

Makine öğrenmesi algoritmaları ile deprem katalogları kullanılarak deprem tahmini

Earthquake prediction using earthquake catalogs with machine learning algorithm

Ertuğrul DEMİRELLİ*^{1,a} , Halil İbrahim SOLAK^{2,3,b} , İbrahim TİRYAKİOĞLU^{1,3,c} 

¹Afyon Kocatepe Üniversitesi, Mühendislik Fakültesi, Harita Mühendisliği Bölümü, 03204, Afyonkarahisar

²Afyon Kocatepe Üniversitesi, Uzaktan Eğitim MYO, Mimarlık ve Şehir Planlama Bölümü, 03200, Afyonkarahisar

³Afyon Kocatepe Üniversitesi, Deprem Uygulama ve Araştırma Merkezi, 03200, Afyonkarahisar

• Geliş tarihi / Received: 21.03.2023

• Kabul tarihi / Accepted: 02.09.2023

Öz

Ülkemiz taşıdığı coğrafi şartlar gereği doğal afetler, özellikle de deprem gerçeği ile düzenli olarak yüzleşmektedir. Can ve mal kayıplarının büyük bölümünün depremlerde meydana geldiği ve ortalama beş yıllık periyotlarla bu coğrafyanın yıkıcı bir depremle sarsıldığı düşünülürse, deprem afeti alınacak önlemler bakımından ilk sırada gelmektedir. Depremler için alınabilecek önlemlerin belirlenmesi için depremlerin önceden tahmin edilebilmesi büyük önem arz etmektedir. Bu bağlamda son yıllarda makine öğrenmesi ile deprem tahmini çalışmaları hız kazanmıştır. Bu çalışmada deprem kataloğu ile jeolojik veriler ve jeodezik verilerin birleştirildiği bir veri seti kullanılarak deprem tahminleri yapılmıştır. Bu veri seti çalışmada kullanılan algoritma modellerini eğitmek ve eğitilen modellerin performansını ölçmek adına test ve eğitim verisi olarak bölünmüştür. Rastgele orman, ekstrem gradyan artırma, karar ağacı ve k en yakın komşu regresyon algoritmaları kullanılarak eğitim seti ile modeller eğitilmiş, eğitilen modeller test verisi ile test edilmiştir. Analiz sonuçları karşılaştırılarak değerlendirilmiştir. Analiz sonuçlarına göre rastgele orman ve ekstrem gradyan artırma regresyon algoritmaları en başarılı sonuçların alındığı algoritmalar olmuştur. Ortalama kare hatası (MSE) değerleri incelendiğinde, en iyi sonuçlar deprem bilgileri, gerinimler ve fay bilgilerinden oluşan veri setinde gözlenmiştir. MSE için rastgele orman ve ekstrem gradyan artırma algoritmaları ile 0.09, karar ağacı algoritması ile 0.16, k en yakın komşu algoritması ile 0.11 değerleri elde edilmiştir. Yapılan bu çalışma makine öğrenmesi ile deprem tahmini çalışmalarında, kullandığı veri seti ile farklı bir bakış açısı getirerek literatüre katkıda bulunmuştur.

Anahtar kelimeler: Deprem, Deprem kataloğu, Makine öğrenmesi, Regresyon

Abstract

Due to its geographical conditions, our country regularly faces the reality of natural disasters, especially earthquakes. Considering that most of the loss of life and property is caused by earthquakes and this region is shaken by a devastating earthquake in an average of five years, earthquake disaster comes first in terms of precautions to be taken. It is of great importance to be able to predict earthquakes in order to determine the precautions that can be taken for earthquakes. In this context, earthquake prediction studies with machine learning have gained momentum in recent years. In this study, earthquake predictions were made using an earthquake catalog and a data set combining geological and geodetic data. This data set is divided into test and training data in order to train the algorithm models used in the study and to measure the performance of the trained models. By using random forest, extreme gradient boosting, decision tree and k nearest neighbor regression algorithms, the models were trained with the training set and the trained models were tested with the test data. Analysis results were compared and evaluated. According to the analysis results, random forest and extreme gradient increment regression algorithms were the algorithms with the most successful results. When the mean square error (MSE) values are examined, the best results are observed in the data set consisting of earthquake information, strains and fault information. This study contributed to the literature by bringing a different perspective to the data set used in earthquake prediction studies with machine learning.

Keywords: Earthquake, Earthquake catalog, Machine learning, Regression

*a Ertuğrul Demirelli; ertugruldemirelli@gmail.com

1. Giriş

1.1. Introduction

İnsanoğlu varolageldiği günden buyana geçen süreçte doğal afetlerle mücadele etmiştir. Afetler insanlığa ve medeniyetlere büyük zararlar vermiştir. İnsanlığa kültürel, sosyal ve ekonomik anlamda büyük zarar veren yüzü aşkın afet çeşidi vardır (Genç, 2007). Toplumların birikimlerini dakikalar hatta saniyeler içinde yok edebilen deprem bu afetlerden en tehlikelisi olarak kabul edilebilir (Gündoğdu vd., 2012). Ülkemiz afet yönetimi ve afetlere yönelik istatistiklerin verildiği (AFAD, 2018) tarafından yapılan çalışmada, ülkemizde afet kaynaklı ölümlerin %60'ı depremler nedeniyle olmuştur. Ortalama olarak yılda 1000 vatandaşımız hayatını kaybetmekte, 2100 vatandaşımız yaralanmakta ve 7000 bina hasar görmektedir. Deprem, yer kabuğunda bulunan ve fay olarak isimlendirilen kırıklarda biriken enerjinin ani boşalması ile açığa çıkan yer değiştirmenin neden olduğu karmaşık elastik dalga hareketleri olarak tanımlanmaktadır. Yapılan bilimsel çalışmalar doğrultusunda depremin oluş nedenlerinin yanı sıra deprem kavramının nicel özellikleri de ortaya çıkarılmıştır. Günümüzde meydana gelmiş bir depremin büyüklüğü, enlem-boylam olarak merkezi, derinliği bilinebilmektedir (Gündoğdu, 1986; Işık vd., 2012).

Dünya üzerindeki nüfusun artması ve endüstriyelleşmeden dolayı toplumlar daha geniş yerleşim alanlarına dağılmış ve bu yerleşim alanları artan bir ivme ile büyümüştür. Toplumların tektonik açıdan aktif veya bu bölgelere komşu bölgelere yerleşmesi, bu bölgelerde yaşayan insanların depremden etkilenme risk durumlarının artmasına neden olmuştur. 1994 yılında Kaliforniya'da meydana gelen Northridge depremi ($M_w = 6.7$), 1995 yılında Japonya'da meydana gelmiş olan Kobe depremi ($M_w = 7.5$) ve Tayland'da meydana gelen Chi-Chi depremi ($M_w = 7.6$) ciddi sayıda can kayıplarına, ekonomide milyarlarca dolarlık kayıplara neden olmuş depremlerdir (Utkucu & Alptekin, 2001; Utkucu vd., 2011). Ülkemiz de coğrafi olarak yükseltili, eğimli ve fay hatlarının bulunduğu bir bölgede yer almasından dolayı doğal afetler sıklıkla yaşanmaktadır (Genç, 2007). 2020 yılında meydana gelmiş olan Elazığ depremi ($M_w = 6.8$), Sisam Adası açıklarında meydana gelen Sisam depremi ($M_w = 6.9$) ve 2023 yılında Kahramanmaraş'ta meydana gelen depremler ($M_w = 7.7$, $M_w = 7.6$) son yıllarda, ülkemizde can ve mal kaybına neden olmuş depremlerdir. Ülkemizde yaşanan depremler ve sonuçları göz önüne alındığında, afet yönetimi üzerine çeşitli çalışmaların yürütülmesi gerekliliği kaçınılmaz bir gerçektir. Afet yönetimi çalışmalarında gelişen bilgisayar teknolojilerinden, son yılların yükselen değerleri olan makine öğrenmesi ve yapay zeka gibi yöntemlerden faydalanmak, analizlerin daha doğru ve hızlı şekilde elde edilmesini sağlayacaktır.

Son dönemde teknoloji alanında meydana gelen hızlı gelişmeler ile birlikte yapay zekâ ve makine öğrenmesi birçok alanda kullanılmaya başlanmıştır. Literatürde makine öğrenmesinin farklı disiplinlerde kullanıldığı görülmektedir. Sağlıkta hastalık teşhisi ve tedavi yöntemleri, finans, bankacılık, siber saldırıların tespiti ve inşaatta depreme dayanıklı mimari tasarım gibi alanlarda kullanıldığı görülmektedir (Başer vd., 2021; Kaya & Yıldız, 2014; Bingöl vd., 2020).

Diyabet hastalığı tahmini için (Özkan vd., 2022) yaptıkları çalışmada k-en yakın komşu, rastgele orman ve karar ağacı algoritmaları yanında farklı algoritmalar kullanılmıştır. Çalışmada diyabet tanısı için birden fazla yaklaşım uygulanmıştır. En iyi sonuçların rastgele orman bayesnet algoritmaları ile elde edildiği görülmüştür. BİST bankacılık endeksinde yer alan bazı bankaların fiyat tahmini için (Duygu & Kılıç, 2019) çoklu regresyon, rastgele orman ve karar ağacı algoritmaları kullanılmıştır. Model eğitimi için teknik ve temel göstergelerden oluşan 2 farklı veri seti kullanılmıştır. Analiz sonuçlarına göre tüm algoritmalar 0.95 R2 skoru üzerinde bir başarı göstermiştir. Bu çalışmaların yanı sıra makine öğrenmesi algoritmaları yer bilimlerinde de kullanılmaktadır. Hava kirliliği tahmini ile ilgili (Ünalı & Yalçın, 2019) tarafından yapılan çalışmada rastgele orman, k-en yakın komşu, karar ağacı ve çoklu doğrusal regresyon gibi algoritmalar kullanılmıştır. Havadaki CO, SO₂, O₃ vb. kirlenici konsantrasyonları bu algoritmalar ile tahmin edilmiştir. Tahmin sonuçlarında en başarılı algoritmaların rastgele orman ve çoklu doğrusal regresyon algoritmalarının olduğu görülmüştür. Literatürde deprem tahmini üzerine de farklı çalışmalar mevcuttur.

Gelecek bir tarihte meydana gelebilecek depremlerin tahmin edilmesi ile ilgili çalışmalar incelendiğinde farklı yöntemler ve veri setleri ile yapılan çalışmaların olduğu görülmüştür. Yunanistan bölgesi için (Moustra, 2011) tarafından, deprem öncesinde ortaya çıktığı düşünülen sismik elektrik sinyalleri kullanılarak deprem tahmini üzerine bir çalışma yapılmıştır. Bir diğer çalışma ise (Rouet-Leduc vd., 2017) tarafından rastgele orman algoritması kullanılarak laboratuvar ortamında yapılmıştır. Çalışmadaki amaç laboratuvar ortamında

depremler oluşturarak fay bloklarının hareketi ile ortaya çıkan akustik sinyallerin elde edilmesi ve bu sinyaller ile deprem tahmini yapılmasıdır. Aynı şekilde (Lubbers vd., 2018) laboratuvar ortamında rastgele orman algoritmasını kullanarak verileri katalog haline getirmişler ve akustik sinyallerinin deprem büyüklüğüne etkisini gözlemlemişlerdir. Bir diğer çalışmada ise (Gitis & Derendyaev, 2019) makine öğrenmesi algoritmaları ve deprem kataloglarını kullanarak depremin zamansal ve mekânsal tahmini üzerine çalışmalar yapmışlardır.

Bu çalışmada makine öğrenmesi algoritmaları kullanılarak ileri bir tarihte meydana gelebilecek depremlere ait yaklaşık büyüklüklerin tahmini amaçlanmıştır. Burada elde edilecek tahmin değerleri ile ileriye dönük riskli bölgelerin tespit edilmesi planlanmaktadır. Makine öğrenmesi ile tahmin süreci ilk olarak veri ön işleme ardından tahmin modellerinin eğitimi ve son olarak da sonuçların doğruluklarının değerlendirilmesi olmak üzere üç adımdan oluşmaktadır. Süreç içinde kullanılan veriler tarihsel deprem katalogları, Türkiye’de yer alan diri fay bilgileri gibi jeolojik veriler ve jeodezik ölçmelerle elde edilmiş gerinimlerden oluşmaktadır. Katalog verileri deprem oluş tarihi, enlem, boylam, derinlik (KM), büyüklük (xM), jeolojik veriler fay uzunluğu ve depremin faya olan uzaklığı, jeodezik veriler ise gerinimlerden oluşmakta olup Tablo 1’de gösterilmiştir. Bu veriler kullanılarak ileri bir tarihte oluşabilecek deprem büyüklüğü tahmin edilmeye ve bunun için en iyi algoritma seçilmeye çalışılmıştır.

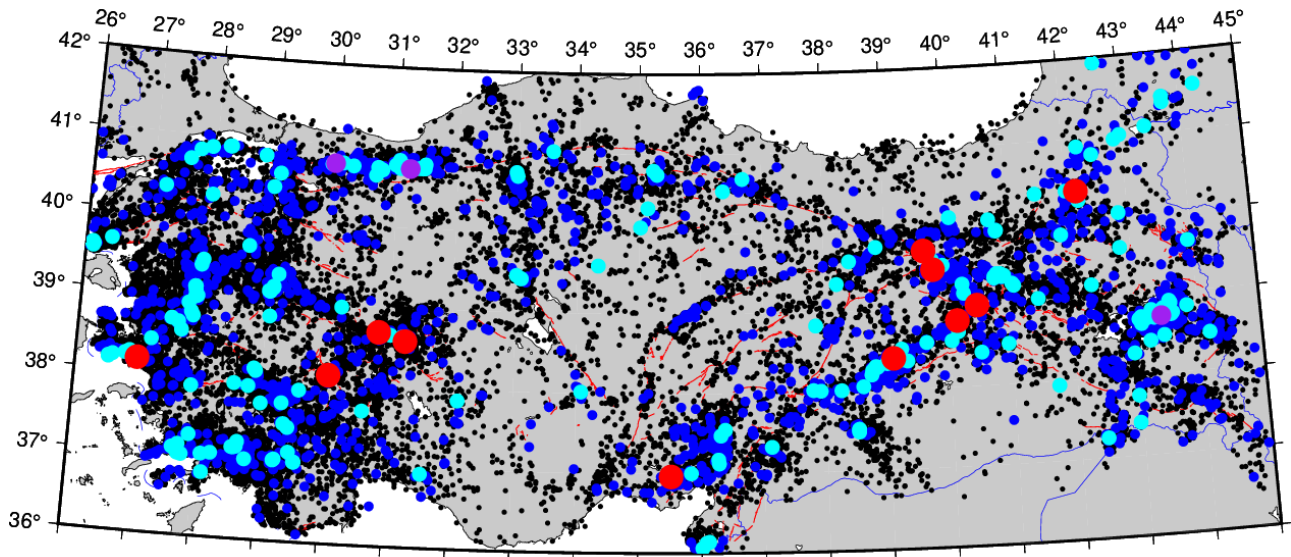
2. Materyal ve metod

2. Material and method

2.1. Veri setinin oluşturulması

2.1. Creating the dataset

Çalışma kapsamında veri seti oluşturabilmek için öncelikle Kandilli Rasathanesi ve Deprem Araştırma Enstitüsü’nün (KRDAE) deprem kataloğunda yer alan depremler filtrelenmiştir. KRDAE kataloğunda 1900 – 2021 yılları arasında ülkemizde meydana gelen depremlere ait ~350000 kayıt yer almaktadır (Şekil 1). Katalog depremleri Şekil 1’de harita üzerinde gösterilmiştir. Bu depremlerden yaklaşık 10500’ü Batı Anadolu’da $37.90^{\circ} - 40.10^{\circ}$ enlemleri ve $26.16^{\circ} - 30.46^{\circ}$ boylamları arasındadır. Bu bölge hem yoğun deprem aktivitesi hem de karmaşık tektonik yapısı nedeniyle çalışma bölgesi olarak seçilmiş ve bu depremler veri setine dahil edilmiştir. Veri setinde bulunan depremler ülkemizde meydana gelmiş ve komşu ülkelerde meydana gelerek ülkemizde de hissedilen depremlerden oluşmaktadır. Deprem kataloğunda bulunan her bir veri için deprem oluş tarihi, depremin enlem ve boylam değerleri, derinliği ve deprem büyüklüğü öznitelik bilgileri bulunmaktadır.



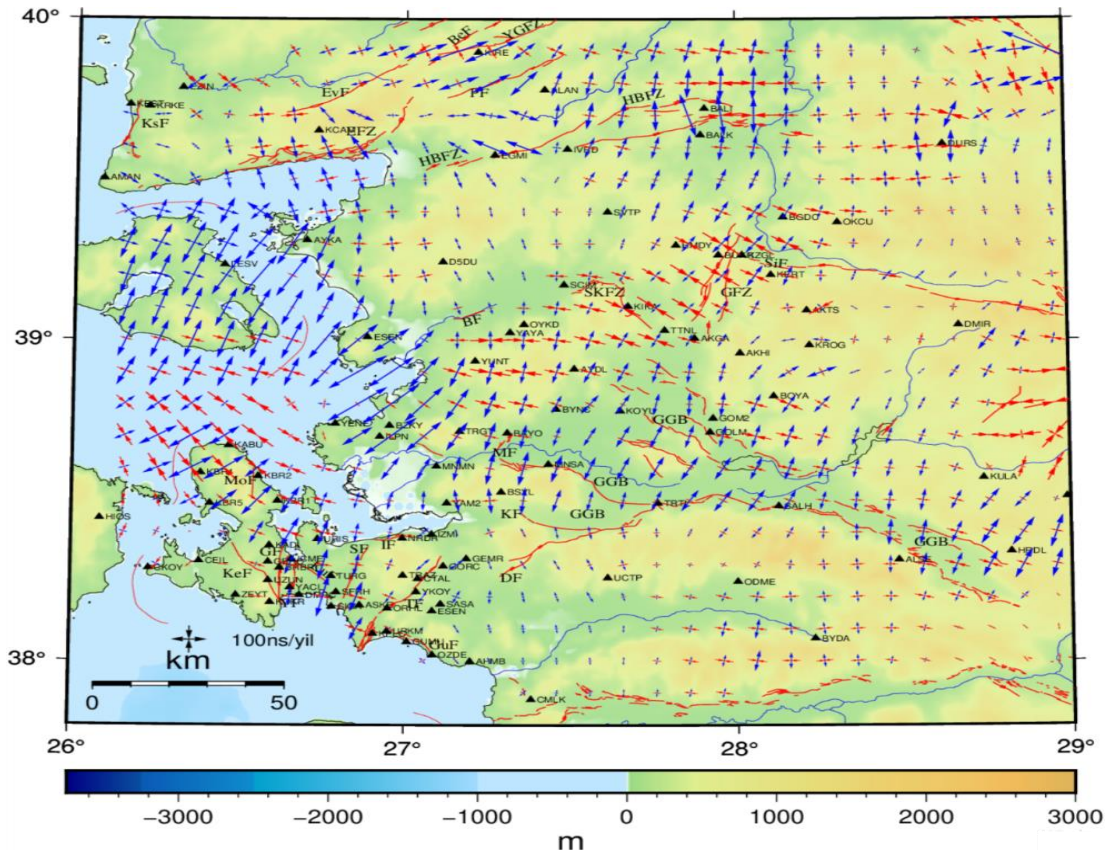
Şekil 1. Türkiye’de aletsel dönemde meydana gelen $M \geq 3$ depremler (KRDAE) (kırmızı çizgi: aktif fay, siyah daire: M:3., mavi daire: M:4., turkuaz daire: M:5., mor daire: M:6., kırmızı daire: M:7)

Figure 1. $M \geq 3$ earthquakes in the instrumental period in the working area (red line: active fault, black circle: M:3., blue circle: M:4., turquoise circle: M:5., purple circle: M:6., red circle: M:7)

Depremlerin oluşumunda en önemli parametre olarak öne çıkan faylara ait veriler ise MTA'nın (Maden Tetkik ve Arama) diri fay haritaları ve ek olarak (Emre vd., 2018)'den elde edilmiştir. Fay uzunluğunun deprem büyüklüğünü önemli ölçüde etkilemesi nedeniyle her bir faya ait uzunluklar, o fay kırıkları arasındaki mesafeler toplanarak km cinsinden elde edilmiştir. Ayrıca her fay için bir id atanmıştır. Bunlara ek olarak fayların tipleri de (RL: Sağ yanal, N: Normal Fay, LL: Sol yanal, R: Ters vb.) (Emre vd., 2018)'den alınmıştır. Fay bilgilerinin model eğitiminde önemli veriler olduğu düşünüldüğünden, hesaplanan fay uzunlukları, deprem-fay arasındaki mesafeler ve fay tipleri veri setine eklenmiştir.

Makine öğrenmesi uygulamalarında modelin başarısını arttırmak için sistemi etkileyen girdi verilerinin iyi tanımlanması gerekmektedir. Bu amaçla günümüzde GNSS verilerinden elde edilen jeodezik gerinimler deprem tehlike analizi çalışmalarında etkin olarak kullanılmaktadır. Gerinim bir cisim üzerinde o ana kadar meydana gelen güncel deformasyon hakkında bilgi vermektedir. Tektonik olarak değerlendirildiğinde bu durum levhalar veya daha küçük bloklar üzerinde meydana gelen stres birikimine karşılık gelir. Gerinimler fayın depremselliği ve kinematığı ile ilgili bilgi veren en önemli parametrelerdir. Bir bölgede gerinim hızlarının büyüklüğü oradaki sismik tehlikeyi ifade etmektedir. (Solak, 2015; Aktuğ, 2017; Araszkiwicz, 2016).

Makine Öğrenmesi için gerçekleştirilecek uygulamaya çalışma bölgesindeki faylara ait bilgiler ile bölgeye ait gerinimlerin de dahil edilmesinin modelin başarısını arttıracak değerlendirilmiştir. Bu amaçla hem modelin başarısını arttırmak hem de gerinimlerin etkisini görmek için ikinci bir veri seti oluşturulmuştur. Bu nedenle (Solak, 2020) tarafından yapılmış olan çalışmada İzmir-Balıkesir bölgesi için verilmiş hız alanları kullanılarak, Geodisuit programı ile bölgenin 0.1° grid aralığında jeodezik gerinimleri hesaplanmıştır (Shen vd. 2015). Elde edilen gerinimler Şekil 2'de gösterilmiştir. Yoğun ağ tasarımı (111 nokta), yüksek hız doğruluğu (max. 1.5 mm/yıl) ve buna bağlı olarak Türkiye geneline kıyasla doğruluğu yüksek gerinim değerleri elde edileceğinden dolayı bu bölge ikinci veri seti için çalışma alanı olarak seçilmiştir.



Şekil 2. Çalışma bölgesi gerinim alanı (kırmızı oklar: sıkışma, mavi oklar: açılma, kırmızı çizgiler: diri fay, mavi kıvrımlı çizgiler: ana nehirler, siyah üçgenler: GNSS noktaları).

Figure 2. Strain rates in the study area (red arrows: compression, blue arrows: extension, red lines: active fault, curved blue lines: main rivers, black triangles: GNSS sites).

Tablo 1. Değişkenler ve veri setinin özellikleri
Table 1. Variables and properties of the dataset

Değişken ismi	Değişken tipi	Değer	Detay
Oluş tarihi	Girdi	Etiket	
Enlem	Girdi	Ondalık	Deprem merkez üssü
Boylam	Girdi	Ondalık	
Der (km)	Girdi	Ondalık	Deprem derinliği
xM	Çıktı	Ondalık	Deprem büyüklüğü
Fay Tipi	Girdi	Etiket	Depreme neden olan fayın türü
Faya uzaklık (km)	Girdi	Ondalık	Depreme neden olan faya uzaklık
Fay uzunluğu (km)	Girdi	Ondalık	Depreme neden olan fayın uzunluğu
Gerinim (ns/yıl)	Girdi	Ondalık	Deprem merkez üssündeki gerinim

2.2. Veri ön işleme

2.2. Data preprocessing

Veri setleri genellikle sonuçları etkileyebilecek hatalı veriler içermektedir ve bu hatalı veriler gürültü olarak tanımlanmaktadır. Veri içerisinde gürültü ne kadar çoksa sonuçların güvenilirliği de o kadar tartışmalı olacaktır (Oğuzlar, 2003). Bu çalışmada 1900 – 1970 yılları arasında meydana gelmiş küçük ölçekli depremlerin teknik yetersizliklerden dolayı katalogta yer almadığı görülmüştür. Veri içerisinde yer alan bu boşluğun sonuçlara olumsuz etkilerini önlemek için 70 yıllık bu bölüm veri setinden çıkartılmış ve çalışmada 1970 yılı sonrası depremler kullanılmıştır. Meydana gelmiş depremler Türkiye sınırları içerisinde, denizde ve ülkemizin komşularında meydana gelmiş depremleri içermesine rağmen kullanılan fay verileri ülke sınırı içindeki karasal fayları içermektedir. Bu nedenle deprem ve depreme neden olan fay eşleştirmelerinde deniz depremleri ve komşu ülkelerde meydana gelen depremlerin fay-deprem eşleştirmelerinden kaynaklanacak hataları elemine etmek için deniz ve ülke sınırı dışında meydana gelmiş depremler de veri seti içinden çıkartılmıştır. Depremlerden sonra meydana gelen artçı depremlerin yarattığı gürültüyü önlemek için artçı depremler veri setinden ayıklanmıştır. Artçı depremlerin ayıklanmasında; M:4. olan depremler için depremin oluş tarihinden itibaren 1 hafta sonrasında ve 5 km yakınında meydana gelmiş depremler, M:5. olan depremler için depremin oluş tarihinden itibaren 2 ay sonrasında ve 15 km yakınında meydana gelmiş depremler, M:6. olan depremler için depremin oluş tarihinden itibaren 3 ay sonrasında ve 20 km yakınında meydana gelmiş depremler ve M:7. olan depremler için depremin oluş tarihinden itibaren 6 ay sonrasında ve 20 km yakınında meydana gelmiş depremler artçı depremler olarak kabul edilmiştir.

37.90 – 40.10 enlemleri ve 26.16 – 30.46 boylamları arasında elde edilmiş olan gerinimler bu enlem ve boylam aralıklarında meydana gelmiş ~10500 depreme eklenmiştir. Veri setinin bir bölümü Tablo 2`de, veri setine ilişkin istatistik bilgileri ise Tablo 3`de gösterilmiştir. Gerinimler eklenirken ters mesafe ağırlıklı enterpolasyon yöntemi kullanılmış ve her depremin içinde bulunduğu gridin köşe noktaları kullanılarak deprem merkez üssü koordinatlarındaki gerinimler hesaplanmıştır. Gerinimlerin, fay bilgilerinin eklenmesi ve artçı depremlerin çıkarılması ile veri seti son halini almıştır.

Tablo 2. Deprem verileri ve öznitelikleri. (NN: Normal, RL: Sağ atımlı. Depremlerin oluş tarihleri jülyen gününe çevrilerek etiketlenmiştir.)

Table 2. Earthquake data and attributes (NN: Normal, RL: Right Lateral. earthquake occurrence dates have been added by converting the julian day)

Olus tarihi (Julian Günü)	Tarih (gg.aa.yy)	Enlem ^o	Boylam ^o	Der(km)	xM	Fay tipi	Faya uzaklık(km)	Fay uzunluk(km)	Gerinim (ns/yıl)
2442450.5	07.02.1975	38.75	28.35	5	4.1	RL	441	56	68.54
2442455.5	12.02.1975	39.14	29	15	4.2	RL	368.8	56	64.29
2442562.5	19.02.1975	39.12	27.68	10	4	NN	514.9	56	107.75
2442562.5	19.02.1975	38.75	27.6	9	4.2	RL	524.1	56	82.49
2459566.5	18.12.2021	39.0075	27.8813	8.9	3.3	RL	492.5	56	74.83
2459570.5	22.12.2021	39.8318	29.1303	4.8	3.2	RL	10.52	29	247.17
2459578.5	30.12.2021	39.2802	28.2493	7.3	3.3	RL	452.6	56	64.49

Ters mesafe ağırlıklı enterpolasyon yöntemde dayanak noktaları ile gerinim değeri belirlenecek noktanın aralarındaki mesafenin tersi ağırlıklandırılarak ilgili noktanın gerinim değeri Denklem 1`de verilen eşitliğe göre hesaplanmıştır.

$$g_e = \frac{\sum_{i=1}^n S_i * G_i}{\sum_{i=1}^n S_i} \quad (1)$$

Eşitlikte hesaplanan g değeri depremin meydana geldiği noktadaki gerinim değerini, G değeri ise dayanak noktalarındaki gerinim değerini ifade etmektedir. Denklemden verilen S değeri ağırlık değerleri, n değeri ise dayanak noktalarının sayısı için kullanılmaktadır. Ağırlık değerlerinin hesaplanması için kullanılan formül Denklem 2`de gösterilmiştir.

$$S_i = 1/d_i, i = 1,2,3,4 \dots \quad (2)$$

Ağırlıkların hesaplanmasında kullanılan d değeri dayanak noktası ve gerinim değeri hesaplanan nokta arasındaki mesafeyi belirtmektedir. Noktalar arasındaki mesafe hesaplanırken python/geopy kütüphanesinden yararlanılmıştır.

Tablo 3. Veri setinin istatistik bilgileri. (Verinin çeyreklik kısımlarının (25%, 50% ve 75%) maksimum değerleri gösterilmiştir.)

Table 3. Statistical information of the data set. (The maximum values of the quartiles (25%, 50% and 75%) of the data are given.)

	count	mean	std	min	25%	50%	75%	max
Oluş tarihi (Julyen günü)	10364	2451168.47	4446.79	2442427.5	2448197.75	2451344.5	2454884.5	2459578.5
Enlem (°)	10364	38.94	0.56	37.9	38.47	39.07	39.32	40.1
Boylam (°)	10364	28.21	1.04	26.16	27.5	27.99	29.06	30.45
Der(km)	10364	7.17	5.32	0.0	4.2	6.6	10.0	30.0
xM	10364	3.26	0.32	3.0	3.0	3.2	3.4	6.0
Faya uzaklık (km)	10364	345.57	214.01	0.33	211.4	377.7	512.5	675.4
Fay uzunluk (km)	10364	53.07	12.51	9.0	56.0	56.0	56.0	90.0
Gerinim (ns/yıl)	10364	76.4	46.44	12.2	49.48	71.48	91.15	389.33

2.3. Algoritma seçimi

2.3. Algorithm selection

Deprem gerçekleşme olayının lineer olarak tanımlanması mümkün olmamasından dolayı lineer çözüm sunan makine öğrenmesi algoritmaları tercih edilmemiştir. Bunun yerine daha komplike modelleme çalışmalarında yüksek doğrulukta sonuçlar sunan K-en yakın komşu (K-nearest neighbor, KNN), Karar ağacı (Decision Tree, DT), Rastgele orman (Random Forest, RF) ve Ekstrem gradyan artırma (Extreme Gradient Boosting, XGBoost) regresyon algoritmaları tercih edilmiştir.

K-en yakın komşu (K-nearest neighbor, KNN): K en yakın komşu algoritması regresyon ve sınıflandırma problemlerinde kullanılan en temel denetimli öğrenme algoritmalarından biridir. Algoritma örnek tabanlı öğrenme esasına göre tahminler yapmaktadır. Bu tip örnek tabanlı olan öğrenme algoritmalarında, eğitim setinde yer alan veriler kullanılarak öğrenme işlemi gerçekleştirilir. Algoritma tahmini istenen örneğin en yakın komşuları ile olan ilişkisine bağlı olarak tahminde bulunmaktadır. En yakın komşular en küçük Öklid mesafesi baz alınarak bulunabildiği gibi Manhattan, Minkowski ve Chebyshev fonksiyonları kullanılarak da bulunabilir. (Taşçı & Onan, 2016; Altunkaynak vd., 2020).

Karar ağacı (Decision Tree, DT): Karar ağacı regresyon ve sınıflandırma uygulamalarında veriyi ağaç tipinde ele alan bir algoritmadır. İşlem sırasında kullanılan veri topluluğu her defasında yapraklar, dallar ve

karar düğümlerinden oluşan alt kümelere bölünür. Öznitelikler düğümler tarafından temsil edilmektedir. Yapraklar ve düğümler ile düğümler arasındaki temsil etmek için ise dallar kullanılmaktadır. Karar ağaçlarının oluşturulmasının temelinde verilere sorulan sorular ile bir karar kurallarının bulunmasıdır. Karar ağaçlarındaki en önemli aşama ağaç yapısındaki dallanmanın hangi kurala veya özneliğe göre şekilleneceğinin belirlenmesidir (Kavzoğlu vd., 2012).

Rastgele orman (Random Forest, RF): Rastgele orman tekniği çok sayıda karar ağacının performansını birleştirerek tahmin veya sınıflandırma yapmaktadır (Rodriguez-Galiano vd., 2015). Bu algoritma kategorik, sürekli veya ikisini birden barındıran veri setlerinde kullanılabilir gibi aynı zamanda küçük veya büyük yoğunluktaki veri setlerinde de rahatlıkla kullanılmaktadır (Akman vd., 2011; Korkmaz vd., 2018). Rastgele orman algoritması diğer algoritmalara göre doğruluk açısından daha tatmin edici olmakla beraber algoritmadaki ağaç yapısının görülememesi dezavantaj olarak görülmektedir (Akin & Terzi, 2020).

Ekstrem gradyan artırma (Extreme Gradient Boosting, XGBoost): XGBoost algoritması 2016 yılında tahmin algoritmalarına alternatif olarak geliştirilmiştir. Algoritma gradyan artırma ve karar ağacı temelli çalışmaktadır. Yarışmalarda gösterdiği başarı ile dikkatleri üzerine çekmiş olan XGBoost algoritması sınıflandırma ve regresyon problemlerinde başarı ile kullanılmaktadır. Öğrenme aşamasında ağaçları oluştururken CART algoritmasını kullanan XGBoost algoritması, kaynak kullanımına dikkat etmesi, hesaplama hızı ve model performansına odaklanması gibi konularda kullanıcıya avantajlar sunmaktadır (Verma vd., 2018; Morde, 2019; Brownlee, 2016; Kurt vd., 2020).

2.4. Model başarısının belirlenmesi

2.4. Determination of modelling performance

Çalışmanın en önemli adımlarından birisi de tahmin doğruluklarının ve kullanılan algoritmaların performanslarının karşılaştırılmasıdır. Makine öğrenmesi algoritmalarında performans ölçmekte kullanılan temel metrikler MSE (Mean Squared Error) ve RMSE (Root Mean Squared Error) olarak sayılabilir (Karasu vd., 2018; Gültepe, 2019). Bu çalışmada performans ölçmekte kullanılan MSE ve RMSE değerlerinde elde edilen düşük değerler ters orantılı olarak performansın ne kadar iyi olduğunu göstermektedir. Bu değerler sıfıra yaklaştıkça modelin daha iyi bir performans sergilediği söylenebilir (Wang & Xu, 2004; Çınaroğlu, 2017).

Algoritmanın yaptığı tahminler ve gerçek değerler arasındaki farkların karelerinin ortalaması karesel ortalama hatayı (MSE) vermektedir. RMSE değeri ise hesaplanan MSE değerinin karekökünün alınması ile bulunmaktadır. MSE değerinin hesaplama formülü Denklem 3'de ve RMSE değerinin hesaplama formülü Denklem 4'de verilmiştir.

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n e_t^2 \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n e_t^2} \quad (4)$$

3. Uygulama ve algoritma performanslarının belirlenmesi

3. Application and determination of algorithm performances

3.1 Uygulama

3.1. Application

Oluşturulan veri setleri ile ML algoritmalarını test etmek için SciKit-Learn kütüphanesi kullanılmıştır. Bu kütüphane Python programlama dili ile yazılmış olup çok sayıda regresyon ve sınıflandırma algoritmalarını içeren bir makine öğrenmesi kütüphanesidir. Python kütüphaneleri olan NumPy ve Matplotlib kütüphaneleri üzerine inşa edilmiştir (Buitinck vd., 2013). Model başarısını artırmak için GridSearchCV ile belirlenen algoritma hiper-parametrelerinin tüm kombinasyonları denenmiş ve en ideal değerler kullanılmıştır. K-en

yakın komşu algoritması için en ideal k komşu sayısı değeri belirlenirken, karar ağacı, rastgele orman ve ekstrem gradyan arttırımı algoritmaları için ise max_depth, min_samples_leaf gibi parametrelerin en ideali kullanılmıştır.

Her iki veri seti de %70 eğitim verisi ve %30 test verisi olmak üzere ikiye bölünmüştür. Öncelikle deprem verileri ve fay bilgilerinden oluşan birinci veri seti, sonrasında ise deprem verileri, fay bilgileri ve gerinimlerden oluşan ikinci veri seti eğitim verisi ile eğitilmiş, sonrasında ise test verileri ile test edilmiştir. Elde edilen sonuçlar yorumlama ve karşılaştırma kolaylığı açısından Tablo 4`de sunulmuştur.

Tablo 4. Algoritmalar ve hata değerleri.
Table 4. Algorithms and error value.

Algoritmalar	1. veri seti (Deprem verileri ve fay bilgileri)			2. veri seti (Gerinimler ve fay bilgileri eklenmiş)		
	MSE	RMSE	Accuracy	MSE	RMSE	Accuracy
	K en yakın komşu	0.11	0.33	0.92	0.11	0.33
Karar ağacı	0.19	0.43	0.90	0.16	0.40	0.92
Rastgele orman	0.14	0.37	0.92	0.09	0.30	0.94
XGBoost	0.14	0.37	0.91	0.09	0.30	0.92

Tablo 4 incelendiğinde deprem verileri ve fay bilgilerinin içeren birinci veri seti için en iyi sonucu sırasıyla RMSE:0.33 ve MSE:0.11 değerlerine sahip olan K en yakın komşu algoritması vermiştir. Rastgele orman, XGBoost ve karar ağacı algoritmaları için bu değerler ise sırasıyla MSE: 0.14-0.14-0.19 ve RMSE: 0.37-0.37-0.43'dür. En yüksek accuracy (doğruluk) değeri ise 0.92 ile k-en yakın komşu ve rastgele orman algoritmalarında elde edilmiştir.

Tablo 4 incelendiğinde günümüzde deprem tehlike analizinde aktif olarak kullanılan fay bilgileri ile gerinimlerin birlikte modele girdi olarak verildiği 2. veri setinde MSE değerlerinin 0.09-0.16, RMSE değerlerinin ise 0.30-0.40 arasında değiştiği görülmektedir. Buna göre en iyi sonucu Rastgele orman ve XGBoost algoritmaları vermiştir (RMSE:0.30 ve MSE:0.09). K en yakın komşu ve Karar ağacı algoritmaları ise sırasıyla MSE: 0.11 – RMSE: 0.33 ve MSE: 0.16 – RMSE: 0.40 değerlerine sahiptir. Bu veri setinde ise en iyi accuracy değeri rastgele orman algoritması ile elde edilmiştir.

Tüm sonuçlar birlikte değerlendirildiğinde Rastgele orman ve XGBoost algoritmalarının ikinci veri setinde en iyi sonuçları verdiği, K en yakın komşu algoritmasının ise birinci veri setinde en iyi değeri verdiği görülmektedir. Ancak burada daha da önemli olan, depremi tanımlayan parametrelerden oluşan girdi sayısı arttıkça sistemin öğrenme becerisinin artmış, hata payının azalmış olmasıdır. Bu durum deprem tekrarlama periyodu, faylara ait tarihsel dönemdeki deprem kayıtları, yüzey kırıkları vb. bilgiler ile daha yüksek doğrulukta sonuç elde edilebileceğini göstermektedir.

4. Tartışma ve sonuçlar

4. Discussion and conclusions

Bu çalışmada deprem kataloglarında yer alan veriler, faylara ait bilgiler ve güncel jeodezik gerinimler kullanılarak deprem kestirimi yapılması amaçlanmıştır. Bu nedenle KRDAE deprem katalogunun (1900-2021) kendisi ve bu katalogdan, artçı depremler çıkartılarak, depremler ile faylar ilişkilendirilerek ve fay bilgilerinin yanına gerinimler eklenerek elde edilen veri seti olmak üzere iki adet veri seti elde edilmiştir. Elde edilen bu iki veri seti üzerinde aynı algoritmalar kullanılmıştır. Dört farklı algoritma ile yapılan testlerde, model performansını belirlemek için MSE ve RMSE değerleri baz alınmıştır. Modelin eğitim ve test aşamalarında, veri setinin %30'u test verisi %70'i eğitim verisi olarak kullanılmıştır. Sıfıra en yakın değerlerin en iyi

performansı verdiği göz önünde bulundurulursa, en başarılı algoritmaların ikinci veri setinde rastgele orman ve ekstrem gradyan arttırma (XGBoost) algoritmalarının, birinci veri setinde ise k en yakın komşu algoritmasının olduğu görülmektedir. En düşük performansı sergileyen algoritmanın ise her iki veri setinde karar ağacı algoritması olduğu görülmüştür. Depremler ve fay bilgilerinin yer aldığı veri seti ile yapılan çalışmada en düşük 0.11 bulunan karesel ortalama hata (MSE) değerinin, gerinimler ve fay bilgilerinin eklendiği veri seti ile 0.09 karesel ortalama hata değerine kadar düştüğü görülmüştür. Gerinim değerleri ve fay bilgilerinin eklenmesi ile rastgele orman ve XGBoost algoritmalarının sonuçlarında gözle görülür bir iyileşme görülmektedir.

Literatürde (Lubbers vd., 2018) tarafından tamamlanan çalışmada elde edilen akustik sinyaller veri setinin büyütülmesi ve (Moustra vd., 2011) tarafından yapılan çalışmalarda kullanılan sismik elektrik sinyalleri verisindeki eksik kısımların tamamlanarak girdi sayısının artırılması gibi müdahaleler ile sonuç doğrulukları artırılmaktadır. Bunun yanında (Leduc vd., 2017) çalışmasında belirtildiği gibi jeodezik verilere ek olarak akustik sinyaller ile laboratuvar ortamında elde edilen veriler Dünya ölçeğinde genişletilerek çalışma doğruluğu artırılabilir. Son olarak (Gitis & Derendyaev, 2019) tarafından tamamlanan çalışmada da değinildiği gibi bu sonuçlar deprem tahmini problemi üzerine yapılan çalışmalara katkıda bulunabilir. Bu çalışmalardan yola çıkarak ileride yapılacak çalışmalarda paleosismolojik verilerin eklenmesi, gerinimlerin tüm ülke ölçeğinde çalışılarak veriye eklenmesi veya veri seti içerisine farklı parametrelerin eklenmesi ile başarı oranının yükseltilebileceği değerlendirilmektedir.

Teşekkür

Acknowledgement

Bu çalışma, 22.FEN.BİL.31 nolu Bilimsel Araştırma Projesi olarak Afyon Kocatepe Üniversitesi tarafından desteklenmiştir. Makalenin inceleme ve değerlendirme aşamasında yapmış oldukları katkılardan dolayı derginin editör ve hakemlerine teşekkür ediyorum.

Yazar katkısı

Author contribution

Bu çalışmanın bölümlerine tüm yazarlar tarafından katkı verilmiştir.

Etik beyanı

Declaration of ethical code

Bu makalenin yazarı, bu çalışmada kullanılan materyal ve yöntemlerin etik kurul izni ve / veya yasal-özel izin gerektirmediğini beyan etmektedir.

Çıkar çatışması beyanı

Conflicts of interest

Yazarlar herhangi bir çıkar çatışması olmadığını beyan ederler.

Kaynaklar

References

- Afet ve Acil Durum Yönetimi Başkanlığı (AFAD), (2018). *Türkiye`de afet yönetimi ve doğa kaynaklı afet istatistikleri*. https://www.afad.gov.tr/kurumlar/afad.gov.tr/35429/xfiles/turkiye_de_afetler.pdf
- Akın, P., & Terzi, Y. (2020). Dengesiz veri setli sağkalım verilerinde cox regresyon ve rastgele orman yöntemlerin karşılaştırılması. *Veri Bilimi*, 3(1), 21-25.
- Akman, M., Genç, Y., & Ankaralı, H. (2011). Random forests yöntemi ve sağlık alanında bir uygulama/random forests methods and an application in health science. *Türkiye Klinikleri Biyoistatistik*, 3(1), 36-48.
- Akşehir, Z. D., & Kılıç, E. (2019). Makine öğrenmesi teknikleri ile banka hisse senetlerinin fiyat tahmini. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 12(2), 30-39.

- Aktuğ, B. (2017). Jeodezik deprem tehlike haritası. *TMMOB Harita ve Kadastro Mühendisleri Odası Türkiye Harita Bilimsel ve Teknik Kurultayı*, Ankara.
- Alptekin, M. U. Ö. (2012). Real-time (gerçek-zamanlı) sismolojisi, dünyadaki gelişimi ve Türkiye'de uygulanabilirliği. *İstanbul Yerbilimleri Dergisi*, 14(1-2).
- Altunkaynak, A., Başakın, E. E., & Kartal, E. (2020). Dalgacık k-en yakın komşuluk yöntemi ile hava kirliliği tahmini. *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 25(3), 1547-1556. <https://doi.org/10.17482/uumfd.809938>
- Araszkievicz, A., Figurski, M., & Jarosiński, M. (2016). Erroneous GNSS strain rate patterns and their application to investigate the tectonic credibility of GNSS velocities. *Acta Geophysica*, 64, 1412-1429. <https://doi.org/10.1515/acgeo-2016-0057>
- Başer, B. Ö., Yangın, M., & Sarıdaş, E. S. (2021). Makine öğrenmesi teknikleriyle diyabet hastalığının sınıflandırılması. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 25(1), 112-120. <https://doi.org/10.19113/sdufenbed.842460>
- Bingol, K., Akan, A. E., Örmecioğlu, H. T., & Er, A. (2020). Artificial intelligence applications in earthquake resistant architectural design: Determination of irregular structural systems with deep learning and ImageAI method. *Journal of the Faculty of Engineering and Architecture of Gazi University* 35:4, 2197-2209. <https://doi.org/10.17341/gazimmfd.647981>
- Brownlee, J. (2016, August 17). *A gentle introduction to xgboost for applied machine learning*. Machine Learning Mastery. <https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/>
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., ... & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project. *arXiv preprint arXiv:1309.0238*.
- Çınaroğlu, S. (2017). Sağlık harcamasının tahmininde makine öğrenmesi regresyon yöntemlerinin karşılaştırılması. *Uludağ Üniversitesi Mühendislik Fakültesi Dergisi*, 22(2), 179-200. <https://doi.org/10.17482/uumfd.338805>
- Emre, Ö., Duman, T. Y., Özalp, S., Şaroğlu, F., Olgun, Ş., Elmacı, H., & Çan, T. (2018). Active fault database of Turkey. *Bulletin of Earthquake Engineering*, 16(8), 3229-3275. <https://doi.org/10.1007/s10518-016-0041-2>
- Genç, F. N., (2007). Türkiye'de doğal afetler ve doğal afetlerde risk yönetimi. *Stratejik Araştırmalar Dergisi*, 5(9), 201-226.
- Gitis, V. G., & Derendyaev, A. B. (2019). Machine learning methods for seismic hazards forecast. *Geosciences*, 9(7), 308. <https://doi.org/10.3390/geosciences9070308>
- Gündoğdu, O. (1986). *Türkiye depremlerinin kaynak parametreleri ve aralarındaki ilişkiler*. [Doktora tezi, İstanbul Üniversitesi Mühendislik Fakültesi]
- Gündoğdu, O., Işık, Ö., & Koç, S. (2012). Marmara ve çevresinde deprem tehlikesi. *Okmeydanı Tıp Dergisi*, 28(2), 71-81. <https://doi.org/10.5222/otd.sup2.2012.071>
- Işık, Ö., Aydınlioğlu, H. M., Koç, S., Gündoğdu, O., Korkmaz, G., & Ay, A. (2012). Afet yönetimi ve afet odaklı sağlık hizmetleri. *Okmeydanı Tıp Dergisi*, 28(2), 82-123. <https://doi.org/10.5222/otd.sup2.2012.082>
- Kavzoğlu, T., Şahin, E. K., & Çölkesen, İ. (2012). Heyelan duyarlılığının incelenmesinde regresyon ağaçlarının kullanımı: Trabzon örneği. *Harita Dergisi*, 147(3), 21-33.
- Kaya, Ç., & Yıldız, O. (2014). Makine öğrenmesi teknikleriyle saldırı tespiti: Karşılaştırmalı analiz. *Marmara University Journal of Science*, 26(3). <https://doi.org/10.7240/mufbed.24684>
- Korkmaz, D., Çelik, H. E., & Kapar, M. (2018). Sınıflandırma ve regresyon ağaçları ile rastgele orman algoritması kullanarak botnet tespiti: Van Yüzüncü Yıl Üniversitesi örneği. *Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 23(3), 297-307.

- Kurt, A., Buldu, B., & Cedimoğlu, İ. H. (2020). Xgboost ve rastgele orman algoritmalarının ağ tabanlı saldırı tespitine yönelik performanslarının karşılaştırılması. *International Marmara Sciences Congress* (pp. 730-736), Kocaeli.
- Lubbers, N., Bolton, D. C., Mohd-Yusof, J., Marone, C., Barros, K., & Johnson, P. A. (2018). Earthquake catalog-based machine learning identification of laboratory fault states and the effects of magnitude of completeness. *Geophysical Research Letters*, 45(24), 13-269. <https://doi.org/10.1029/2018GL079712>
- Morde, V. (2019, Apr 8). *Xgboost algorithm: long may she reign!* Towards Data Science. <https://towardsdatascience.com/https-medium-com-vishalmorde-xgboost-algorithm-long-she-may-rein-edd9f99be63d>
- Moustra, M., Avraamides, M., & Christodoulou, C. (2011). Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals. *Expert systems with applications*, 38(12), 15032-15039. <https://doi.org/10.1016/j.eswa.2011.05.043>
- Oğuzlar, A. (2003). Veri ön işleme. *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, (21), 67-76.
- Özkan, Y., Yürekli, B. S., & Suner, A. (2022). Diyabet tanısının tahminlenmesinde denetimli makine öğrenme algoritmalarının performans karşılaştırması. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, 12(1), 211-226.
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., & Chica-Rivas, M. J. O. G. R. (2015). Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geology Reviews*, 71, 804-818. <http://dx.doi.org/10.1016/j.oregeorev.2015.01.001>
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, 44(18), 9276-9282. <https://doi.org/10.1002/2017GL074677>
- Shen, Z. K., Wang, M., Zeng, Y., & Wang, F. (2015). Optimal interpolation of spatially discretized geodetic data. *Bulletin of the Seismological Society of America*, 105(4), 2117-2127. <https://doi.org/10.1785/0120140247>
- Solak, H. İ. (2015). *GNSS hızları ile güneybatı anadolu'daki gerinim alanlarının zamansal değişimi*. [Yüksek lisans tezi, Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü].
- Solak, H. İ. (2020). *İzmir-Balıkesir transfer zonu ve çevresindeki güncel deformasyonların GNSS yöntemi ile incelenmesi*. [Doktora tezi, Afyon Kocatepe Üniversitesi Fen Bilimleri Enstitüsü].