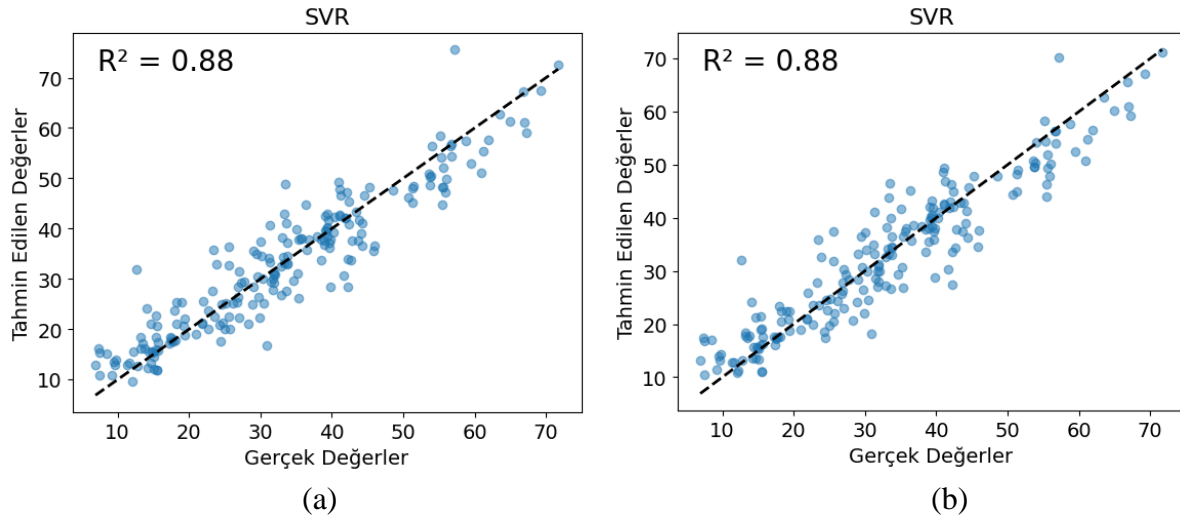
Şekil 6. AdaBoost modeline ait öznelik önem grafikleri



Şekil 7. LightGBM modeline ait öznelik önem grafikleri



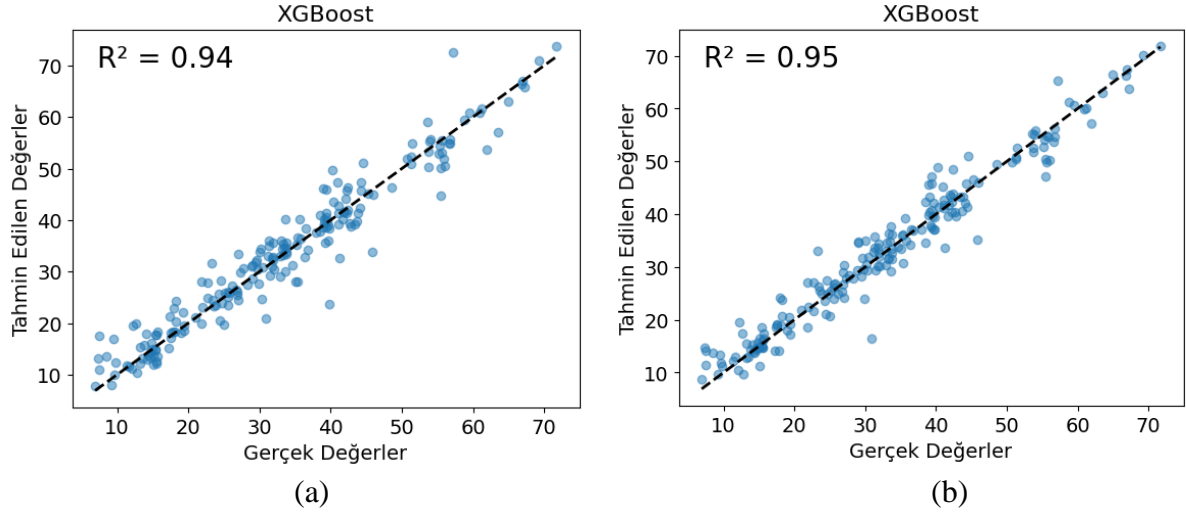
Şekil 8. RF modeline göre gerçek ile tahmin edilen değerler arasındaki ilişki; a) ilk veri seti, b) yeni özniteliklerin eklendiği veri seti



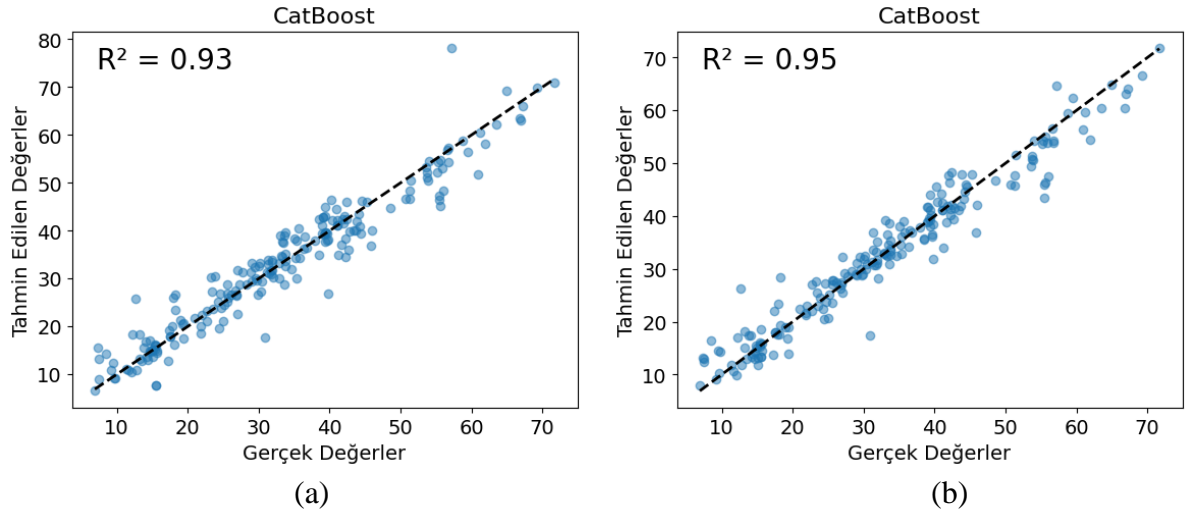
Şekil 9. SVR modeline göre gerçek ile tahmin edilen değerler arasındaki ilişki; a) ilk veri seti, b) yeni özniteliklerin eklendiği veri seti

Basınç dayanımı tahminlemede çalışma kapsamında kullanılan makine öğrenimi algoritmalarının performansları Tablo 2’de verilmiştir. Her iki senaryo içinde R2 değerleri 0,87 - 0,95 arasında, MSE değerleri 10,47 - 29,00 arasında, RMSE değerleri 3,23 - 5,39 arasında, MAE değerleri 2,31 - 4,23 arasında, MAPE değerleri ise 0,09 - 0,17 arasında çıkmış olup doğru tahmin etme yeteneği açısından en iyi sonuçları XGBoost ve LightGBM algoritmaları göstermiştir. En düşük performansı ise AdaBoost ve SVR algoritmaları göstermiştir. Bu algoritmaların başarısı geniş hiperparametre seçenekleriyle veri setinin karmaşıklığını anlamaları, gürültüye dayanıklılıkları ve büyük veri setlerinde hız avantajı sağlamalarından gelmektedir. XGBoost, paralel işlem desteği ve düzenleme (L1, L2) yetenekleriyle model karmaşıklığını dengede tutarken LightGBM, histogram tabanlı öğrenme ve bellek verimliliğiyle özellikle büyük veri setlerinde üstün performans sergilemektedir. Bu özellikler, her iki algoritmanın karmaşık ve büyük veri setlerinde daha iyi sonuçlar elde etmesini sağlar. Diğer yandan AdaBoost ve SVR, yapılarına bağlı kısıtlamalar nedeniyle daha düşük performans göstermiştir. AdaBoost, her iterasyonda yanlış sınıflandırılan örnekleri aşırı vurguladığı için gürültüye karşı hassastır. SVR ise kernel ve

hiperparametre ayarlarına oldukça bağımlıdır. Bu nedenle veri setindeki gürültü ve yüksek boyutluluk durumlarında performansları düşebilmektedir. Bu nedenle veri setinin yapısı ve algoritmanın özellikleri performans üzerinde belirleyici olmuştur [43-46].



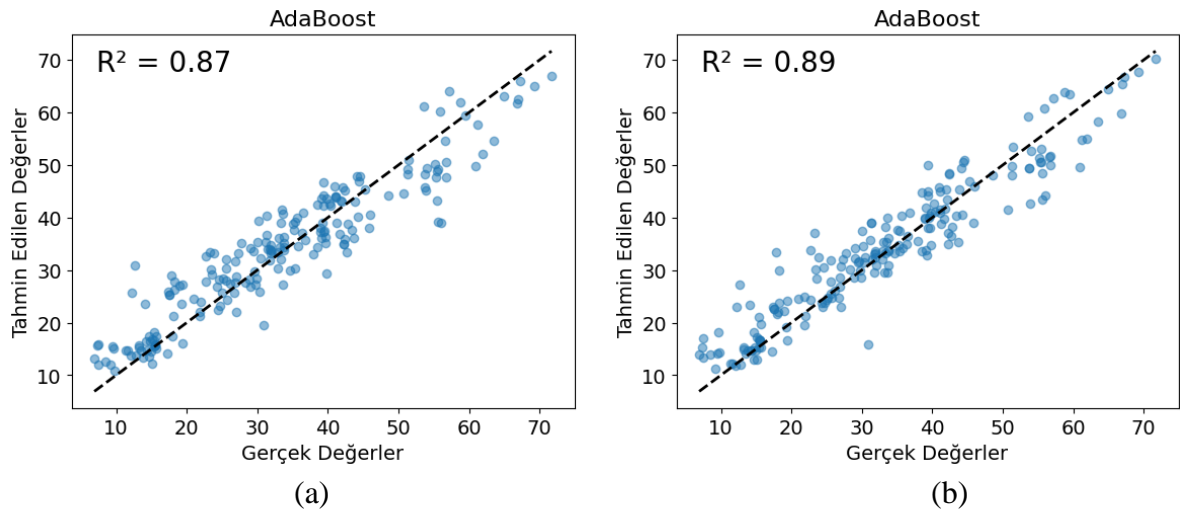
Şekil 10. XGBoost modeline göre gerçek ile tahmin edilen değerler arasındaki ilişki; a) ilk veri seti, b) yeni özneliklerin eklendiği veri seti



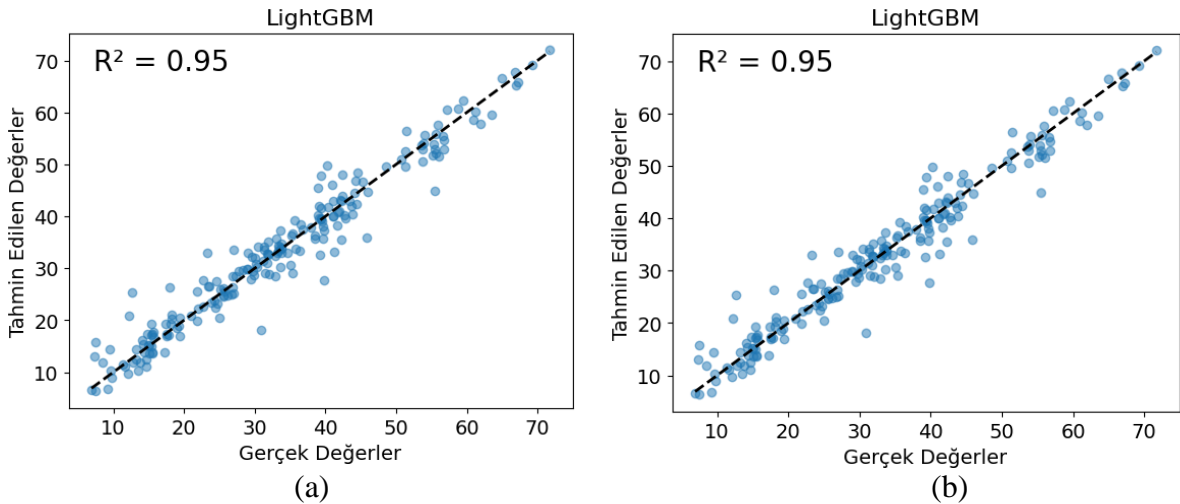
Şekil 11. CatBoost modeline göre gerçek ile tahmin edilen değerler arasındaki ilişki; a) ilk veri seti, b) yeni özneliklerin eklendiği veri seti

Zaman açısından bakıldığında her iki senaryo içinde analiz süreleri 1-59 dakika arasındadır. Süre açısından en iyi sonucu SVR, ardından sırasıyla AdaBoost, LightGBM, XGBoost ve CatBoost algoritmaları vermiştir. Analiz süresi en uzun olan algoritma ise RF algoritmasıdır. Gradyan artırma yöntemi kullanan algoritmalar her iterasyonda hatayı iyileştirmeye çalıştığından dolayı daha hızlı sürmektedir. RF modelinde, ağaçların sayısı arttıkça iyileşme görülmektedir. Yani kendi kendini optimize etmemektedir. Bundan dolayı analiz süresi uzayabilmektedir. XGBoost ve LightGBM, hata metrikleri (R^2 , MSE, RMSE, MAE, MAPE) ve işlem süresi açısından en iyi performansı göstermiştir. Her iki algoritma da en yüksek R^2 değerine (%0,95) ulaşmış ve en düşük hata oranlarını sağlamıştır. LightGBM, işlem süresi açısından daha hızlı çalışarak özellikle büyük veri setleri için

avantajlıdır. Diğer algoritmalarından AdaBoost, hatalı örneklere odaklanması nedeniyle daha yüksek hata oranlarına sahiptir; SVR, hızlı çalışmasına rağmen doğruluk oranı ve hata metriklerinde zayıf kalmıştır. CatBoost, hata metriklerinde XGBoost ve LightGBM'e yakın olsa da işlem süresi oldukça uzundur. Bu üstünlükler XGBoost ve LightGBM'in esnek hiperparametre ayarları, güçlü düzenleme mekanizmaları ve verimli hesaplama yöntemlerinden kaynaklanmaktadır [43-46]. Tüm sonuçlar değerlendirildiğinde süre ve hata metrikleri açısından en iyi sonucu XGBoost ve LightGBM algoritmaları göstermiştir. Nguyen ve diğ. [2] tarafından yapılan çalışmada aynı veriseti üzerinde SVM, XGBoost, MLP ve GBR olmak üzere dört farklı algoritma kullanılmış, GBR ve XGBoost modellerinin daha iyi performans sergilediği belirlenmiştir. Feng ve ark. [47] tarafından yapılan başka bir çalışmada SVM, ANN, RF ve XGBoost algoritmaları arasından en iyi sonucu XGBoost algoritmasının verdiği tespit edilmiştir.



Şekil 12. AdaBoost modeline göre gerçek ile tahmin edilen değerler arasındaki ilişki; a) ilk veri seti, b) yeni özniteliklerin eklendiği veri seti



Şekil 13. LightGBM modeline göre gerçek ile tahmin edilen değerler arasındaki ilişki; a) ilk veri seti, b) yeni özniteliklerin eklendiği veri seti

Tablo 2. Basınç dayanımı tahminlemede farklı makine öğrenimi algoritmalarının performansları

		RF	SVR	XGBoost	CatBoost	AdaBoost	LightGBM
R²	v1	0,93	0,88	0,94	0,93	0,87	0,94
	v2	0,93	0,88	0,95	0,95	0,89	0,95
MSE	v1	15,58	27,73	14,25	16,23	29,00	13,42
	v2	15,39	27,37	10,47	12,42	24,29	11,64
RMSE	v1	3,95	5,27	3,78	4,03	5,39	3,66
	v2	3,92	5,23	3,23	3,52	4,93	3,41
MAE	v1	2,84	3,92	2,63	2,77	4,23	2,60
	v2	2,82	3,90	2,31	2,45	3,71	2,38
MAPE (%)	v1	12	15	10	11	17	11
	v2	12	15	9	10	15	10
Zaman (sn)	v1	180	36	180	240	60	180
	v2	180	42	240	300	60	180

IV. SONUÇ

Bu çalışmada, yüksek performanslı betonun basınç dayanımını tahmin etmek amacıyla RF, SVR, XGBoost, CatBoost, AdaBoost ve LightGBM olmak üzere altı farklı makine öğrenme algoritması kullanılmıştır. Çalışma, aynı veri seti ile iki farklı senaryoda yapılmıştır. İlk senaryoda mevcut veri seti kullanılmış, ikinci senaryoda ise mevcut veri setindeki öznitelikler kullanılarak iki yeni öznitelik daha türetilmiştir. Her iki durumda da XGBoost ve LightGBM algoritmaları hem zaman açısından hem de yüksek doğruluk - düşük hata oranı açısından en iyi performansı göstermiştir. Veri setine iki yeni öznitelik daha eklenmesi kullanılan makine öğrenimi algoritmalarının doğru tahmin etme yeteneğini arttırmıştır.

Farklı makine ve derin öğrenme algoritmaları ile daha geniş veri setleri kullanılarak çalışmaların yapılması, çok daha hızlı ve doğru tahmin yeteneği olan uygulamaların geliştirilmesi, beton teknolojisi ve inşaat mühendisliği açısından oldukça önemlidir. Buna ilaveten topluluk öğrenme modelleri kullanılması da doğru tahmin yeteneğinin artırılması konusunda alternatif oluşturabilir. Betonun en önemli özelliği olan basınç dayanımının haricinde farklı mekanik ve durabilite özelliklerinin, yine benzer şekilde farklı özel beton türlerinin özelliklerinin makine öğrenimi algoritmaları ile tahmin edilmesine ihtiyaç bulunmaktadır.

Öte yandan Dünya’da yıllık tüketimi yaklaşık 35 milyar ton olan betonun sektör hacmi en büyük alanlardan birisidir. Beton tasarım, üretim ve kalite kontrol aşamalarını etkili ve verimli bir şekilde yönetmek, deneysel numune sayılarını minimize ederek doğal kaynak tüketimini azaltmak, iş gücü ve zamandan tasarruf ederek ve yine aynı şekilde standart sapmayı minimize ederek ekonomik kazanç sağlamak ve en nihayetinde insan kaynaklı hataları azaltarak ekonomik ve ekolojik kazançların haricinde büyük felaketlerde yaşanabilecek can ve mal kayıplarını engellenmek amacıyla yapay zeka ve istatistik tabanlı beton kalite tahminlemelerin yapılması oldukça önemlidir.

KAYNAKLAR

1. Kandiri A, Golafshani EM, Behnood A (2020) Estimation of the compressive strength of concretes containing ground granulated blast furnace slag using hybridized multi-objective ANN and salp swarm algorithm. Construction and Building Materials 248:118676. <https://doi.org/10.1016/j.conbuildmat.2020.118676>

2. Nguyen H, Vu T, Vo TP, Thai HT (2021) Efficient machine learning models for prediction of concrete strengths. *Construction and Building Materials* 266:120950. <https://doi.org/10.1016/j.conbuildmat.2020.120950>
3. AlShareedah O, Nassiri S (2021) Pervious concrete mixture optimization, physical, and mechanical properties and pavement design: A review. *Journal of Cleaner Production* 288:125095. <https://doi.org/10.1016/j.jclepro.2020.125095>.
4. Mardani-Aghabaglou A, Bayqra SH, Özen S, Altun MG, Faqiri ZA, Ramyar K (2020) Silindirle sıkıştırılmış beton karışımlarının tasarım yöntemleri ve yapılan çalışmalar. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi* 26(3):419-431. <https://doi.org/10.5505/pajes.2019.93530>.
5. Han Q, Gui C, Xu J, Lacidogna G (2019) A generalized method to predict the compressive strength of high-performance concrete by improved random forest algorithm. *Construction and Building Materials* 226:734-742. <https://doi.org/10.1016/j.conbuildmat.2019.07.315>.
6. Elemam WE, Abdelraheem AH, Mahdy MG, Tahwia AM (2020) Optimizing fresh properties and compressive strength of self-consolidating concrete. *Construction and Building Materials* 249:118781. <https://doi.org/10.1016/j.conbuildmat.2020.118781>.
7. Mardani-Aghabaglou A, Tuyan M, Yılmaz G, Arıöz Ö, Ramyar K (2013) Effect of different types of superplasticizer on fresh, rheological and strength properties of self-consolidating concrete. *Construction and Building Materials* 47:1020-1025. <https://doi.org/10.1016/j.conbuildmat.2013.05.105>
8. Mardani-Aghabaglou A, Sezer Gİ, Ramyar K (2014) Comparison of fly ash, silica fume and metakaolin from mechanical properties and durability performance of mortar mixtures view point. *Construction and Building Materials* 70:17-25. <https://doi.org/10.1016/j.conbuildmat.2014.07.089>
9. Aydogmus HY, Erdal HI, Karakurt O, Namli E, Turkan YS, Erdal H (2015) A comparative assessment of bagging ensemble models for modeling concrete slump flow. *Computers and Concrete*. 16(5):741-757. <https://doi.org/10.12989/cac.2015.16.5.741>.
10. Banthia N, Sheng J (1996) Fracture toughness of micro-fiber reinforced cement composites. *Cement and Concrete Composites* 18:251-269. [https://doi.org/10.1016/0958-9465\(95\)00030-5](https://doi.org/10.1016/0958-9465(95)00030-5).
11. Altun MG, Oltulu M (2020) Effect of different types of fiber utilization on mechanical properties of recycled aggregate concrete containing silica fume. *Journal of Green Building* 15(1):119-136. <https://doi.org/10.3992/1943-4618.15.1.119>.
12. Güneyisi E, Gesoglu M, Özbay E (2009) Evaluating and forecasting the initial and final setting times of self-compacting concretes containing mineral admixtures by neural network. *Materials and Structures* 42:469-484. <https://doi.org/10.1617/s11527-008-9395-5>.
13. Avcı E, Altun MG, (2023) Betonun çökme ve basınç dayanımının makine öğrenmesi modelleri kullanılarak tahmin edilmesi. *ICSHSR 4th International Conference on Health, Engineering and Applied Sciences*.
14. Yeh IC (1998) Modeling of strength of high-performance concrete using artificial neural networks. *Cement and Concrete Research* 28:1797-1808. [https://doi.org/10.1016/S0008-8846\(98\)00165-3](https://doi.org/10.1016/S0008-8846(98)00165-3).
15. Nguyen-Sy T, Wakim J, To QD, Vu MN, Nguyen TD, Nguyen TT (2020) Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. *Construction and Building Materials* 260:119757. <https://doi.org/10.1016/j.conbuildmat.2020.119757>.
16. Yörübulut S, Dogan O, Erdugan F, Yörübulut S (2019) Tahribatsız Yöntem Verileri Kullanılarak Yapay Sinir Ağı ve Regresyon Yöntemi ile Beton Basınç Dayanımının Tahmin Edilmesi. *International Journal of Engineering Research and Development* 12(2):769-776. <https://doi.org/10.29137/umagd.734655>
17. Topçu İB, Boğa AR, Hocaoğlu FO (2009) Modeling corrosion currents of reinforced concrete using ANN. *Automation in Construction* 18(2):145-152. <https://doi.org/10.1016/j.autcon.2008.07.004>.
18. Boğa AR, Öztürk M, Topçu İB (2013) Using ANN and ANFIS to predict the mechanical and chloride permeability properties of concrete containing GGBFS and CNI. *Composites:Part B* 45:688-696. <https://doi.org/10.1016/j.compositesb.2012.05.054>.
19. Ofuyatan OM, Agbawhe OB, Omole DO, Igwegbe CA, Ighalo JO (2022) RSM and ANN modelling of the mechanical properties of self compacting concrete with silica fume and plastic waste as partial constituent replacement. *Cleaner Materials* 4:100065. <https://doi.org/10.1016/j.clema.2022.100065>.
20. Hammoudi A, Moussaceb K, Belebchouche C, Dahmoune F (2019) Comparison of artificial neural network (ANN) and response surface methodology (RSM) prediction in compressive strength of recycled concrete aggregates. *Construction and Building Materials* 209:425-436. <https://doi.org/10.1016/j.conbuildmat.2019.03.119>.
21. Açikgenç M, Ulaş M, Alyamaç KE (2015) Using an artificial neural network to predict mix compositions of steel fiber-reinforced concrete. *Arabian Journal for Science and Engineering* 40:407-419. <https://doi.org/10.1007/s13369-014-1549-x>
22. Hossain MM, Uddin MN, Hossain MAS (2023) Prediction of compressive strength ultra-high steel fiber reinforced concrete (UHFRFC) using artificial neural networks (ANNs). *Materials Today: Proceedings*. <https://doi.org/10.1016/j.matpr.2023.02.409>

23. Köksal F, Şahin Y, Beycioğlu A, Gencel O, Brostow W (2012) Estimation of fracture energy of high-strength steel fibre-reinforced concrete using rule-based Mamdani-type fuzzy inference system. *Science and Engineering of Composite Materials* 19(4):373-380. <https://doi.org/10.1515/secm-2012-0017>
24. Neville AM (1997) *Properties of Concrete*. London: Wiley
25. Mehta PK, Monteiro P (1997) *Concrete: microstructure, properties, and materials*. McGraw-Hill Publishing.
26. Baradan B, Türkel S, Yazıcı H, Ün H, Yiğiter H, Felekoğlu B, Tosun K, Aydın S, Yardımcı MY, Topal A Öztürk AU (2012) *Beton*. Dokuz Eylül Üniversitesi Mühendislik Fakültesi No:334, İzmir, Türkiye.
27. Daemen J, Rijmen V, Daemen J, Rijmen V (2020). Correlation matrices. *The design of Rijndael: the advanced encryption standard (AES)*, 91-113.
28. Awad M, Khanna R, Awad M, Khanna R (2015) Support vector regression. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, pp. 67-80.
29. Parbat D, Chakraborty M (2020) A python based support vector regression model for prediction of COVID19 cases in India. *Chaos, Solitons & Fractals*, 138, 109942.
30. Biau G, Scornet E (2016) A random forest guided tour. *Test*, 25:197-227.
31. Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).
32. Brownlee J (2016) *XGBoost With python: Gradient boosted trees with XGBoost and scikit-learn*. Machine Learning Mastery.
33. Collins M, Schapire RE, Singer Y (2002) Logistic regression, AdaBoost and Bregman distances. *Machine Learning* 48:253-285.
34. Gao R, Liu Z (2020) An improved adaboost algorithm for hyperparameter optimization. In *Journal of Physics: Conference Series* 1631(1), 012048. IOP Publishing.
35. Beskopylny AN, Stel'makh SA, Shcherban EM, Mailyan LR, Meskhi B, Razveeva I, Beskopylny N (2022) Concrete strength prediction using machine learning methods CatBoost, k-nearest neighbors, support vector regression. *Applied Sciences* 12(21), 10864. <https://doi.org/10.3390/app122110864>.
36. Xiang W, Xu P, Fang J, Zhao Q, Gu Z, Zhang Q (2022) Multi-dimensional data-based medium-and long-term power-load forecasting using double-layer CatBoost. *Energy Reports* 8:8511-8522. <https://doi.org/10.1016/j.egy.2022.06.063>.
37. Al-Kasassbeh M, Abbadi MA, Al-Bustanji AM (2020) LightGBM algorithm for malware detection. In *Intelligent Computing: Proceedings of the 2020 Computing Conference* 3:391-403. Springer International Publishing.
38. Ju Y, Sun G, Chen Q, Zhang M, Zhu H, Rehman MU (2019) A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. *Ieee Access* 7:28309-28318. <https://doi.org/10.1109/ACCESS.2019.2901920>
39. Tatachar AV (2021) Comparative assessment of regression models based on model evaluation metrics. *International Journal of Innovative Technology and Exploring Engineering* 8(9):853-860.
40. Chicco D, Warrens MJ, Jurman G (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj Computer Science* 7, e623. <https://doi.org/10.7717/peerj-cs.623>
41. Zhou Z, Hooker G (2021) Unbiased measurement of feature importance in tree-based methods. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15(2):1-21.
42. Jeon H, Oh S (2020) Hybrid-recursive feature elimination for efficient feature selection. *Applied Sciences* 10(9), 3211. <https://doi.org/10.3390/app10093211>
43. Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp 785-794).
44. Chen C, Zhang Q, Ma Q, Yu B (2019) LightGBM-PPI: Predicting protein-protein interactions through LightGBM with multi-information fusion. *Chemometrics and Intelligent Laboratory Systems* 191:54-64. <https://doi.org/10.1016/j.chemolab.2019.06.003>.
45. Solomatine DP, Shrestha DL (2004) AdaBoost. RT: a boosting algorithm for regression problems. In *2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541)* (Vol 2, pp 1163-1168). IEEE.
46. Awad M, Khanna R, Awad M, Khanna R (2015) Support vector regression. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*. Springer Nature. pp 67-80.
47. Feng J, Zhang H, Gao K, Liao Y, Gao W, Wu G (2022) Efficient creep prediction of recycled aggregate concrete via machine learning algorithms. *Construction and Building Materials* 360,129497. <https://doi.org/10.1016/j.conbuildmat.2022.129497>.