



## Arařtırma Makalesi

# GELİŐMİŐ ÖZELLİK MÜHENDİSLİĐİ VE MAKİNE ÖĐRENİMİ İLE İSTANBUL DEPREMİ RİSK ANALİZİ

Hüseyin PARMAKSIZ<sup>1</sup>, , , Zeynep ÖZER<sup>1</sup>, 

<sup>1</sup> Bilecik Őeyh Edebalı Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Biliřim Sistemleri, Bilecik, Türkiye

 Sorumlu Yazar: [huseyin.parmaksiz@bilecik.edu.tr](mailto:huseyin.parmaksiz@bilecik.edu.tr)

Geliř tarihi: 26/09/2024

Kabul tarihi: 10/11/2024

**Özet:** İstanbul'un coĐrafi konumu onu aktif fay hatları üzerinde konumlandırmakta ve depreme yatkın hale getirmektedir. Bölgenin jeolojik özellikleri ve nüfus yoğunluĐu, olası bir depremin sosyal ve ekonomik sonuçlarını artırmakta ve deprem riskinin önlenmesi ve yönetimi için kapsamlı bir strateji gerektirmektedir. Tarihte görüldüĐü üzere, 1509 yılında meydana gelen deprem "Kıyamet-i SuĐra" olarak bilinmektedir. Bu depremde 4000 ila 13000 kiři ölmüő, 10000'den fazla kiři yaralanmıő ve Osmanlı İmparatorluĐu'nun başkenti Konstantinopolis'te yaklaşık 1070 ev yıkılmıőtır. Daha yakın bir tarihte, 1999 Kocaeli depreminde İstanbul'da çok sayıda yapı hasar görmüő ve çok sayıda insan hayatını kaybetmiőtir. Bu çalıőma, Büyük İstanbul Depremi'ndeki olası can kaybını deĐerlendirmek için makine öĐrenimi tekniklerini kullanmaktadır. Tahminlere göre, deprem sırasında en fazla can kaybının hangi bölgelerde yaőanacaĐının belirlenmesi, bu bölgelerde önceden önlem alınması, olası İstanbul depremi için risk yönetim planlarının oluőturulması ve acil durum taktiklerinin geliőtirilmesi amaçlanmaktadır. Çalıőmada Rastgele Orman Regresyonu (RF), Doğrusal Regresyon (LR), AŐırđ Gradyan Artırma (XGBoost), Destek Vektör Makinesi (SVM), Karar AĐacı Regresyonu (DT), Ekstra AĐaçlar (ET), Kategorik Artırma (CatBoost) ve Kısıtlı AŐırđ ÖĐrenme Makinesi (CELM) gibi tahmin modelleri kullanılmıőt ve aralarındaki iliőkiler incelenmiőtir. LR modelinin kullanılan diĐer modellere kıyasla daha etkin olduĐu gözlemlenmiőtir. mRMR, Boruta ve KarŐılıklı Bilgi (MI) yöntemleri özelliklerin daha verimli sečilmesi için kullanılmıőtir. Karar mekanizmalarını aydınlatmak için Açıklanabilir Yapay Zeka olarak Shapley Toplamalı Açıklamalar (SHAP) analizi kullanılmıőtir. Çalıőma, deprem riskine karŐı afet yönetimi stratejileri ve önlemlerine odaklanarak İstanbul'da can ve mal kaybını en aza indirecek model ve yöntemleri deĐerlendirmektedir.

**Anahtar Sözcükler:** İstanbul deprem analizi, makine öĐrenmesi, özellik mühendisliĐi, açıklanabilir yapay zekâ

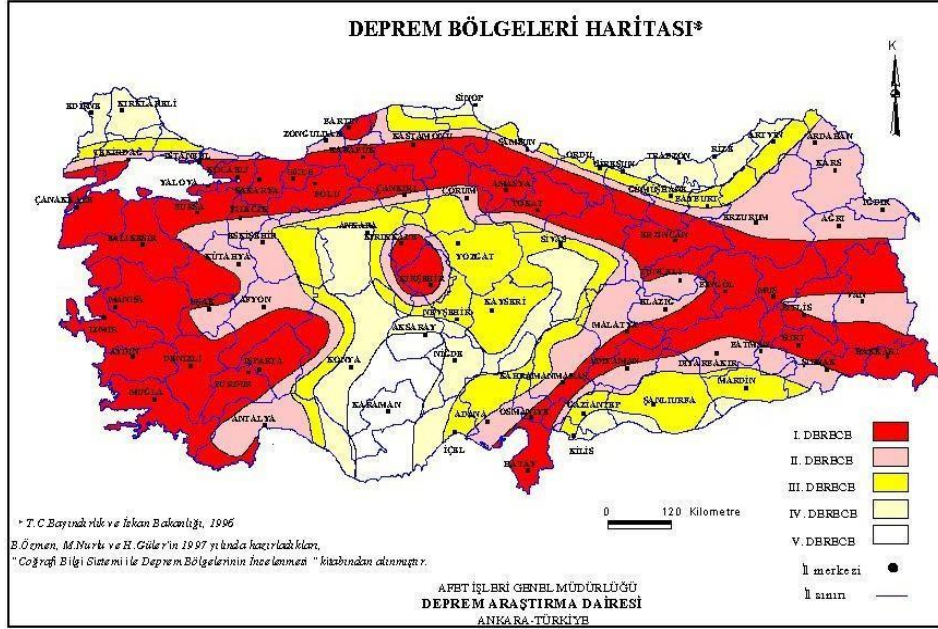
## ISTANBUL EARTHQUAKE RISK ANALYSIS WITH ADVANCED FEATURE ENGINEERING AND MACHINE LEARNING

**Abstract:** Istanbul's geographical location places it on active fault lines, making it prone to earthquakes. The geological attributes and population density of the region amplify the social and economic consequences of a potential earthquake, demanding a comprehensive strategy for earthquake risk prevention and management. As seen in history, the earthquake in 1509 is known as the "Little Apocalypse." This earthquake killed between 4000 and 13000 people, injured over 10000, and destroyed around 1070 homes in Constantinople, the Ottoman Empire's capital. More recently, during the 1999 Kocaeli earthquake, numerous structures in Istanbul were damaged, and many people died. This study uses machine learning techniques to assess the possible loss of life in the Great Istanbul Earthquake. According to the forecasts, the goal is to determine which places will have the greatest loss of life during the earthquake, to take safeguards in these areas ahead of time, to construct risk management plans for the potential Istanbul earthquake, and to develop emergency tactics. The study used prediction models such as Random Forest Regression, Linear Regression, XGBoost, SVM, Decision Tree Regression, Extra Trees, CatBoost, and Constrained ELM, and examined their interrelationships. It is observed that the LR model is more efficient than the other models used. mRMR, Boruta, and Mutual Information (MI) methods were used to pick features more efficiently. SHapley Additive ExPlanations (SHAP) analysis as Explainable Artificial Intelligence (XAI) was used to elucidate the decision mechanisms. The study focuses on disaster management strategies and measures against earthquake risk and evaluates models and methods to minimize the loss of life and property in Istanbul.

**Keywords:** Istanbul earthquake analysis, machine learning, feature engineering, XAI

## 1. Giriş

Deprem, yer kabuğundaki hızlı kırılmalar sonucu açığa çıkan enerjinin sismik dalgalar yoluyla yayılması ve dünyayı sarsmasıyla oluşmaktadır. Türkiye'nin aktif fay hatları üzerinde yer alması ve uzun bir deprem geçmişine sahip olması, onu depreme eğilimli bir bölge haline getirmektedir (Kundak ve Türkoğlu, 2007). Türkiye'nin jeolojik yapısı Dördüncü Jeolojik Zaman (Kuvaterner) boyunca gelişmiştir. Türkiye, Arap, Avrasya ve Afrika plakalarının birleştiği aktif bir tektonik bölgede yer almakta ve bu da ülkeyi depremlere açık hale getirmektedir. Bu levhaların hareketleri yer kabuğunda gerilmelere neden olarak fay hattı kırılmalarına ve depremlere yol açmaktadır. Arap ve Avrasya plakalarının çarpışması Türkiye'yi depremlere karşı özellikle savunmasız hale getirmiştir. Sismik tehlikeler ülkenin ekonomik ve sosyal yaşamı için ciddi sonuçlar doğurabileceğinden, özellikle İstanbul gibi büyük kent merkezleri için kapsamlı bir risk değerlendirmesi kritik önem taşımaktadır (Eyidoğan v.d., 2010). İstanbul, Kuzey Anadolu Fay Hattı (KAF), Çınarcık Fay Hattı ve Tekirdağ Fay Hattı da dahil olmak üzere birçok fay hattının kesiştiği noktada yer almaktadır. Şehir, zengin coğrafi tarihi ve köklü geçmişiyle birçok medeniyete ev sahipliği yapmıştır. Şekil 1'de Türkiye'nin sismik fay hatları verilmektedir.



Şekil 1. Türkiye'deki Deprem Fay Hatları (Magdeburger Sigorta, 2024)

Depremler, büyük hasara ve can kaybına yol açarak acil durum müdahale sürecinde ölümlerin hızlı bir şekilde değerlendirilmesini gerekli kılmaktadır. Bu bağlamda, ölüm sayılarının hızlı değerlendirilmesi ve tahminlerin dinamik olarak güncellenmesi, etkin bir acil durum karar verme süreci için kritik öneme sahiptir. Liu, Wang ve Zhang (2024), 6 Şubat 2023'te meydana gelen MS 8.0 ve MS 7.9 büyüklüğündeki Kahramanmaraş depremlerini ele alarak, AFAD tarafından bildirilen ölüm sayılarına dayalı olarak Hızlı Kaba Tahmin (QRE) yapmış ve bu tahminleri yeni veriler ışığında dinamik bir şekilde revize etmişlerdir. Bu çalışmada kullanılan Yarı Doğrusal ve Uyarlanabilir Tahmin (QLAE) yöntemi, nihai ölüm sayısını hızlı bir şekilde uyarlayarak sonraki ölümleri tahmin edebilmekte ve depremin ardından yalnızca 13 saat içinde doğruluğu yüksek tahminler sunabilmektedir. Bu tür adaptif tahmin yaklaşımlarının başarıyla uygulanması, İstanbul'da beklenen büyük deprem gibi olası afetlerin etkilerini azaltma noktasında da faydalı olabilir.

İstanbul ve çevresindeki deprem tehlikesi, tarihsel ve sismolojik veriler kullanılarak kapsamlı bir şekilde değerlendirilmektedir (Eyidoğan, 2010). Marmara Bölgesi'nin sismik aktivitesi, 1894 yılında İzmit Çınarcık Havzası'nda meydana gelen 7.0 büyüklüğündeki depremle başlayarak önemli bir düzeye ulaşmıştır. Bunun yanı sıra, 1999 Kocaeli/Gölcük depremi de bölgeye ciddi zararlar vermiştir. Kuzey Anadolu Fay Hattı (KAF) üzerindeki bu depremler, yer kabuğunda biriken enerjinin

kritik bir seviyeye ulaştığını ve yeni bir büyük deprem olasılığının arttığını göstermektedir. Liu, Wang ve Zhang (2024) tarafından geliştirilen tahmin yöntemlerinin İstanbul'da beklenen büyük depremin etkilerini öngörmek amacıyla kullanılması, sismolojik verilerin ışığında bölgenin deprem potansiyelinin daha iyi anlaşılmasına ve afet öncesi hazırlık süreçlerinin güçlendirilmesine katkı sağlayabilir.

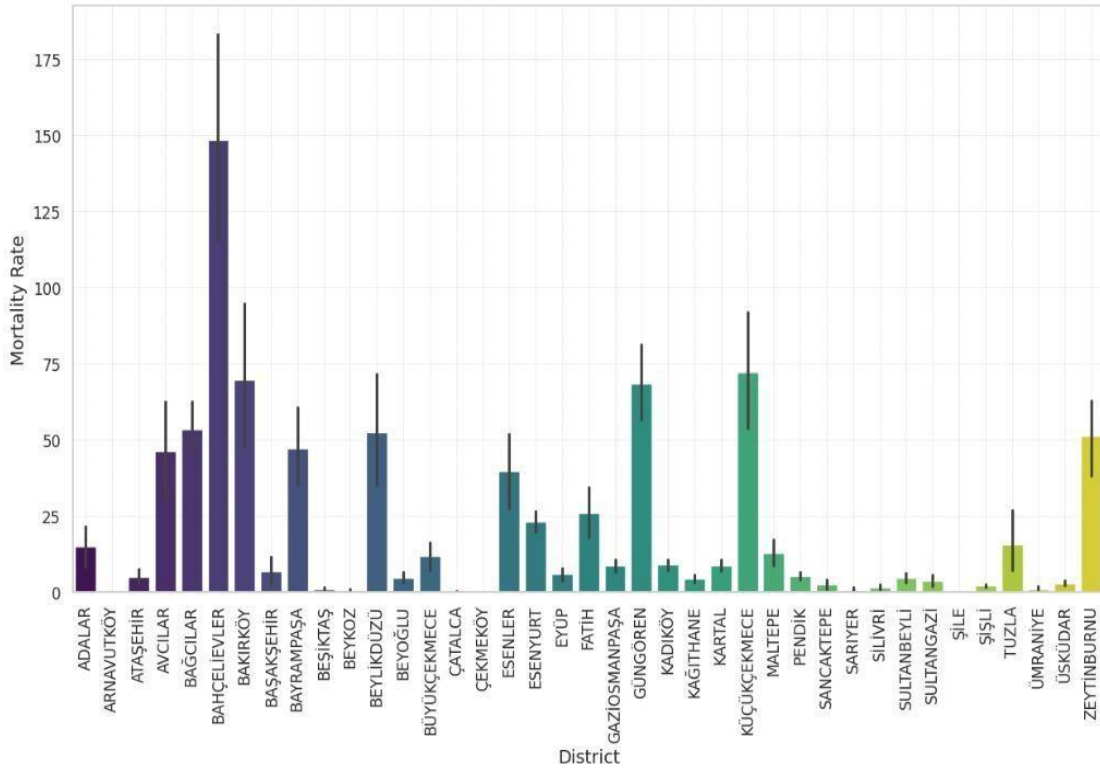
Literatürde deprem sonucunda oluşabilecek altyapı hasarlarının ve bu hasarların neden olabileceği ikincil risklerin (yangın, su kıtlığı, elektrik kesintisi, haberleşme vb.) kapsamlı bir şekilde incelenmemiş olması, mevcut risk değerlendirmelerinin kapsamını sınırlamaktadır. Deprem sırasında veya sonrasında elektrik, doğal gaz ve su gibi hayati altyapı hizmetlerinde meydana gelen kesintiler kamu sağlığını ve güvenliğini tehlikeye atmaktadır. Elektrik kesintileri iletişim sistemlerinin çökmesine ve yangın riskinin artmasına neden olabilirken, su kesintileri sağlık hizmetlerini aksatmaktadır ve salgın hastalık riskini artırmaktadır. Atık su sistemlerinin zarar görmesi çevresel kirlenmeye ve sağlık sorunlarına neden olmaktadır. Bu tür ikincil tehditler deprem risk değerlendirmelerinde daha kapsamlı bir şekilde ele alınmalıdır. İstanbul'un deprem riskini etkin ve tam olarak tahmin etmek için mevcut veri ve modelleme, altyapı hasar değerlendirmesi ve Çoklu Risk Analizine (Aydınoglu ve Taştan, 2015) öncelik verilmelidir. Bu yetersizliklerin giderilmesi, İstanbul'un sismik dayanıklılığının artırılması ve afet kayıplarının azaltılması yönünde atılacak kritik bir adım olacağı düşünülmektedir. Geçmişte İstanbul için yaşanacak depremlerde can kaybı sayısının tahmini ve olumsuz etkisini azaltabilmek için çeşitli çalışmalar yapılmıştır (Kaptan, 2013).

Bu çalışma, İstanbul'un deprem riskinin daha iyi anlaşılması ve yönetilmesi amacıyla makine öğrenmesi ve gelişmiş özellik mühendisliği yaklaşımlarıyla ele alınmasına odaklanmıştır. Çalışmada RF, LR, XGBoost, SVM, DT, ET, CatBoost ve Kısıtlı Aşırı Öğrenme Makinesi (CELM) modellerinin performansları değerlendirilmiştir. CELM modelinin bu alanda daha önceden kullanılmamış olması özgünlük açısından çalışmanın önemini vurgulamaktadır (Parmaksız ve Karakuzu, 2022). Deprem kaynaklı can kayıplarını minimize etmek, ilçelerin depreme hazır olma düzeylerini belirlemek ve sonuç olarak daha etkin acil durum planlarının oluşturulmasına katkı sağlamak çalışmanın ana hedeflerini yansıtmaktadır. Çalışmada, İstanbul'un 23 ilçesi ve 613 mahallesi için kapsamlı bir veri seti kullanılmıştır (Kundak ve Türkoğlu, 2007). Veri setindeki bağımlı ve bağımsız değişkenler arasındaki doğrusal ilişkiyi modellemek için regresyon analizi yöntemi kullanılmıştır. Veri seti kalitesini artırmak için eksik veri analizi yapılmıştır, yüksek korelasyonlu özellikler belirlenmiştir ve bu özellikler hariç tutularak modelin genelleme kabiliyeti güçlendirilmiştir. CELM algoritmasının özellikle ilçe bazlı analizler olmak üzere İstanbul'daki deprem riskini etkili bir şekilde tahmin edebildiğini göstermiştir. Bu durum, makine öğrenimi yöntemlerinin doğal afetlerin etkilerini etkili bir şekilde azaltabileceğine dair güçlü kanıtlar sunmaktadır. Özellik mühendisliği ve hiper parametre optimizasyon yöntemlerinin kombinasyonu modelin performansını önemli ölçüde artırmış ve bilimsel katkıyı güçlendirmiştir. Veri seti üzerinde yapılan kapsamlı analizler, sonuçların güvenilirliğini artırarak daha etkili ve doğru deprem riski tahmini yapılmasını sağlamıştır.

## 2. Veri Seti Analizi ve Özellik Mühendisliği

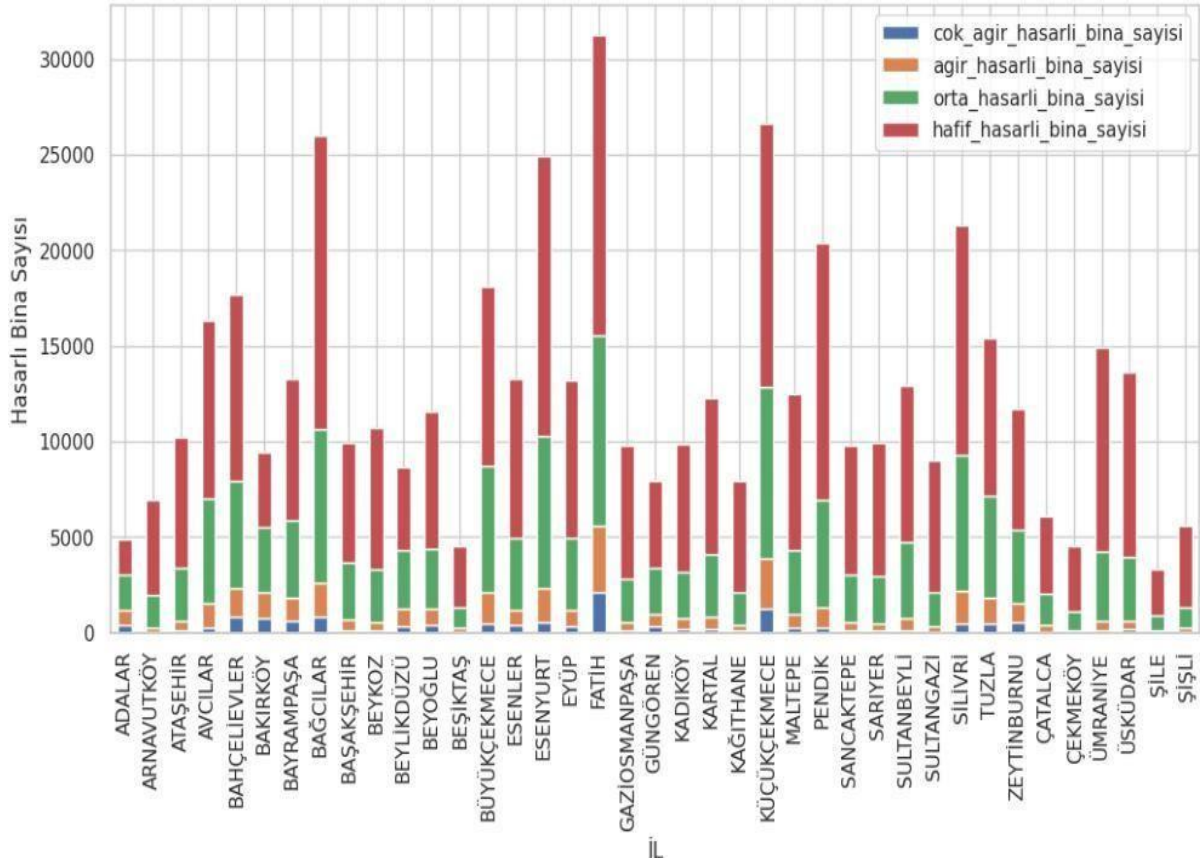
İstanbul Büyükşehir Belediyesi (İBB) tarafından sağlanan güncel ve gerçek verilere sahip veri seti kullanılmaktadır (İstanbul Büyükşehir Belediyesi, 2024). İlgili veri seti, deprem sonrası ortaya çıkan hasarın kapsamlı bir analizini sunmak amacıyla, binalardaki hasar seviyelerinden (çok ağır, ağır, orta, hafif) can kayıplarına, yaralı sayılarından altyapı hasarlarına (doğalgaz, içme suyu, atık su boruları) ve geçici barınma ihtiyaçlarına kadar geniş bir yelpazede değişkenleri içermektedir. Veri setinde yer alan 959 kayıt ve 15 özellik, İstanbul'un farklı ilçelerindeki deprem etkilerini detaylı bir şekilde ortaya koymaktadır. Betimsel istatistikler, depremin etkilerinden coğrafi farklılıkları vurgulamaktadır. Özellikle can kaybı sayısının standart sapması oldukça yüksek (28.84) bulunmuştur. Bu durum, ilçeler arasında can kaybı oranlarında önemli farklılıklar olduğunu göstermektedir. Asgari (0) ve azami (230) can kaybı değerleri arasındaki geniş aralık ise, depremin bazı bölgelerde daha yıkıcı sonuçlar doğurduğunu açıkça ortaya koymaktadır.

Şekil 2'de sunulan ilçe bazlı ölüm oranları dağılımı, bu coğrafi farklılıkları görsel olarak desteklemektedir. İBB veri seti, İstanbul'daki depremin sosyo-ekonomik ve fiziksel etkilerini anlamak için değerli bir kaynak sunmaktadır. Veri setindeki yüksek standart sapma ve geniş aralık değerleri, deprem risk yönetimi ve kentsel dönüşüm planlamalarında dikkate alınması gereken önemli bir husus olarak öne çıkmaktadır.



Şekil 2. İlçe bazlı ölüm oranları.

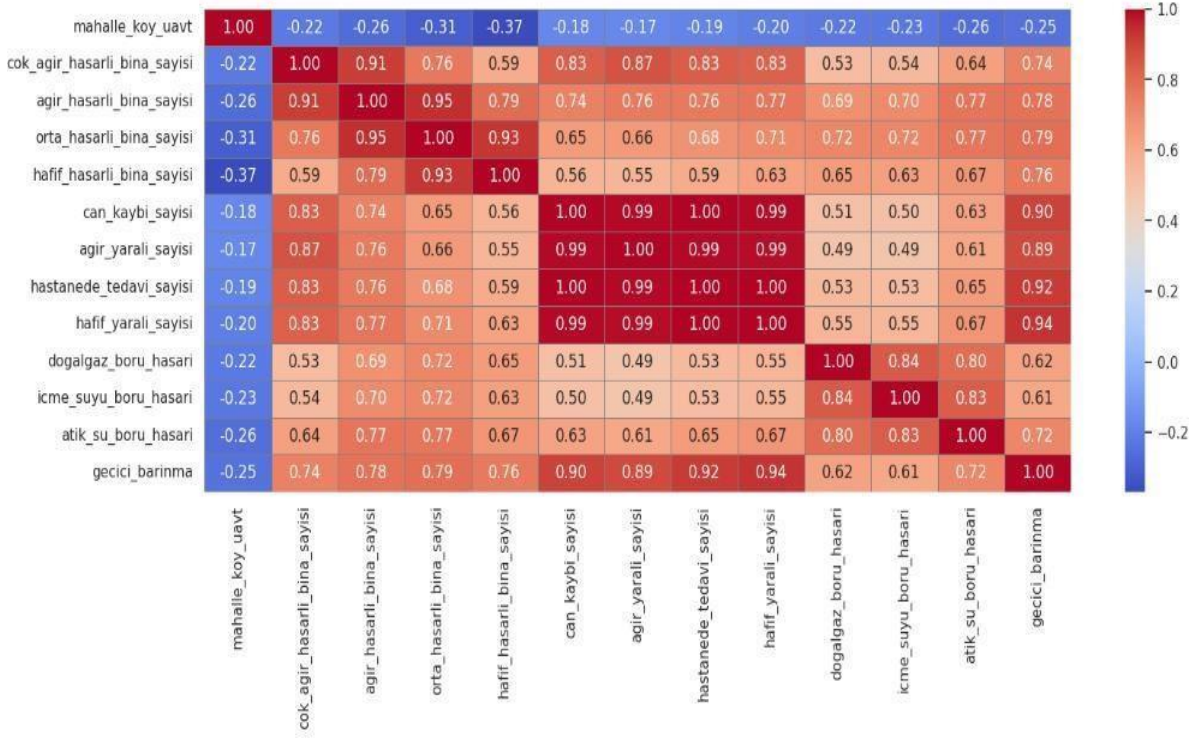
İstanbul'da yapılan hasar tespit çalışmaları, hafif ve orta hasarlı binaların yaygın bir dağılım gösterdiği karmaşık bir deprem riski profili ortaya koymuştur. Bu durum, İstanbul'daki yapı stokunun önemli bir kısmının depreme karşı yeterince güçlendirilmediğini göstermektedir. Bununla birlikte, ağır ve çok ağır hasarlı binaların dağılımı daha yerel ve heterojen olup, bazı mahallelerde bu tür binaların daha yoğun olması deprem riskini artırmaktadır. Bu durum, kentsel dönüşüm projelerinin planlanması ve acil müdahale stratejilerinin belirlenmesi açısından büyük önem taşımaktadır. Hasarlı binaların dağılımı, kentsel dönüşüm çalışmalarının önceliklendirilmesi ve risk yönetim planlarının geliştirilmesi için kritik bir veri kaynağıdır. Şekil 3'te ilçe bazlı hasarlı binaların dağılımı görselleştirilmektedir. Depreme dayanıklılığını artırmak için riskli binaların güçlendirilmesi veya yıkılması, acil durum hazırlıkları ve vatandaşların bilinçlendirilmesi gibi önlemlerin uygulanması elzem olarak ortaya çıkmaktadır.



Şekil 3. Hasarlı binaların ilçe bazlı dağılımı.

Şekil 4'te sunulan korelasyon ısı haritası incelendiğinde, çok ağır hasarlı bina sayısı ile can kaybı sayısı arasında yüksek pozitif bir korelasyon olduğu görülmektedir. Benzer şekilde, hastanede tedavi gören kişi sayısı ile can kaybı sayısı arasında da yüksek bir korelasyon saptanmıştır. Bu bulgular, özellikle ağır hasarlı bina sayısının yüksek olduğu ilçelerde acil risk planlarının hazırlanması gerektiğini ortaya koymaktadır. Bu bağlamda, bina dayanıklılıklarının artırılması, yaşı büyük binaların ise yıkım kararı ile yenilenmesi gerekliliği önemli bir sonuç olarak karşımıza çıkmaktadır.





Şekil 4. Korelasyon ısı haritası.

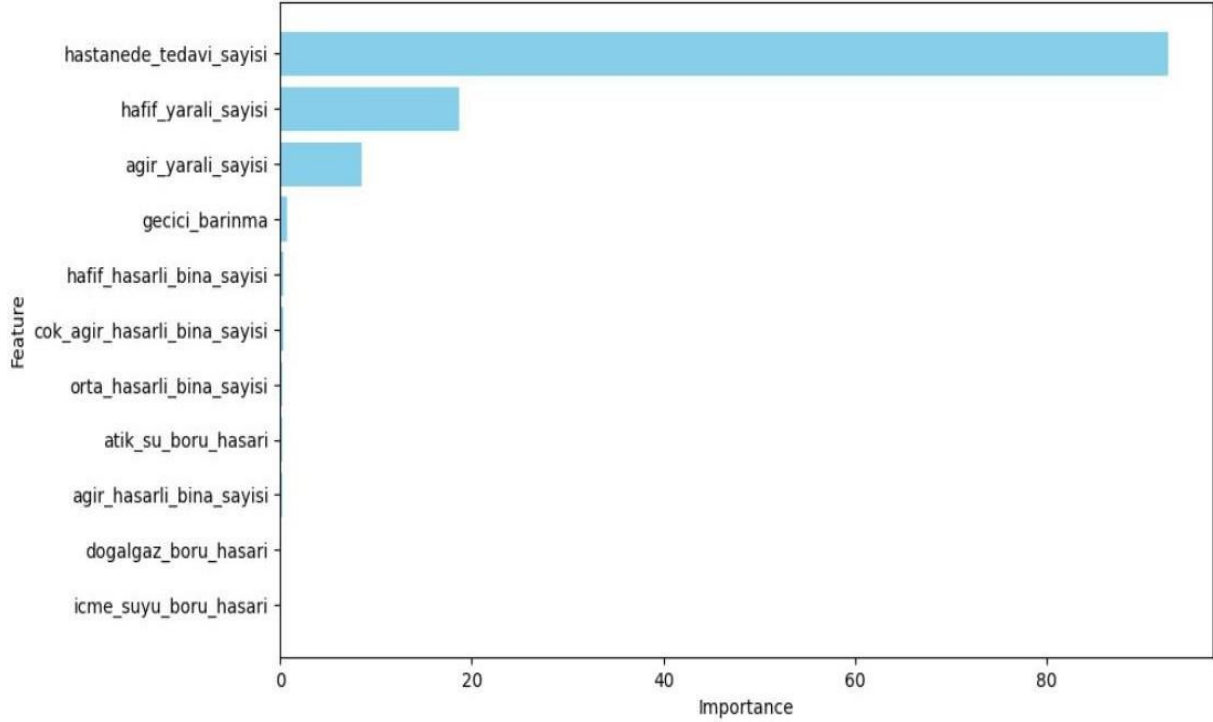
### 3. SHAP-XAI ile Model Açıklama

Makine öğrenimi modellerinde özellik seçimi, gereksiz ifadeleri göz ardı ederek performansı optimize etmektedir. Model, çeşitli yöntemler kullanılarak ölçülebilen yorumlanamaz bir çıktı ürettiğinden, makine öğrenimi uygulamalarında bu çok önemlidir. Bu durum, makine öğrenimi modelini kara kutudan çıkararak ve daha iyi yorumlamaya olanak tanıyan SHAP gibi açıklanabilir yapay zekâ (XAI) yöntemlerinin geliştirilmesine yol açmıştır.

(Slundberg, 2024). SHAP analizi, modellerin açıklanabilirliğini geliştirmek için SHapley oyun teorisinin mantığını kullanır. SHAP, model içindeki yapıyı anlayarak karar verme sürecinde en etkili metriklerin belirlenmesine yardımcı olmaktadır ve modelin çıktıya katkısının daha iyi anlaşılmasını sağlamaktadır.

2024 yılında Huanjing Wang vd. kredi kartı dolandırıcılığı tespit modellerinde SHAP tabanlı ve önem tabanlı özellik seçimi yöntemlerini karşılaştırmıştır ve önem tabanlı yöntemlerin SHAP tabanlı yöntemlerden daha iyi performans gösterdiğini açıklamıştır (Wang, 2024). Ziqi Li'nin 2022 tarihli makalesinde, makine öğrenimi modellerinden mekânsal etkileri çıkarmak için SHAP yöntemi araştırılmış ve XGBoost modeli geleneksel mekânsal istatistiksel modellerle karşılaştırılmıştır (Li, 2022). Ferenc Szigeti, Edoardo Mosca, Stella Tragianni, Daniel Gallagher ve Georg Groh'un 2022 tarihli makalesinde, Doğal Dil İşleme (NLP) modellerinin yorumlanabilirliğini iyileştirmek için SHAP tabanlı yöntemler incelenmiş, 41 farklı yöntem incelenmiş ve NLP araştırmacıları için öneriler sunulmuştur. M. Toğaçar vd. Göğüs röntgen görüntülerini kullanarak zatürre tespiti için derin öğrenmeye dayalı bir model geliştirmiştir ve %99.41 doğruluk oranına ulaşmıştır (Toğaçar, 2020).

Şekil 5'te SHAP ile modelin açıklanabilirliğini göstermek için en etkili özelliklerin sıralaması verilmektedir. Burada SHAP, RF modeli için "hastanede\_tedavi\_sayisi", "hafif\_yarali\_sayisi" ve "agir\_yarali\_sayisi" özelliklerinin modelin başarımında en etkin olduğunu göstermektedir.



Şekil 5. RF modelinde SHAP ile özellik açıklanabilirliği.

#### 4. Boruta, MI ve mRMR ile Özellik Seçiminin Önemi

Literatürde özellik seçiminde Boruta, MI ve mRMR'nin yaygın kullanım örnekleri mevcuttur. mRMR, veri seti içindeki hangi değerlerin bağımlı değişken ile daha çok ilişkisinin olduğunu bularak gereksiz benzerlikleri de azaltmaya yardımcı olmaktadır. Özellikle veri setinin çok boyutlu olduğu zamanlarda etkili bir yöntem olarak tercih edilmektedir. 2013 yılında Nicolas De Jay ve meslektaşları, sağlam tahminler oluşturmak ve tahmin doğruluğunu artırmak için birden fazla özellik seti kullanarak mRMR yöntemine bir topluluk yaklaşımı getirmiştir. Çalışma, topluluk mRMR uygulamalarının klasik mRMR'den daha iyi performans gösterdiğini ve biyolojik bağlamla daha ilgili genleri tanımladığını göstermiştir (De Jay, 2013).

2010 yılında Miron B. Kursa ve Witold R. Rudnicki, bir veri kümesindeki tüm ilgili değişkenleri bulan bir özellik seçim algoritması olan Boruta paketini geliştirmiştir. Boruta, rastgele orman sınıflandırma algoritması etrafında bir sarmalayıcı olarak tasarlanmıştır ve rastgele problemler kadar ilgili olmayan özellikleri kaldırmakta kullanılmaktadır (Kursa ve Rudnicki, 2010). 2013 yılında Kursa ve Rudnicki, bir bilgi sistemindeki gerçekten önemli değişkenlerin tam kümesini belirlemek için algoritmanın geliştirilmiş bir versiyonunu sunmuştur. Algoritma, orijinal niteliklerin önemini rastgele kopyalarıyla iteratif olarak karşılaştırmaktadır. Algoritmanın performansı sentetik veri örnekleri ve kısa RNA dizilerinin aptamerik aktivitesi için önemli olan dizi motifleri üzerinde analiz edilmiştir (Kursa v.d., 2010).

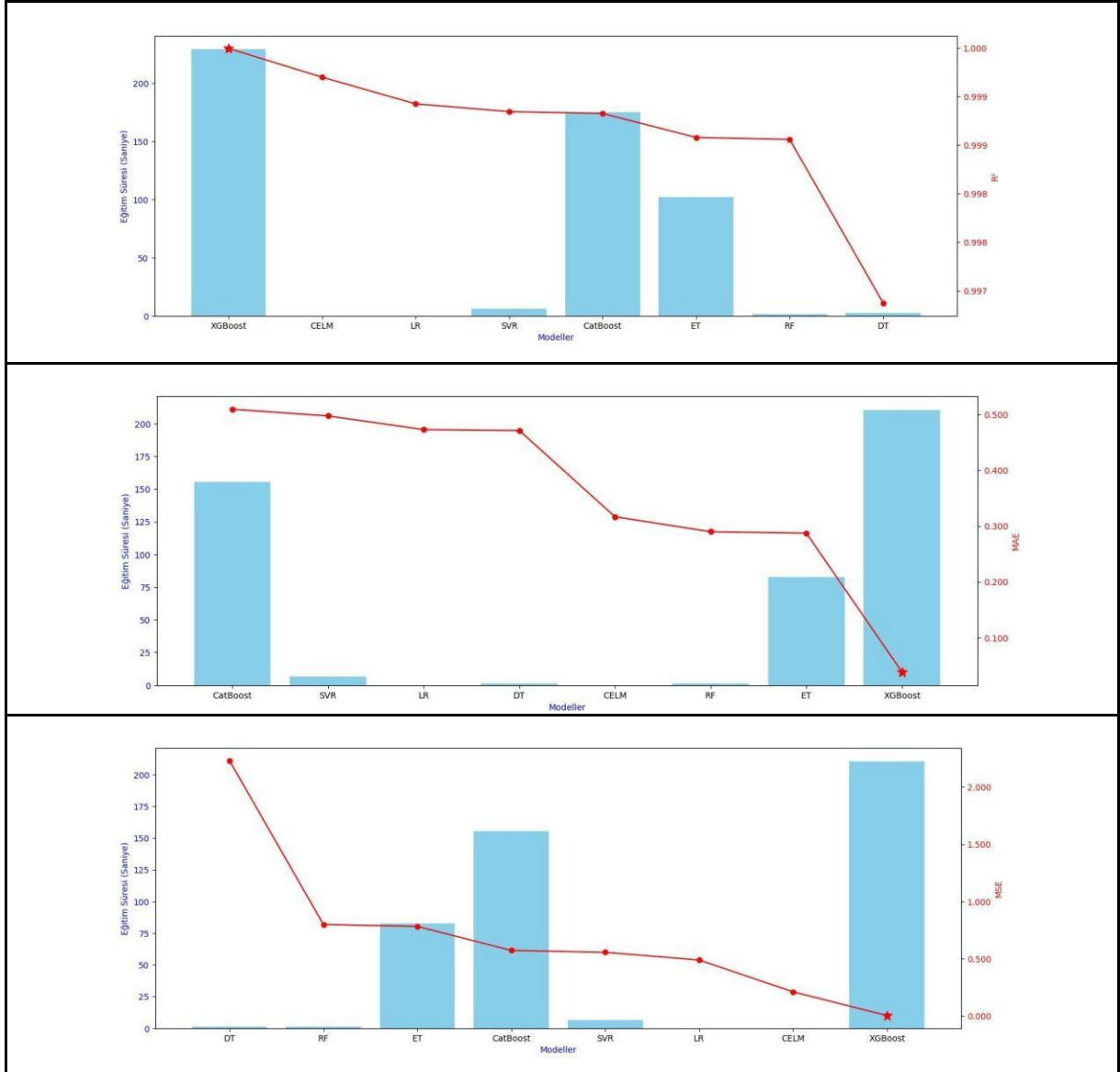
MI, iki değişkenin birbirleriyle ne kadar ilişkili olduklarını ölçümlenmektedir. Hangi özelliğin hedef değişkeni daha iyi tanımladığını bulmaya yardımcı olmaktadır. 2014 yılında Jorge R. Vergara ve Pablo A. Estévez, alaka düzeyi, fazlalık ve tamamlayıcılık gibi kavramları açıklayarak MI dayalı özellik seçim yöntemlerini gözden geçirerek başarılı sezgisel kriterler için teorik bir çerçeve sunmuşlardır (Vergara ve Estévez, 2014). 2011 yılında Fatemeh Amiri, Mohammad Mahdi Rezaei Yousefi, Caro Lucas, Azadeh Shakery ve Nasser Yazdani, özellikle R2L ve U2R saldırıları için saldırı tespit doğruluğunu artırmak için MI tabanlı özellik seçimi kullanan yeni bir saldırı tespit sistemi önermiştir. Çalışma, MI tabanlı yöntemin izinsiz girişleri daha yüksek doğrulukla tespit etmedeki etkinliğini göstermiştir (Amiri, 2011). 2006 yılında Rajasekaran ve arkadaşları dalgacık dönüşümleri ve temel bileşen analizi kullanarak yapay deprem ve tepki spektrumları oluşturmak için beş sinir ağı tabanlı model önermişlerdir. Modeller, eğitim için 25 ve test için 4 depremden elde edilen veriler kullanılarak

Lee ve Han'ın modelleri ile karşılaştırılmıştır. Bu yaklaşım, yapısal dinamik problemlerine çözüm tasarlamak için kullanışlıdır ve önemli bir araçtır (Rajasekaran v.d., 2006). 2015 yılında G. Asencio-Cortés ve arkadaşları, veri boyutunu azaltmak ve deprem tahmini için yeni veri kümeleri oluşturmak amacıyla temel bileşen analizini tanıtmıştır. Şili'nin en çok tehdit altındaki şehirleri olan Santiago ve Pichilemu'yu temel bileşen analizi ile çeşitli sınıflandırıcılar kullanarak incelemiştir ve sonuçlarda önemli gelişmeler elde etmişlerdir (Cortés ve Gualberto, 2015). Aydın Demir, Selim Günay, Marko Marinković vd. deprem mühendisliğini ve politikalarını değerlendirmek için 1999 ve 2022 Düzce depremlerini analiz etmiştir ve deprem kodlarına uyum ve yapısal güçlendirme çabalarının yapılarda düşük hasara yol açtığını bulmuştur (Demir v.d., 2022).

## 5. Model Geliştirme ve Değerlendirme

Can kaybı sayısının tahmini için, RF, LR, XGBoost, SVM, DT, ET, CatBoost ve CELM olmak üzere sekiz farklı model kurulmuştur, bu modellerin eğitim ve test veri setindeki analizlerde harcadıkları süreler ile MAE, MSE ve  $R^2$  değerleri hesaplanmıştır. Keşifsel veri ve korelasyon analizi incelemeleri ile veriler arasında doğrusal ilişki olduğu saptanmıştır, örneğin hasarlı bina sayısı arttıkça can kaybı sayısının da artış gösterdiği belirlenmiştir. Eğitim veri seti için elde edilen sonuçlar Şekil 6'da ve Tablo 1'de, test veri setinde elde edilen sonuçlar ise Şekil 7'de ve Tablo 2'de verilmiştir. En iyi modelin verdiği sonuçların daha net gösterimi için yıldız simgesi ile işaretlenmiştir.





Şekil 6. Modellere ait  $R^2$ , MAE ve MSE değerlerine karşın işlem süreleri (eğitim kümesi %70)

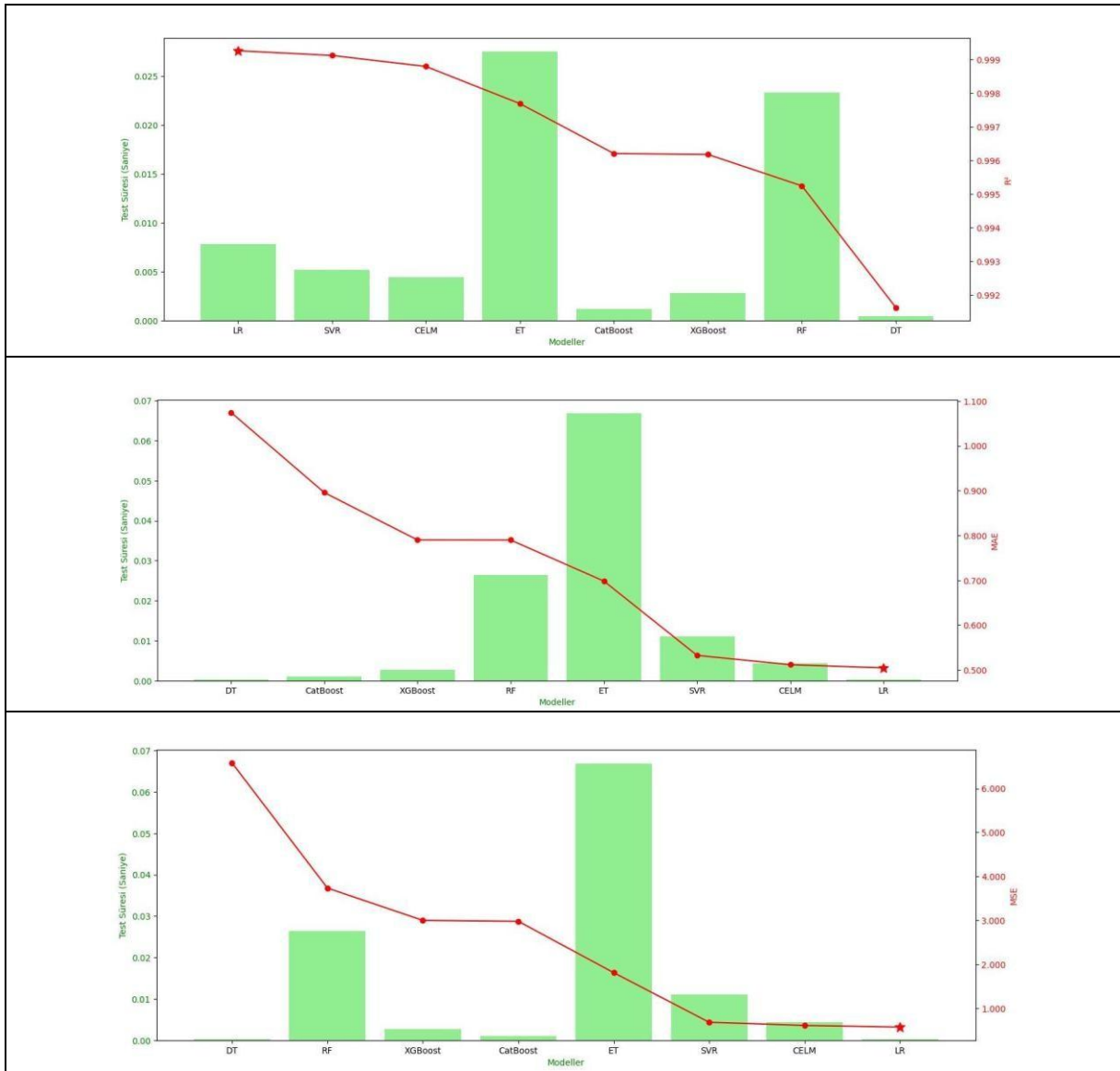
Tablo 1. Modellerin Performans Metrikleri (eğitim kümesi %70)

Modeller	Eğitim Süre <sup>a</sup>	MAE	MSE	$R^2$
XGBoost	210.6259	0.0384	0.0035	0.9999
CELM	200.0010	0.0390	0.0039	0.9998
LR	0.0133	0.4730	0.4885	0.9994
SVR	6.6825	0.4976	0.5562	0.9993
CatBoost	155.6197	0.5096	0.5727	0.9993
ET	82.6432	0.2875	0.7825	0.9990
RF	1.0881	0.2900	0.7991	0.9990
DT	1.2474	0.4715	2.2307	0.9973

<sup>a</sup> Eğitim süresi saniye cinsinden verilmiştir.

Zaman analizi, makine öğreniminde modellerin eğitim ve test aşamalarındaki verimliliğini ölçmek için kullanılmaktadır. Bir modelin veri setini öğrenmesi için gereken süreyi ve tahminlerde bulunma hızını ölçmektedir. Zaman analizinin temel amacı, bir modelin oluşturulması ve değerlendirilmesi için geçen süreyi izlemek, optimizasyonu sağlamak ve genel performansı iyileştirmektir. Bu, potansiyel darboğazların belirlenmesine yardımcı olarak kaynak gereksinimlerini belirlemede etkin rol oynamaktadır.

Makine öğrenimi performans ölçütleri arasında  $R^2$  (göreceli açıklanabilirlik), MAE (ortalama mutlak hata) ve MSE (ortalama karesel hata) bulunmaktadır.  $R^2$ , bir modelin bağımlı değişken tarafından ne kadarının açıklanabileceğini ölçer ve yüksek bir  $R^2$  değeri (0-1 aralığında) daha iyi açıklanabilirliğini gösterir, ancak model verileri iyi ezberlediğinde aşırı uyum oluşur, bu da eğitim bölümünde yüksek performansa, ancak test bölümünde düşük performansa neden olmaktadır. Uyumsuzluk, model verileri iyi öğrenemediğinde ve hem eğitim hem de test verilerinde düşük performans gösterdiğinde ortaya çıkmaktadır. Dengeli uyum, modelin hem eğitim hem de testte verileri iyi öğrenmesi ve yüksek performans metriklerine ulaşmasıdır. MAE regresyon problemlerinde kullanılan, tahmin edilen ve gerçek değerler arasındaki mutlak tahmin hatalarının ortalaması alınarak hesaplanan bir metriktir. MSE, tahmin edilen ve gerçek değerler arasındaki farkın ortalama karesidir ve ne kadar küçükse model performansının iyi olduğu ifade edilmektedir.



Şekil 7. Modellere ait  $R^2$ , MAE ve MSE değerlerine karşı işlem süreleri (test kümesi %30)

**Tablo 2.** Modellerin Performans Metrikleri (test kümesi %30)

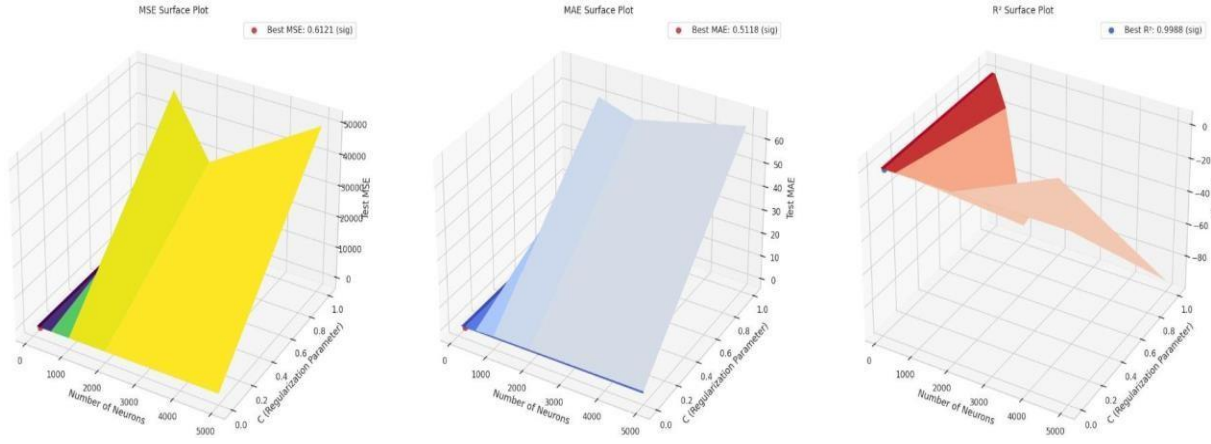
Modeller	Test Süre <sup>a</sup>	MAE	MSE	R <sup>2</sup>
XGBoost	0.0026	0.7899	3.0016	0.9961
CELM	0.0090	0.540	0.050	0.9936
LR	0.0130	0.5046	0.5760	0.9992
SVR	0.0110	0.5324	0.6855	0.9991
CatBoost	0.0010	0.8957	2.9807	0.9962
ET	0.0668	0.6982	1.8126	0.9976
RF	0.0263	0.7898	3.7364	0.9952
DT	0.0003	1.0743	6.5837	0.9916

<sup>a</sup> Test süresi saniye cinsinden verilmiştir.

XGBoost, 2016 yılında Tianqi Chen ve Carlos Guestrin tarafından, özellikle büyük veri kümelerinde makine öğreniminde ağaç güçlendirme yöntemlerinin verimliliğini ve ölçeklenebilirliğini artırmak için yazılmıştır. XGBoost sisteminin tasarımını, algoritmik optimizasyonlarını ve dağıtık hesaplama yeteneklerini detaylandırmaktadır. XGBoost, makine öğrenimi yarışmalarında ve gerçek dünya uygulamalarında gösterdiği üstün performans ile ağaç güçlendirme yöntemleri için popüler bir seçim haline gelmektedir (Tianqi ve Guestrin, 2016).

Çalışmada uyarlanan XGBoost modeli için eğitim aşamasında amprik olarak hiper parametreler;  $n\_estimators$  değeri (100,200,300),  $learning\_rate$  (0.001, 0.1, 0.2),  $max\_depth$  (3,5,7),  $subsample$  (0.8, 0.9, 1.0) ve  $gamma$  (0, 0.1, 0.2) değerlerinden seçilmektedir. Eğitim süresi (225.4562) olarak hesaplanmıştır. Eğitim için MSE (0.0035), MAE (0.0384) ve  $R^2$  değeri (0.9999) olarak hesaplanmıştır. XGBoost modelinde test aşamasında ise hiper parametreler;  $n\_estimators$  (100, 200, 300),  $learning\_rate$  (0.01, 0.1, 0.2),  $max\_depth$  (3,5,7),  $subsample$  (0.8, 0.9, 1.0) ve  $gamma$  (0, 0.1, 0.2) değerlerinden seçilmiştir. Test süresi (0.0029), veri sayısına bağlı olarak eğitime kıyasla daha hızlı bir tahminle gösterdiğini belirtebiliriz. MSE (3.0016), MAE (0.7899) ve  $R^2$  değeri (0.9961) olarak bulunmuştur. Elde edilen sonuçlar eğitim kısmına kıyasla performans oldukça düşük seyretmektedir. Burada modelin aşırı uyum sağladığı ve test verileri üzerinde tahmin yapmakta zorlandığını göstermektedir.

CELM, ELM'nin sinir ağları ile birleşmesi sonucu oluşmuş bir yapıdır. ELM, yapay sinir ağlarının eğitiminde gizli katmanların ağırlıklarını sabit tutarak daha hızlı bir eğitim süreci oluşmasını sağlamaktadır. ELM'nin yapısında giriş katmanı, gizli katman ve çıkış katmanı olmak üzere üç katmanlı bir yapısı bulunmaktadır. CELM ile ELM birleşerek daha hızlı bir şekilde eğitim ve aşırı öğrenmeye karşı bir direnç oluşturmaktadır. CELM, ELM'nin avantajlarını korurken genelleme performansını artırmaktadır. Ağırlık kısıtlamaları, model karmaşıklığını azaltarak aşırı uyum olasılığını düşürmektedir (Wang v.d., 2022).



**Şekil 8.** CELM modeli değerlendirilmesi (a- MSE yüzeyi, b- MAE yüzeyi ve c- $R^2$  yüzeyi / test verisi %30)

CELM, eğitim sürecini sadece 0.0516 saniyede tamamlayarak aşırı uyumla etkili bir şekilde mücadele eden dengeli bir öğrenme yöntemidir. Modelin hiper parametreleri deneysel olarak belirlenmiş ve 500 nöron, sigmoidal aktivasyon fonksiyonu ve 100 düzenlilik faktörü kullanılmıştır. CELM, eğitim sürecini 0.0044 saniyede tamamlayarak yüksek performans göstermiş ve eğitim süresine kıyasla daha hızlı performans göstermiştir. Test verileri, eğitim verilerine kıyasla hata oranlarında artış göstermiştir, ancak test aşamasında  $R^2$  değerinin 0.9988 olması, bir miktar doğruluk kaybı olmasına rağmen, eğitim ve test aşamaları arasındaki performans farkının sınırlı olduğunu ve modelin genel doğruluğunu koruduğunu göstermektedir. Modelin performans değerlendirme ölçütleri arasında MAE 0.5118, MSE 0.6121 ve  $R^2$  0.9994 yer almaktadır. Şekil 8’de CELM’in genel olarak iyi performans gösterdiği görülmektedir.

Çalışmada bağımlı ve bağımsız değişkenler arasındaki ilişkiyi incelemek için LR kullanılmıştır. Modelin performans değerlendirme ölçütleri arasında MAE 0.4730, MSE 0.4885 ve  $R^2$  0.9994 yer almaktadır ve bu da eğitim verileri üzerinde iyi bir uyum olduğunu göstermektedir. Test aşamasında kullanılan hiper parametreler şunlardır: fit\_intercept: True, copy\_X: True ve n\_jobs: Yok şeklindedir. Test süresi 0.0003 saniyedir ve modelin MAE değeri 0.5046, MSE 0.5760 ve  $R^2$  değeri 0.9992’dir; bu da kabul edilebilir hata oranlarıyla yüksek doğruluk ve güvenilir performansa işaret etmektedir.

SVR modeli, marj aralığındaki maksimum noktada en küçük hataya sahip bir çizgi veya eğri çizmeyi amaçlayan, modelin doğruluğunu artırmayı ve aşırı uyum riskini azaltmayı hedefleyen bir yöntemdir. Eğitim aşamasında seçilen hiper parametreler C (0.1, 1, 10), epsilon (0.01, 0.1, 0.2) ve kernel ('linear', 'rbf')’dir. Eğitim süresi 150.2417 olarak hesaplanmış, düşük MAE (0.4976) ve MSE (0.5562) değerleri düşük hata payını göstermiştir.  $R^2$  (0.9993) yüksek açıklanabilirliğe işaret etmektedir. Test aşamasında hiper parametreler C (0.1, 1, 10), epsilon (0.01, 0.1, 0.2) ve kernel ('linear', 'rbf') olarak belirlenmiştir. Test süresi 0.0051 olarak hesaplanmıştır ve bu da test verileriyle yüksek tahmin hızını göstermektedir. MAE, daha düşük mutlak hatayı gösteren 0.5324 ve MSE, eğitim bölümüne kıyasla daha yüksek olduğunu gösteren 0.6855 idi.  $R^2$  (0.9991) olarak hesaplanmıştır.

CatBoost modeli, Yandex tarafından arama motoru yeteneklerini geliştirmek için geliştirilen bir gradyan artırma kütüphanesidir. Aşırı öğrenmeyi en aza indirmek için topluluk öğrenme tekniklerini kullanarak sınıflandırma ve regresyon problemleri için giderek daha fazla kullanılmaktadır. Çalışmada CatBoost modelinde, modeli eğitmek ve test etmek için iterasyonlar, öğrenme oranı, derinlik ve l2\_leaf\_reg gibi hiper parametreler kullanılmaktadır. CatBoost modeli eğitim ve test aşamasında kullanılan hiper parametreler; iterations (100, 200, 300), learning\_rate (0.01, 0.1, 0.2), depth (3,5,7), l2\_leaf\_reg (1, 3, 5) amprik seçilmiştir.

Eğitim süresi 150.0010, MSE (0.5727), MAE (0.5096) ve  $R^2$  (0.9993) olarak bulunmuştur. Test süresi 0.0011, MSE (2.9807), MAE (0.8957) ve  $R^2$  (0.9962) olmuştur.

ET modeli, RF'ye benzer şekilde, performansı artırmak ve eğitim süresinden tasarruf etmek için düğümlerin rastgele seçimini kullanır. Model eğitiminde kullanılan hiper- parametreler arasında `n_estimators`, `max_depth`, `min_sample_split`, `min_sample_leaf` ve `max_features` yer almaktadır. Eğitim MSE (0.7825), MAE (0.2875) ve  $R^2$  (0.9990) olarak bulunmuştur. Eğitim süresi 73.2488'dir ve modelin açıklanabilirliğini etkileyen yüksek hata metrikleri %99.90 doğrulukla sonuçlanmıştır. Test aşamasında uygulanan hiper parametreler ise; `n_estimators`: `randint` (100, 500), `max_depth`: `randint` (3, 20), `min_sample_split`: `randint` (2, 10), `min_sample_leaf`: `randint` (1, 10), `max_features`: `uniform` (0.1, 1.0) olarak belirlenmiştir. Test aşamasında, model MSE (1.8126), MAE (0.6982) ve  $R^2$  değerleri (0.9976) ile hızlı bir şekilde eğitilmiştir. Ancak, test verilerinin düşük olması aşırı uyum olasılığını artırmakta ve ET modelinin tahmin için yeterli olmayabileceğini göstermektedir.

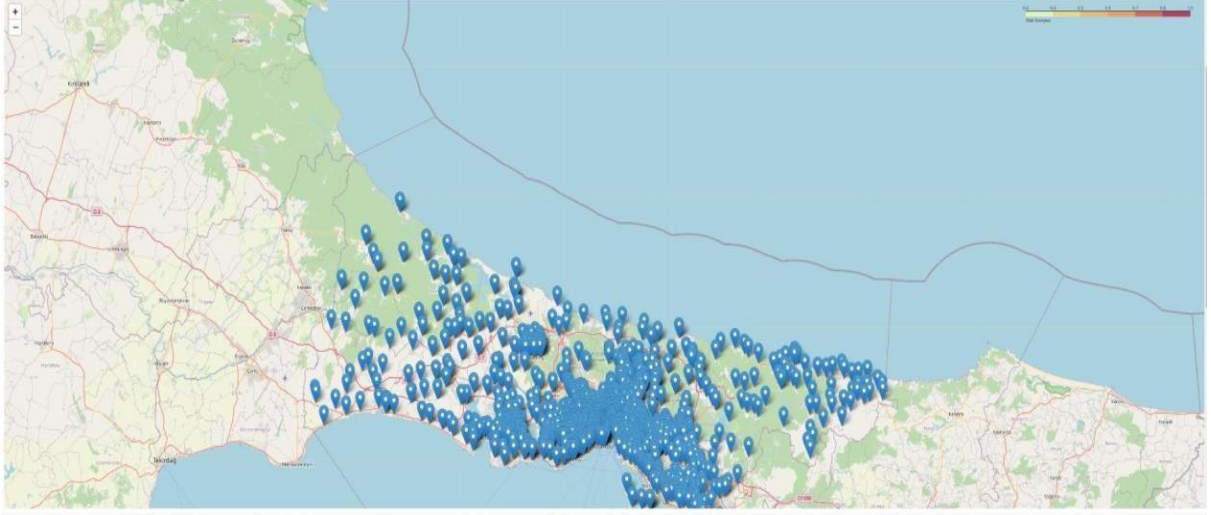
RF modeli, daha doğru ve sağlam tahminler yapmak için birden fazla karar ağacını birleştirmektedir. Her ağaç, verilerin farklı bir alt kümesi üzerinde eğitilir ve sonuç, bu ağaçların tahminlerinin ortalamasının alınmasıyla oluşmaktadır. Bu yaklaşım, aşırı uyumu önlemeyi ve tahmin hatalarından kaçınmayı amaçlamaktadır. Eğitim aşamasında MSE (0.7991), MAE (0.2900) ve  $R^2$  değeri (0.9990) gibi metriklerle 300 karar ağacı oluşturmak için `n_estimators` (30) ve `random_state` (11) gibi hiper parametreler kullanılmaktadır. Test aşamasında da aynı parametreler kullanılmış ve test verilerini analiz etmek için harcanan süre (0.0233) olmuştur; bu da eğitim verilerine kıyasla daha hızlı tahmin yapıldığını göstermektedir. MSE (3.7364), MAE (0.7898) ve  $R^2$  değeri (0.9952) olmuştur. Bununla birlikte, modelin güvenilirliği ve performansı iyi değildir ve öğrenilen modelle tahminler yapmakta zorlanmaktadır.

DT modeli verileri inceler ve değişkenleri bölmeler kullanarak gruplara ayırır ve ağaç dalları oluşturmaktadır. Model görselleştirme sağlar ve anlaşılmasını kolaylaştırır, ancak çok fazla dallanma olduğunda aşırı uyum olasılığını artırma gibi bir dezavantajı da vardır. Eğitim aşamasında `max_depth` (3, 10), `min_samples_split` (2, 10) ve `min_samples_leaf` (1, 10) gibi hiper parametreler kullanılmış ve eğitim süresi (2.1474) olmuştur. MSE (2.2307), MAE (0.4715) ve  $R^2$  değeri (0.9973) olarak bulunmuştur. Test aşamasında, model eğitim verilerine kıyasla daha hızlı tahminler sağlar, ancak tahmindeki sapmalardan muzdariptir ve düşük aşırı uyum göstermiştir. MSE (6.5837), MAE (1.0743) ve  $R^2$  değeri (0.9916) olarak bulunmuştur. Modelin performansı eğitim verilerine kıyasla daha düşüktür ve farklı modellerin etkinliğini değerlendirirken hem eğitim hem de test verilerini dikkate almak önemlidir.

## 6. İlçe Bazlı Can Kaybı Yoğunluğunun Haritalama Sistemi ile Gösterimi

Çalışma içerisinde haritalama sistemi kullanılarak, deprem öncesinde İstanbul'un çeşitli ilçelerinde yaşanabilecek can kaybı oranlarını görsel bir şekilde gösterimi sağlanmıştır. Oluşturulan haritalama sistemi önceden yüksek can kaybı sayısını tahmin ederek bölgede zemin yapısı incelenip önlem almak için drenaj, zemin enjeksiyonu vb. gibi işlemler ile can kaybı sayısı azaltılmaya çalışılmaktadır.





Şekil 9. İstanbul ilçe bazlı deprem risk haritası

Şekil 9’da yer alan haritalama sistemi, olası İstanbul depreminde can kaybı sayısını önceden tahmin ederek risk yönetim planlarının yapılmasına ve aynı şekilde acil durum tekniklerinin geliştirilmesine katkı sağlamaktadır. Bu tür hassas, ilçe bazlı gösterimler felaket zamanlarında kritik öneme sahiptir. Harita oluştururken ilk adım Json dosyasından İstanbul ilinin ilçelerinin sırasıyla enlem ve boylam bilgileri pandas kütüphanesi ile okutularak bir GeoDataFrame oluşturulur. Seçilen en iyi model (LR) ile İstanbul’un ilçe bazlı can kaybı tahmini yaptırılır ve bu bir CSV dosya formatına dönüştürülerek her ilçe için deprem risk seviyeleri belirlenir. Daha sonra Folium kullanılarak İstanbul için bir harita oluşturulur. Harita en sonunda bir HTML dosyasına kaydedilip HTML çerçevesi ile embed edilerek tarayıcı üzerinden görüntüleme sağlanılmaktadır.

2021 yılında Boğaziçi Üniversitesi Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü tarafından deprem yer sarsıntısını ve bina hasar dağılımı harita üzerinde gösterilmiştir (Boğaziçi Üniversitesi Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü, 2021).

## 7. Sonuçlar ve Gelecek Çalışmalar

Çalışmada, İstanbul’un deprem riskini anlamak ve yönetmek için makine öğrenimi ve gelişmiş özellik mühendisliği teknikleri kullanılmıştır. LR, SVM ve CELM dahil olmak üzere sekiz regresyon modeli, depremle ilgili kayıpları tahmin etmedeki doğruluk açısından değerlendirilmiştir. Bu sonuçlar, farklı ilçelerdeki depreme hazırlık seviyelerini belirleyerek etkili acil durum planlarının hazırlanmasına yardımcı olmuştur. Bulgular, şehrin afet yönetimi stratejilerine rehberlik etmesi hedeflenmiştir. Gelecekteki araştırmalar, makine öğrenimi algoritmalarının performansını artırmak ve daha kapsamlı iç görüler sağlamak için derin öğrenme yöntemlerini entegre etmelidir. Derin öğrenme modelleri deprem sonrası can kaybı ve hasar oranlarını daha detaylı tahmin edilmesi beklenmektedir. Yapay zekâ tabanlı analizler, bina yapısı, yaşı, nüfus yoğunluğu ve altyapı dayanıklılığı gibi faktörleri dikkate alarak kentsel planlama süreçlerine entegre edilebilir. Gelişmiş tahmin modelleri ve haritalama teknikleri depreme müdahale kapasitesini artırabilir ve güvenli bina tasarımları yoluyla kırılganlığı azaltabilir. Afet yönetim süreçlerini daha kapsamlı hale getirmek için fay hattı konumu, bölgesel zemin özellikleri ve demografik yapılar gibi detaylar daha fazla çalışma ile ele alınabilir.

### Çıkar Çatışması

“Yazarlar çıkar çatışması olmadığını beyan etmişlerdir.”

### Yazarların Katkı Oranı

“Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan etmişlerdir.”

### Etik Beyan

“Bu çalışmada sunulan veri, bilgi ve belgeler akademik ve etik kurallar çerçevesinde elde edilmiştir.”

### Finansal Destek

“Çalışmada finansal destek alınmamıştır.”

## Kaynakça

- Amiri, F. (2011). Mutual information-based feature selection for intrusion detection systems. *Journal of Network and Computer Applications*, 34(4), 1184-1199. <https://doi.org/10.1016/j.jnca.2011.01.002>
- Aydınoglu, A. Ç., & Taştan, B. (2015). Çoklu afet ve risk yönetiminde tehlike ve zarar görebilirlik belirlenmesi için gereksinim analizi. *Marmara Coğrafya Dergisi*, 31, 1-19.
- Boğaziçi Üniversitesi Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü. (2021). *İstanbul İleri Erken Uyarı ve Risk Yönetim Sistemi (İERREWS) Raporu*. İstanbul Büyükşehir Belediyesi. <https://deprezmemin.ibb.istanbul/uploads/prefix-ierrews-rapor-66a6a12ee02d7.pdf>
- Cortés, A., & Gualberto. (2015). Improving earthquake prediction with principal component analysis: Application to Chile. In *Hybrid Artificial Intelligent Systems: 10th International Conference, HAIS 2015, Bilbao, Spain, June 22-24* (pp. 123-134). [https://doi.org/10.1007/978-3-319-19644-2\\_33](https://doi.org/10.1007/978-3-319-19644-2_33)
- Demir, A., Günay, S., & Marinković, M. (2022). Turkey earthquake: Advances in the past 2 decades, lessons learned, and future projections. *Bulletin of Earthquake Engineering*. <https://doi.org/10.1007/s10518-024-01984-z>
- De Jay, N. (2013). mRMRe: An R package for parallelized mRMR ensemble feature selection. *Bioinformatics*, 29(18), 2365-2368. <https://doi.org/10.1093/bioinformatics/btt383>
- Eyidoğan, H., & Balamir, M. (2010). Türkiye'de ve İstanbul'da deprem riskleri. *ResearchGate*. [https://www.researchgate.net/publication/343962745\\_Turkiye%27de\\_ve\\_Istanbul%27da\\_Deprem\\_Riskleri](https://www.researchgate.net/publication/343962745_Turkiye%27de_ve_Istanbul%27da_Deprem_Riskleri)
- Eyidoğan, H. (2007). Marmara bölgesinin ve İstanbul kentinin deprem tehlikesi üzerine bir derleme. In *TMMOB Afet Sempozyumu Bildiri Kitabı* (ss. 15-29). Mattek Matbaacılık.
- İstanbul Büyükşehir Belediyesi. (2024). Veri seti: Deprem senaryosu analiz sonuçları. <https://data.ibb.gov.tr/en/dataset/deprem-senaryosu-analiz-sonuclari>
- Kaptan, K. (2013). İstanbul'un deprem hazırlığı için 'sıfır' can kaybı projesi. *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 28(1), 119-126.
- Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, 36, 1-13. <https://doi.org/10.18637/jss.v036.i11>
- Kursa, M. B., Jankowski, A., & Rudnicki, W. R. (2010). Boruta—A system for feature selection. *Fundamenta Informaticae*, 101(4), 271–285. <https://doi.org/10.3233/FI-2010-288>
- Kundak, S., & Türkoğlu, H. (2007). İstanbul deprem risk analizi. *İTÜ Dergisi/A*, 6(2), 37–46.
- Li, Z. (2022). Extracting spatial effects from machine learning models using local interpretation methods: An example of SHAP and XGBoost. *Computers, Environment and Urban Systems*, 96, 101845. <https://doi.org/10.1016/j.compenvurbsys.2022.101845>
- Liu, Y. (2022). Diagnosis of Parkinson's disease based on SHAP value feature selection. *Journal Name*, 42(3), 856–869. <https://doi.org/10.1016/j.bbe.2022.06.007>
- Liu, Y., Wang, Z., & Zhang, X. (2024). Estimating the final fatalities using early reported death count from the 2023 Kahramanmaraş, Türkiye, MS 8.0–7.9 earthquake doublet and revising the estimates over time.

- Earthquake Research Advances*, 100331. <https://doi.org/10.1016/j.eqrea.2024.100331>
- Magdeburger Sigorta. (2024). Türkiye'deki deprem bölgeleri. *Magdeburger Sigorta Blogu*. <https://www.magdeburger.com.tr/blog/turkiye-deki-deprem-bolgeleri-nerelerdir>
- Marcílio, W. E., & Eler, D. M. (2020). From explanations to feature selection: Assessing SHAP values as a feature selection mechanism. In *Proceedings of the 33rd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. <https://doi.org/10.1109/SIBGRAPI51738.2020.00053>
- Mosca, E., Szigeti, F., Tragianni, S., Gallagher, D., & Groh, G. (2022, October). SHAP-based explanation methods: A review for NLP interpretability. In *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 4593–4603).
- Parmaksız, H., & Karakuzu, C. (2022). Performance analysis of Extreme Learning Machine classifiers on radio frequency fingerprinting. *BSEU Journal of Engineering Research and Technology*.
- Rajasekaran, S., Latha, V., & Lee, S. C. (2006). Generation of artificial earthquake motion records using wavelets and principal component analysis. *Journal of Earthquake Engineering*, 10(5), 665–691.
- Slundberg, S. (n.d.). SHAP GitHub source. GitHub. <https://github.com/slundberg/shap>.
- Tianqi, C., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. <https://doi.org/10.1145/2939672.2939785>
- Toğaçar, M. (2020). A deep feature learning model for pneumonia detection applying a combination of mRMR feature selection and machine learning models. *Journal Name*, 41(4), 212–222. <https://doi.org/10.1016/j.irbm.2019.10.006>
- Vergara, J. R., & Estévez, P. A. (2014). A review of feature selection methods based on mutual information. *Neural Computing and Applications*, 24, 175–186.
- Wang, H. (2024). Feature selection strategies: A comparative analysis of SHAP-value and importance-based methods. *Journal Name*, 11(1), 44. <https://doi.org/10.1186/s40537-024-00905-w>
- Wang, J., Lu, S., Wang, S. H., & diğerleri. (2022). Aşırı öğrenme makinesi üzerine bir inceleme. *Multimedia Tools and Applications*, 81(4), 41611–41660. <https://doi.org/10.1007/s11042-021-11007-7>