



Makale / Research Paper

Çarpışma Riski Bulunan Asteroitlerin Makine Öğrenmesi ile Tespiti

Ömer Can ESKİCİOĞLU^{1a*}, Ali Hakan IŞIK^{1b}, Onur SEVLİ^{1c}

¹Mühendislik-Mimarlık Fakültesi, Burdur Mehmet Akif Ersoy Üniversitesi, Burdur, 15030, TÜRKİYE
*oceskioglu@mehmetakif.edu.tr

Received/Geliş: 25.06.2022

Accepted/Kabul: 07.09.2022

Öz: Asteroitler geçmişten günümüze kadar insanların dikkatini çekmektedir. Kadim medeniyetlerin inanç ve kültürlerinde de geniş yer almaktadır. İnsanoğlunun keşfetme ve merak duygusu bu cisimlere olan ilgisinin artmasına neden olmaktadır. Teknolojinin belirli bir seviyeye gelmesiyle asteroitlerin tespiti, teşhisi ve materyalleri net bir şekilde bulunabilmektedir. Bu cisimlerin izleyecekleri rota ve çarpışma etkileri sürekli olarak gözlem gerektirmektedir. Çalışmamızda Kaggle’da bulunan ve kaynağı NASA-JPL olan bir asteroit veri seti kullanılarak Dünya’ya çarpma ihtimali olan asteroitlerin sınıflandırılması yapılmıştır. Veri setinde 4687 asteroit verisi bulunmaktadır. Veriler üzerinde eksik verilerin doldurulması, anomali tespit etme ve normalizasyon gibi ön işleme aşamaları uygulanmıştır. Daha sonra korelasyon yardımıyla tehlikelilik durumları için verisetinden 19 adet öznelik tespit edilmiştir. Öznelikler ile Karar Ağacı, Naive Bayes, Lojistik Regresyon, Rastgele Orman, Destek Vektör Makineleri, K-En Yakın Komşu, Xgboost ve Adaboost makine öğrenmesi algoritmaları kullanılarak asteroit sınıflandırması yapılmıştır. Farklı nöron ve katman sayılarına sahip derin sinir ağı ile veriler eğitilmiş ve sınıflandırma algoritmaları ile karşılaştırılmıştır. Karşılaştırma sonucunda en yüksek başarıyı %99,80 ile AdaBoost algoritması ile sağlanmıştır. Çalıştırılan tüm sınıflandırma algoritmalarında ızgara-arama yöntemi kullanılarak hiperparametre optimizasyonu yapılmıştır. Böylelikle sürekli gözlem gerektiren ve yüksek miktardaki verilerin daha performanslı bir şekilde işlenmesini sağlayan bir yöntem önerilmiştir.

Anahtar kelimeler: Asteroit Sınıflandırması, Anomali Tespiti, Karar Ağaçları, Derin Sinir Ağları.

Machine Learning Detection of Collision-Risk Asteroids

Abstract: Asteroids have attracted people's attention from the past to the present. It has a wide place in the beliefs and cultures of ancient civilizations. The sense of discovery and curiosity of human beings causes an increase in their interest in these objects. With the technology coming to a certain level, the detection, diagnosis and materials of asteroids can be found clearly. The route and collision effects of these objects require constant observation. In our study, asteroids that are likely to hit the Earth have been classified using an asteroid data set in Kaggle and the source of which is NASA-JPL. The dataset contains 4687 asteroid data. Pre-processing steps such as filling in missing data, anomaly detection and normalization were applied on the data. Then, with the help of correlation, 19 features were determined from the dataset for dangerous situations. Asteroid classification was made by using Decision Tree with features, Naive Bayes, Logistic Regression, Random Forest, Support Vector Machines, K-Nearest Neighbor, Xgboost and Adaboost machine learning algorithms. With the deep neural network with different number of neurons and layers, the data were trained and compared with classification algorithms. As a result of the comparison, the highest performance was achieved with the AdaBoost algorithm with 99.80%. Hyperparameter optimization was performed using the grid-search method in all the classification algorithms that were run. Thus, a method that requires continuous observation and enables the processing of large amounts of data in a more efficient way has been proposed.

Keywords: Asteroid Classification, Anomaly Detection, Decision Trees, Deep Neural Networks.

Bu makaleye atıf yapmak için

Eskicioğlu Ö.C., Işık A.H., Sevli O., “Çarpışma Riski Bulunan Asteroitlerin Makine Öğrenmesi ile Tespiti” El-Cezeri Fen ve Mühendislik Dergisi, 2022, 9(4); 1431-1449.

How to cite this article

Eskicioğlu Ö.C., Işık A.H., Sevli O., “Machine Learning Detection of Collision-Risk Asteroids” El-Cezeri Journal of Science and Engineering, 2022, 9(4); 1431-1449.

ORCID ID: ^a0000-0001-5644-2957; ^b0000-0003-3561-9375; ^c0000-0002-8933-8395

1. Giriş

Yaklaşık 5 milyar yıl önce Güneş Sisteminin bulunduğu bölgede yıldız kalıntılarından oluşan bir toz bulutu vardı. Bu kalıntılardan oluşan toz bulutları zamanla çarpışıp birleşmeye başlamışlardır. Ağır elementler birleştikçe büyük kütleleri oluşturdular. Büyük kütleli cisimlerin kendi yer çekimi alanları vardı ve kütlesi yerçekimi alanından kaynaklı sürekli olarak artmaktaydı. Ağır elementler içeren ve büyük kütleli bu cisim Güneşin çekirdeğini oluşturmuştur. Güneş Sisteminin gezegenleri de kayaç, metaller ve gazlardan meydana gelmektedir [1]. Asteroitler ise Güneş Sisteminin şekillenmesinin daha sonra ve gezegen oluşturmak için birleşemeyecek olan kayaç ve metallere oluşmaktadır [2]. Asteroitler düzensiz şekillere sahiptir ve genellikle kendi eksenini etrafında dönerler.

Asteroitlerin yörüngeleri Dünyadaki gözlem evleri tarafından sıklıkla izlenmektedir. Güneş Sisteminde bilinen 1.076.651 asteroit vardır [3]. Bu asteroitlerin gezegenlere çarpma ihtimali bulunmaktadır. Jupiter'in devasa kütle çekimi her ne kadar dünya için bir kalkan görevi görse de bazı zamanlarda Dünya'nın tehlikeli olarak yakınından geçmektedir. Dünya için uzun, uzay için kısa zaman periyotlarında asteroitler Dünya'ya çarpmaktadır. Çalışmamızda mevcut bilinen asteroitlerin parametrelerine göre tehlikelilik durumları tahmin edilmektedir.

Literatürdeki çalışmalara bakıldığında asteroit teşhisi ve sınıflandırması ile ilgili kullanılan yöntemler ve teknikler incelenmiştir. Yapılan literatür araştırmasında makine öğrenmesi algoritmalarıyla asteroit sınıflandırılması ve asteroitlerle ilgili çalışmalar arasında şunlar gösterilebilir;

VISTA-VHS (VISTA Yarımküre Araştırması)' den elde edilen araştırmalar ışığında Y, J, H ve Ks filtreleri kullanarak asteroitler için taksonomik sınıflandırma yapılması hedeflenmiştir. MOVIS verilerinden yararlanarak 18265 asteroit için taksonomik sınıflandırma yapılmıştır. Olasılık tabanlı bir yöntem ve KNN (k-en yakın komşu) algoritmaları kullanılmıştır. Çalışmada kullanılan yöntemlerde genellikle benzer sonuçlar ile karşılaşıldığı ve renk hatalarının $(Y - J) \text{ err} \leq 0,118$ ve $(J - Ks) \text{ err} \leq 0,136$ verdiği rapor edilmiştir [4].

NASA'nın OSIRIS-REx görevindeki görüntüler kullanılarak Bennu asteroitindeki düz alanlar bulabilmek için makine öğrenmesi yöntemlerinden faydalanmaktadır. Rastgele Orman algoritmasıyla %90 oranında başarımla gösterilmektedir [5].

Güney Afrika, Sutherland, Kore Mikrolensing Teleskop Ağı'nın (KMTNet) ile gözlenen 39 Dünya'ya yakın asteroitin spektrofotometrisini sunmaktadır. Scikit-learn modülünü kullanarak KNN makine öğrenme algoritması ile 39 hedef asteroitten 31'ini S-, C-, X- ve D- tipi olarak sınıflandırılması yapılmıştır [6].

Asteroit Karasal Etkili Son Uyarı Sistemi (ATLAS), asteroitlere gezegeni savunma görevini büyük bir tarama yaparak gerçekleştirmektedir. 142 milyon yıldızın ışık eğrilerini analiz edip veri kümesi oluşturulmuştur. 169 sınıftan oluşan yıldız verileri çalışmada paylaşılmıştır. Doğrulanmamış adayları 15 farklı kategoriye ayırmak için 3 katmanlı ve 400-800-400 düğümlü derin sinir ağı kullanılmaktadır [7].

Asteroitleri tanımlamak için K en yakın komşu, Karar Ağaçları, Gradyan Arttırma ve Lojistik Regresyon uygulamaktadır. Yaklaşık %99 doğrulukta sonuç vermektedir. Ayrıca makine öğrenme metodolojisi kullanarak 404 yeni asteroit belirlenmiştir [8].

Düzensiz küçük cisimler etrafındaki yerçekimi ivmesini hesaplamak ve tahmin etmek için model tabanlı ve veriye dayalı bir yaklaşımı araştırmışlardır. Uzay aracı konumu ve yerçekimi ivmesi

arasındaki iliřkiyi öđrenebilen Tek Katmanlı İleri Besleme Ađları (SLFN) tasarlayıp eđitmek ve deđerlendirmek için Ekstrem Öđrenme Makinesi (ELM) teorisi kullanılmıřtır. ELM tabanlı SLFN'nin hem küresel olarak hem de asteroit yüzeyine yakın seilmiř alanlarda istenilen iliřkiyi öđrenebildiđini alıřmada gösterilmiřtir [1].

Dinamik olarak seilmiř 75 hareketsiz kuyruklu yıldız adayları ile 6 asteroitin hareketleri rapor edilmiřtir. Belirlenen nesnelerin sınıflandırmasını spektrofotometrik deđerlere dayanarak yapılmıřtır. Parametre olarak öklid uzunluđu ile KNN algoritması kullanılarak %92,2 başarımla elde edilmiřtir [9].

Gönüllü kiřiler tarafından desteklenmiř bilim projesi AstreoidZoo anlatılmaktadır. En verimli asteroit arařtırması olarak gösterilen Cataline Sky Survey(CSS) tarafından elde edilen verileri, bir makine öđrenmesi algoritması ile eksiksiz olarak tespit etmesi hedeflenmiřtir. Temel hedef, bu verilerle makine öđrenme algoritmaları için yařanılan zorlukları ařıp bir eđitim ortamı hazırlamaktır [10].

Asteroitleri tespit etmek ve incelemek için binlerce parametre nokta kaynađını incelemek gerektirmektedir. Yüksek parametrelerden oluřan elde edilen veriler için daha verimli teknikler ve yöntemler gerekir. Bu nedenden dolayı asteroit keřfi ve tespiti için makine öđrenimi tekniklerinden denetimli öđrenme uygulanmıřtır [11].

Dünya'ya yakın göktařı arařtırma sistemi olan (ATLAS) ile elde ettiđi verilerdeki optik ve elektronik kusurları ayırmak için bir sinir ađı modeli uygulanması önerilmektedir. Modelin ATLAS verilerindeki gerek asteroitlerde %99,6 oran ile başarımla gösterildiđi belirtilmektedir. Modelin ATLAS'ta konuřlandırılması Dünya'ya yakın nesne (NEO) adaylarının miktarının %90 oranında azalmasına neden olması ve daha verimli alıřmasına olanak sađlaması amaçlanmaktadır [12].

Makine öđrenmesi algoritmalarıyla keřfedilmemiř tehlikeli potansiyele sahip asteroitler (PHA) için yörünge parametreleri yardımıyla tahmin etmeyi hedeflemiřtir. Önerdiđi yaklařımda PHA da bulunan Dünya'ya yakın asteroitler (NEA) gruplarını SVM'in RBF çekirdeđi ile elde olmayan verileri yapay veri oluřturarak tahmin etmeyi amaçlamaktadır [13].

NEO kapsamında olan Bennu asteroitinde bulunan kimyasal ve fiziksel nesnelerin OSIRIS-Rex görevindeki görüntü iřleme grubu, mozaik oluřturma tekniđi ve 3'lü özellik tanıma tekniđi kullanarak tehlikeli nesnelerin haritalanması önerilmiřtir. Ayrıca CosmoQuest kullanarak internet tabanlı amatör sınıflandırma ve makine öđrenimi araçları da kullanılması gerekmektedir. alıřmanın sonucunda Bennu misyonunun başarısını sađlamak için tehlike haritaları sayesinde hızlı ve güvenli bir řekilde oluřturulması hedeflenmiřtir [14].

Büyük Asteroit Mücadelesi (AGC) olarak tanımlanan ok disiplinli iř birlikleri ile asteroit tehditlerini bulmaya ve yeniliki angajman mekanizmaları ile birlikte bir alıřma grubu oluřturulmuřtur. Asteroit Yönlendirme Misyonu (ARM) bir asteroiti yakalamak ve yeniden yönlendirmek için mevcutta bulunan robotik teknolojilerinden yararlanılarak en yüksek düzeyde fayda sađlanması hedeflenmektedir. alıřmada 2013 ve 2017 arasında AGC faaliyetlerini aıklamaktadır [15].

Nesne tespiti, makine öđrenmesi ve derin öđrenme ile görüntü sınıflandırarak asteroit, yıldız ve galaksi gibi oluřumların ESA sky bilim arřivinden aık olarak nasıl kullanılabileceđi ve yararlanılabileceđi hakkında bir alıřma hazırlanmıřtır [16].

Yapay zeka ve makine öğrenmesinin astronomi üzerindeki kullanımları hakkında bilgi vermektedir. Yıldız, asteroit, öte gezegen ve kuasarlar gibi astronomik objelerin günümüz teknolojisi ışığında yapay zeka teknikleri ile tespit ve teşhis edileceği hakkında derleme bir çalışma hazırlanmıştır [17].

İtalyan Uzay Ajansı (ASI) tarafından yönetilen LICIACube, gezegenler arası seyir yaptıktan sonra Didymos ikili asteroitine uçuş yörüngesi boyunca otonom olarak yönlendirilecektir. Bu süre boyunca asteroitten birkaç görüntü dünyaya aktaracaktır. 2021 sonunda fırlatılması beklenen LICIACube, teknolojik görev zorluğu ve otonom olarak hareket edebilmesi projeyi daha önemli bir hale getirmektedir [18].

Bus-DeMeo taksonomik sınıflandırma şemasında Temel Bileşen Analizi (PCA) ile bir makine öğrenme algoritması olan KNN'i karşılaştırmıştır. KNN veri kümesinde >%80 doğruluk oranıyla metodolojiyi doğrular, ancak mevcut yöntemleri değiştirmek için gereken eşğin(>%95) altında kalmaktadır [19].

Asteroit yüzeylerinde 0,3 ile 1,1 µm dalga boyu aralığındaki spektral özelliklere bakarak birkaç geniş gruba sınıflandırılabilirliğini önermektedir. En son geliştirilen 3 taksonomi, analiz tekniğine bağlı olarak 9,11,14 farklı sınıfa ayrılmaktadır. Çalışmada belirtilen sınıflar üzerinden anlatım yapılmaktadır [20].

3 parametrelilik taksonomik sistemi tanımlamak için yüksek kaliteli U, V, x görsel ile hem de yüksek kaliteli IRAS albedos ile 357 asteroitin sınıflandırma listesini literatüre sunmaktadır. 11 sınıf bulunmaktadır. 4 sınıf uzun zamandır bilinmektedir. 6 sınıf ise son yıllarda kabul edilmiştir. Geriye kalan son sınıf ise yeni olarak öngörülmektedir [21].

Asteroitlerin mevcut renk ve albedo verilerinin üzerine geliştirilen ve uygulanan 3 taksonomik sınıflandırma şeması oluşturulmuştur. Oluşturulan şemaların ilk 2'sinde asteroit taksonomik tablolarla detaylı şekilde açıklanmıştır [22].

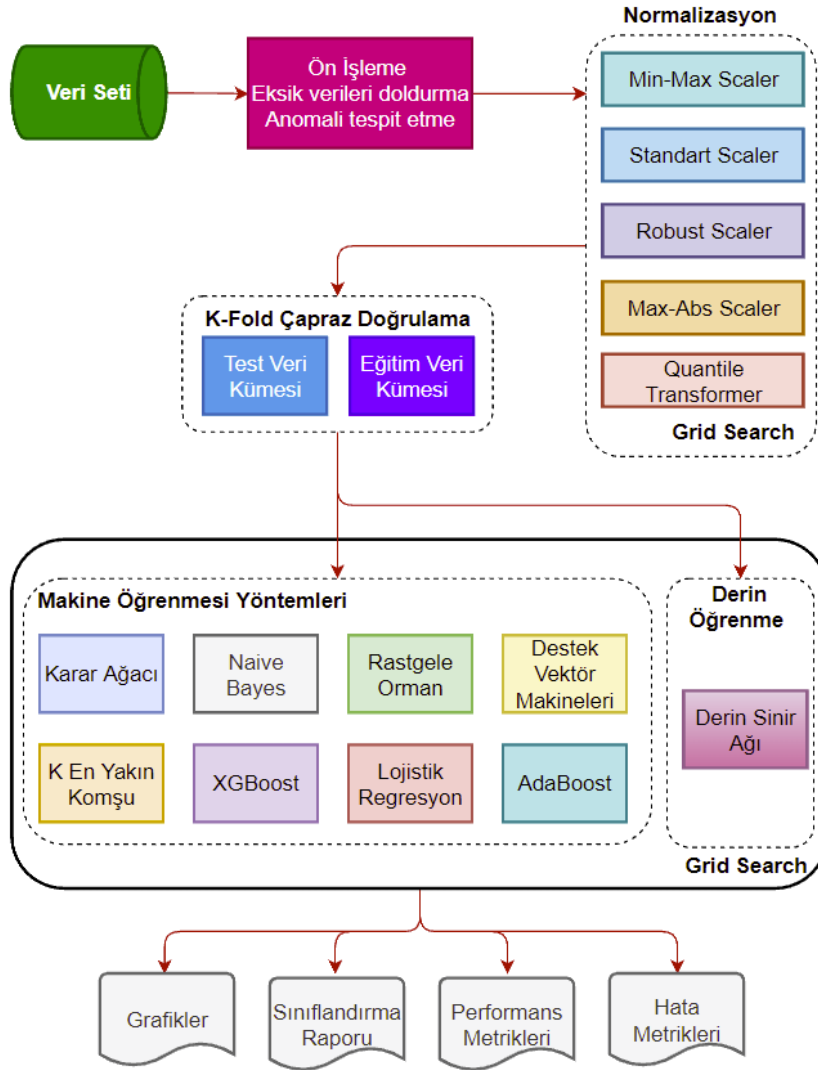
Literatürdeki çalışmalara bakıldığında makine öğrenmesinin ve derin öğrenmenin birçok astronomi çalışmasında aktif olarak kullanıldığını ve kabul edilebilir sonuçlar verdiği görülmektedir. Asteroit sınıflandırma üzerine literatürde belirli çalışmalar yapılmıştır.

Çalışmamızda, Dünya'ya potansiyel tehdit olabilecek yörüngesi ve teknik özellikleri bilinen asteroitlerin sınıflandırması yapılmıştır. Çarpışma riski yüksek olan asteroitler "tehlikeli olanlar" olarak, çarpışma riski düşük olanlar ise "tehlikeli olmayanlar" olarak sınıflandırılmıştır. Böylelikle içeriği ve özellikleri bilinen bir asteroiti, çarpışmadan çok daha önce yüksek doğrulukta sınıflandırılması ve erken tedbir alınması amaçlanmaktadır. Literatürde yürütülen çeşitli asteroit önleme mekanizmaları ile yörüngesinden saptırılması veya uzak mesafeden imha edilmesi etiketleme işleminden sonra uygulanacağı öngörülmektedir. Böylece Dünya yaşamına tehdit olan potansiyel asteroitlerin mevcut tehlikelilik durumunu ortadan kaldırılması düşünülmektedir.

2. Yöntem

Çalışmamızda Kaggle'da bulunan ve kaynağı NASA-JPL olan bir asteroit veri seti kullanılmıştır. Veri setinde 4687 asteroit verisi bulunmaktadır. Bu verilerde korelasyon yardımıyla tehlikelilik durumları için bağıl özellikler bulunup, sınıflandırma yöntemleriyle tahmin edilmesi hedeflenmiştir. Veriseti üzerinde eksik verileri doldurma, anomali tespiti ve normalizasyon yapılmıştır. Min-Max, Standart, Robust, Max-Abs ve Quantile Dönüşüm ölçekleme kullanarak ızgara-arama algoritması ile normalizasyon yapılmıştır. Uygulanan normalizasyon algoritmasının, sınıflandırmada doğruluk, kesinlik ve duyarlılık üzerindeki etkileri gözlemlenmiştir. Veri ön işleme aşamasında test ve eğitim kümesinin seçiminde K-Fold çapraz doğrulama kullanılmıştır. Veri seti K-Fold ile 3 parçaya ayrılıp

dinamik olarak eğitim ve test kümesi oluşturulmuştur. Çalışmanın genel yapısı Şekil 1’de gösterilmektedir.



Şekil 1. Çalışmanın Akış Diyagramı

2.1. Veri Seti

Veri setimizde asteroitlere ait 4687 adet veri kullanılmıştır. Tehlikelilik durumuna göre tahmin yapılacağı için tehlikeli olanlar ve tehlikeli olmayanlar olarak veri seti 2 parçaya ayrılmıştır. Bu bağlamda 3922 adet tehlikeli olmayan ve 755 adet tehlikeli olan sınıflarımız bulunmaktadır. Veri setinde toplam 40 adet özellik bulunmaktadır. Özellikler arası korelasyon bulunmasından sonra tahmin için 19 adet özellik kullanılması önerilmiştir. Bu özellikler Tablo 1 ‘de verilmektedir.

Tablo 1. Veri Setinin Özellikleri

Veri Setinin Özellikleri	Açıklama	Veri Tipi
Absolute Magnitude	Mutlak Büyüklük	float64
Est Dia in M(min)	M cinsinden Tahmini Çap (dak)	float64
Miles per hour	Saatteki mil	float64
Miss Dist.(kilometers)	Kayıp uzaklık(kilometre)	float64

Orbit Uncertainty	Yörünge Belirsizliği	int64
Minimum Orbit Intersection	Minimum Yörünge Kesişimi	float64
Jupiter Tisserand Invariant	Jüpiter Değişmez Tisserandı	float64
Epoch Osculation	Tur Salınımı	float64
Eccentricity	Eksantriklik	float64
Semi Major Axis	Yarı büyük eksen	float64
Inclination	Eğim / Eğilim	float64
Asc Node Longitude	Asc. Düğüm Boylamı	float64
Orbital Period	Yörünge dönemi	float64
Perihelion Distance	Günberi Mesafesi	float64
Perihelion Arg	Günberi Argümanı	float64
Aphelion Dist	Aphelion Mesafesi	float64
Perihelion Time	Günberi Saati	float64
Mean Anomaly	Ortalama Anomali	float64
Mean Motion	Ortalama Hareket	float64

2.2. Makine Öğrenmesi Modelleri

Çalışmamızda NASA-JPL asteroit veri setini kullanarak makine öğrenmesi yöntemleriyle sınıflandırma yapılmıştır. Sınıflandırma aşamasında Karar Ağacı, Naive Bayes, Lojistik Regresyon, Rastgele Orman, Destek Vektör Makineleri, K-En Yakın Komşu, Xgboost ve Adaboost makine öğrenmesi sınıflandırıcıları kullanılmıştır.

- Karar ağacı sınıflandırıcısı

Karar ağaçlarında temel fikir, herhangi birçok amaçlı yaklaşımlarda karmaşık bir karar mekanizmalarında problemin nihai çözümü için birkaç basit kararın birliği sayesinde istenilen çözümü bulmayı amaçlayan yaklaşımdır [23]. Karar Ağacı sınıflandırıcısı, karmaşık işlemleri basite indirgeyerek problemi çözmeye odaklanmıştır.

- Naive bayes sınıflandırıcısı

Naive Bayes verilen özelliklerin belirli sınıftan bağımsız olduğunu varsayarak öğrenmeyi büyük ölçüde basitleştirmektedir [24]. Temelde yöntem 18. Yüzyılda yaşamış olan İngiliz matematikçi Thomas Bayes'in ortaya attığı teoremine dayanmaktadır (1). Bu modelde her özellik birbirinden bağımsız olarak kabul edilmektedir.

$$P(A|B) = \frac{(P(B|A) * P(A))}{P(B)} \quad (1)$$

$P(A|B)$ = B'nin doğru olduğu bilindiğinde A'nın olma olasılığı

$P(B|A)$ = A'nın doğru olduğu bilindiğinde B'nin olma olasılığı

$P(A)$ = A'nın olma olasılığı

$P(B)$ = B'nin olma olasılığı

- Lojistik regresyon sınıflandırıcısı

Fazla sayıda değişkenin varlığında olasılık oranını bulabilmek için lojistik regresyon kullanılmaktadır. Yanıt değişkeninin iki terimli olması haricinde çoklu doğrusal regresyona benzemektedir. Her bir değişkenin, gözlemlenen olayın olasılık oranı üzerindeki etkisidir. Tüm

deđişkenlerin ilişkisini analiz etmektedir [25]. Bir sonucu belirleyen ve bir ya da daha fazla bağımsız deđişken bulunan veri kümelerini analiz etmek için kullanılmaktadır. İstatistiksel bir yöntemdir.

- Rastgele orman sınıflandırıcısı

Birden fazla sayıda karar ağaçlarının birleşmesinden oluşan modele denir. Ormandaki her ağacın, tüm ağaçlar için aynı dağılıma sahip ve bağımsız olarak örneklenen rastgele bir vektörün değerlerine dayanan ağaç tahmin edicilerinin birleşimlerine denir [26]. Karar ağaçlarında sıkça karşılaşılan ezberleme sorunun birden fazla ağaç üzerinde işleyerek çözmektedir.

- Destek vektör sınıflandırıcısı

1990'larda Vapnik ve ekibi tarafından geliştirilmiştir. İki gruplu sınıflandırma problemleri için bir gözetimli öğrenme makinesidir [27]. Temeli istatistiksel öğrenme teorisinde dayanmaktadır ve mevcut sınıfların en iyi şekilde ayrılmasını sağlayabilmek için karar sınırlarının yerini belirlemeye çalışmaktadır [28].

- K-En yakın komşu sınıflandırıcısı

1951 yılında Evelyn Fix ve Joseph Hodges tarafından geliştirilmiştir. Regresyon ve sınıflandırma için kullanılmaktadır. En basit makine öğrenmesi algoritması olarak gösterilebilir. Ancak sınıflandırmada etkili ve güçlü bir yöntemdir [29]. Herhangi bir eğitim aşaması bulunmamaktadır. Girilen K değerine en yakın noktalara göre sınıf tespitini yapılmasını sağlamaktadır. Yakın noktalar hesaplanırken çeşitli uzunluk ölçütleri kullanılabilir (Öklid, Manhattan vb.). Parametre seçimleri eldeki verilerin dağılıma göre seçilmelidir.

- XGBoost sınıflandırıcısı

2016 yılında Tianqi Chen ve Carlos Guestrin tarafından geliştirilmiştir [30]. Temel olarak Gradient Boosting algoritmasının optimize edilmiş halidir. Daha yüksek performans ile çalışmaktadır. Xgboost algoritması aşırı öğrenmenin önüne geçmesi, yüksek tahmin gücü ve boş verileri yönetebilme kapasitesiyle diğer makine öğrenmesi algoritmaları arasında öne çıkmaktadır. Ayrıca diğer makine öğrenme algoritmalarından 10 kat daha hızlı çalıştığı gözlemlenmiştir [31].

- Adaboost sınıflandırıcısı

Adaboost algoritması 1996 yılında Yoav Freund ve Robert E. Schapire tarafından geliştirilmiştir [32]. Algoritmanın temel çalışma prensibi, fazla sayıda zayıf öğreniciyi bir araya getirerek güçlü bir öğrenici oluşturmaktır. Genellikle arttırma metodu hedefi, tahmin edicileri kümülatif olarak eğitmektir [33]. İlk başarılı arttırma algoritmalarındandır ve Gödel ödülünü kazanmıştır.

2.3. Derin Sinir Ağları

İnsan beynindeki nöronlardan esinlenerek oluşturulmuş karmaşık işlemleri yapılabilmesini olanak sağlayan bir yapay nöronlardan meydana gelen bir ağıdır. Derin sinir ağları 3 bölüme ayrılmaktadır. Giriş katmanı, gizli katman ve çıkış katmanından oluşmaktadır [34]. İnsanların öğrenme sürecini matematiksel bir şekilde yorumlanması sonucunda ortaya çıkmıştır. Finans, görüntü işleme, bilgisayarda görü, biyoloji, enerji, otomotiv, havacılık, üretim, savunma sanayi, sensör teknolojileri ve telefon gibi çeşitli sektörler üzerinde hayatımızda yoğun olarak kullanılmaktadır [35].

3. Bulgular ve Tartışma

Çalışmamızda 8 makine öğrenme algoritmasını ve derin sinir ağını kullanarak asteroitlerin tehlikelilik sınıflandırılması yapılmıştır. Tüm sınıflandırıcılarda ve oluşturulan derin sinir ağı modelinde maksimum performans için ızgara-arama algoritması kullanılmıştır. En yüksek başarımları veren hiperparametreler eğitilmiştir. Eğitim kapsamında Adaboost sınıflandırıcısı %99.80 doğruluk ile bu probleme en iyi çözüm sağlayan algoritma olduğu kanısına varmıştır. Eğitilen tüm modeller iyi bir performans değerlerine sahiptir. Asteroitlerin tehlikelilik durumuna göre sınıflandırmasında hepsi kullanılabilir.

3.1. Makine Öğrenmesi Algoritmalarının Performans Sonuçları

Çalışmamızda toplamda 8 makine öğrenmesi algoritması kullanılmıştır. Bunlar; Xgboost, AdaBoost, K-En Yakın Komşu, Destek Vektör, Rastgele Orman, Lojistik Regresyon, Karar Ağaçları ve Naive Bayes'tir. Her bir algoritmada ızgara-arama yöntemi kullanılmıştır.

Karar Ağaçları sınıflandırıcısında criterion, max_depth, min_samples_split ve ölçekleme fonksiyonları ızgara arama algoritmasına verilerek Tablo 2'deki sonuçlar elde edilmiştir. Bu hiperparametreler ile en yüksek %99,74 doğruluk oranı elde edilmiştir.

Tablo 2. Karar Ağacı ile Izgara-Arama Algoritmasının Sonuçları

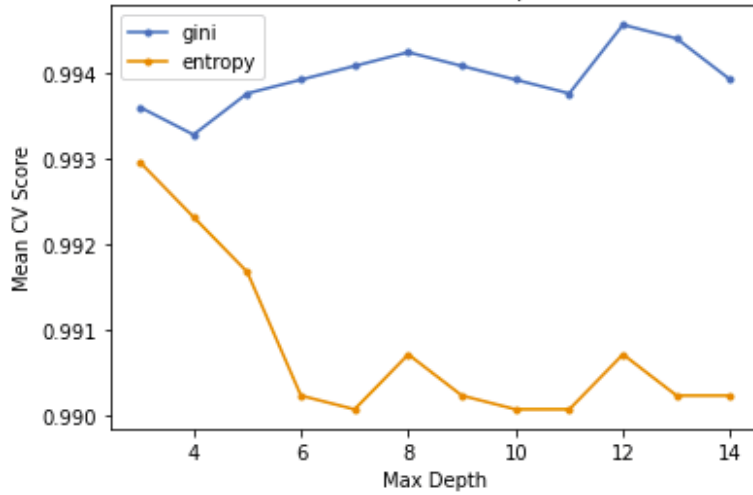
criterion	Hiperparametreler			Doğruluk Oranı	Kesinlik	Duyarlılık	F1 Skoru	MSE
	Max depth	min_samples_split	Ölçekleme					
entropy	11	3	MinMaxScaler	0,9942	0,9962	0,9970	0,9966	0,0057
gini	7	2	MinMaxScaler	0,9961	0,9977	0,9977	0,9977	0,0038
gini	8	2	MinMaxScaler	0,9974	0,9977	0,9992	0,9985	0,0025
entropy	8	2	StandardScaler	0,9955	0,9977	0,9970	0,9974	0,0044
gini	7	2	StandardScaler	0,9961	0,9977	0,9977	0,9977	0,0038
gini	8	2	StandardScaler	0,9974	0,9977	0,9992	0,9985	0,0025
entropy	8	2	RobustScaler	0,9955	0,9977	0,9970	0,9974	0,0044
gini	7	2	RobustScaler	0,9961	0,9977	0,9977	0,9977	0,0038
gini	8	2	RobustScaler	0,9974	0,9977	0,9992	0,9985	0,0025
entropy	8	2	MaxAbsScaler	0,9955	0,9977	0,9970	0,9974	0,0044
gini	7	2	MaxAbsScaler	0,9962	0,9977	0,9977	0,9977	0,0038
gini	8	2	MaxAbsScaler	0,9974	0,9977	0,9992	0,9985	0,0025
entropy	8	2	Quantile Transformer	0,9955	0,9977	0,9970	0,9974	0,0044
gini	7	2	Quantile Transformer	0,9955	0,9977	0,9970	0,9974	0,0044
gini	8	2	Quantile Transformer	0,9967	0,9970	0,9992	0,9981	0,0032
			MinMaxScaler	0,9959	0,9974	0,9974	0,9974	0,0040
			StandardScaler	0,9963	0,9978	0,9978	0,9976	0,0036
			RobustScaler	0,9963	0,9978	0,9978	0,9976	0,0036
			MaxAbsScaler	0,9963	0,9978	0,9978	0,9976	0,0036
			Quantile Transformer	0,9959	0,9974	0,9974	0,9974	0,0040

Tablo 2'de verilen performans sonuçları kapsamında en yüksek değerler Tablo 3'de verilmiştir.

Tablo 3. Karar Ağacı ile En Performanslı Sonuçlar

Hiperparametreler				Ölçekleme	Dođruluk Oranı
critierion	max_depth	min_samples_split			
gini	8	2		MinMaxScaler	0,9974
gini	8	2		StandardScaler	0,9974
gini	8	2		RobustScaler	0,9974
gini	8	2		MaxAbsScaler	0,9974

Karar Ağacı için en iyi hiperparametreler: gini, 8 ve 2 olarak bulunmaktadır. Şekil 2’de Karar Ağacı algoritmasının kriterine ve derinliğine göre karşılaştırma grafiđi görünmektedir. Grafiđe göre gini criterion hiperparametresi entropy parametresine göre daha uygun bir seçim olduđu anlaşılmaktadır. Entropy kriter parametresinde maksimum derinliđin düşük olması daha iyi bir sonuç verdiđi görülmüştür. Grafikteki en yüksek ortalama skor, gini kriterinde maksimum derinliđi 8 veya 12 parametrelerinde elde edilmiştir.



Şekil 2. Karar Ağacı Algoritmasının Performans Deđerleri

Naive Bayes sınıflandırıcısında var_smoothing ve ölçekleme fonksiyonları ızgara arama algoritmasına verilerek Tablo 4’deki sonuçlar elde edilmiştir. Bu hiperparametreler ile en yüksek %94,90 dođruluk oranı elde edilmiştir.

Tablo 4. Naive Bayes ile Izgara-Arama Algoritmasının Sonuçları

HiperParametreler		Dođruluk Oranı	Kesinlik	Duyarlılık	F1 Skoru	MSE
var_smoothing	Ölçekleme					
0,001873	Quantile Transformer	0,9136	0,9448	0,9527	0,9487	0.0863
0,005336	Quantile Transformer	0,9148	0,9511	0,9459	0,9485	0,0851
0,000351	Quantile Transformer	0,9225	0,9508	0,9573	0,9541	0,0774
	MinMaxScaler	0,9490	0,9702	0,9686	0,9694	0,0474
	StandardScaler	0,9488	0,9698	0,9683	0,9691	0,0482
	RobustScaler	0,9484	0,9696	0,9676	0,9687	0,0496
	MaxAbsScaler	0,9490	0,9702	0,9686	0,9694	0,0474
	Quantile Transformer	0,9426	0,9667	0,9637	0,9652	0,0546

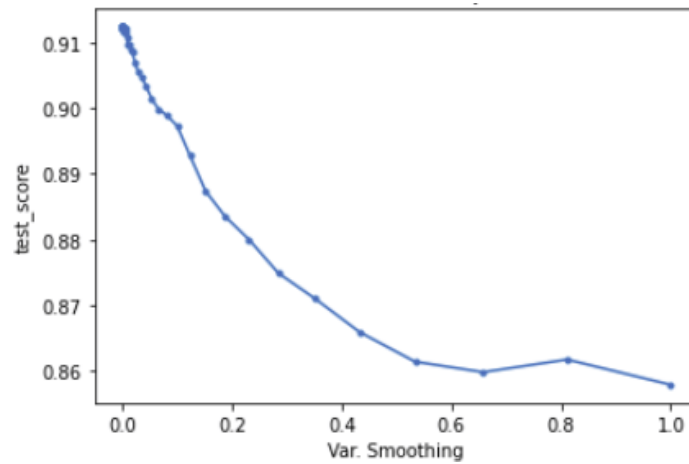
Tablo 4’de verilen performans sonuçları kapsamında en yüksek deđerler Tablo 5’de verilmiştir.

Tablo 5. Naive Bayes ile En Performanslı Sonuçlar

HiperParametreler		Doğruluk
var_smoothing	Ölçekleme	Oranı
	MinMaxScaler	0,9490
	MaxAbsScaler	0,9490

Naive Bayes için en iyi hiperparametreler: MinMaxScaler ve MaxAbsScaler kullanılarak bulunmaktadır. Şekil 3’de Naive Bayes algoritmasının başarımına ve var_smoothing parametresine göre karşılaştırma grafiği görülmektedir.

Grafikte var_smoothing parametresinin yükseldiği takdirde başarımın düştüğü gözlemlenmektedir. Bu parametrenin 0’a en yakın durumunda maksimum performans elde edileceği sonucuna varılabilir.



Şekil 3. Naive Bayes Algoritmasının Performans Değerleri

Tablo 6. Lojistik Regresyon ile Izgara-Arama Algoritmasının Sonuçları

HiperParametreler			Doğruluk	Kesinlik	Duyarlılık	F1	MSE
C	penalty	Ölçekleme	Oranı			Skoru	
1000,0	12	MinMaxScaler	0,9552	0,9766	0,9707	0,9736	0,0447
1000,0	12	MinMaxScaler	0,9571	0,9744	0,9736	0,9740	0,0428
1000,0	12	MinMaxScaler	0,9545	0,9773	0,9692	0,9733	0,0454
10,0	12	StandardScaler	0,9571	0,9773	0,9722	0,9747	0,0428
100,0	12	StandardScaler	0,9603	0,9767	0,9752	0,9760	0,0396
10,0	12	StandardScaler	0,9545	0,9766	0,9699	0,9732	0,0454
1000,0	12	RobustScaler	0,9366	0,9652	0,9601	0,9627	0,0633
1000,0	12	RobustScaler	0,9430	0,9627	0,9680	0,9654	0,0569
10,0	12	RobustScaler	0,9411	0,9697	0,9610	0,9654	0,0588
1000,0	12	MaxAbsScaler	0,9545	0,9758	0,9707	0,9732	0,0454
1000,0	12	MaxAbsScaler	0,9558	0,9736	0,9728	0,9732	0,0441
1000,0	12	MaxAbsScaler	0,9545	0,9773	0,9692	0,9733	0,0454
1,0	12	QuantileTransformer	0,9225	0,9856	0,9274	0,9556	0,0774
10,0	12	QuantileTransformer	0,9244	0,9689	0,9412	0,9549	0,0755
1,0	12	QuantileTransformer	0,9238	0,9849	0,9293	0,9563	0,0761
		MinMaxScaler	0,9556	0,9726	0,9722	0,9716	0,0443
		StandardScaler	0,9573	0,9785	0,9764	0,9780	0,0396
		RobustScaler	0,9402	0,9691	0,9580	0,9644	0,0602
		MaxAbsScaler	0,9549	0,9786	0,9704	0,9756	0,0442
		QuantileTransformer	0,9236	0,9862	0,9288	0,9580	0,0726

Lojistik Regresyon sınıflandırıcısında C, penalty ve ölçekleme fonksiyonları ızgara arama algoritmasına verilerek Tablo 6'daki sonuçlar elde edilmiştir. Bu hiperparametreler ile en yüksek %96,03 doğruluk oranı elde edilmiştir.

Tablo 6'da verilen performans sonuçları kapsamında Lojistik Regresyon sınıflandırıcısı için en iyi hiperparametre: 100,0, 12 ve StandardScaler olarak bulunmaktadır.

Rastgele Orman sınıflandırıcısında criterion, max_depth, max_features, n_estimators ve ölçekleme fonksiyonları ızgara arama algoritmasına verilerek Tablo 7'deki sonuçlar elde edilmiştir. Bu hiperparametreler ile en yüksek %99,67 doğruluk oranı elde edilmiştir.

Tablo 7. Rastgele Orman ile Izgara-Arama Algoritmasının Sonuçları

criterion	HiperParametreler				Doğruluk Oranı	Kesinlik	Duyarlılık	F1 Skoru	MSE
	max_depth	max_features	n_estimators	Ölçekleme					
gini	4	auto	400	MinMaxScaler	0,9923	0,9992	0,9917	0,9955	0,0076
gini	6	auto	400	MinMaxScaler	0,9967	0,9992	0,9969	0,9981	0,0032
gini	6	auto	200	MinMaxScaler	0,9967	0,9992	0,9970	0,9981	0,0032
gini	4	auto	400	StandardScaler	0,9923	0,9992	0,9917	0,9955	0,0076
gini	6	auto	400	StandardScaler	0,9967	0,9992	0,9970	0,9981	0,0032
gini	6	auto	200	StandardScaler	0,9967	0,9992	0,9970	0,9981	0,0032
gini	4	auto	400	RobustScaler	0,9923	0,9992	0,9917	0,9955	0,0076
gini	6	auto	400	RobustScaler	0,9967	0,9992	0,9970	0,9981	0,0032
gini	6	auto	200	RobustScaler	0,9967	0,9992	0,9970	0,9981	0,0032
gini	4	auto	400	MaxAbsScaler	0,9923	0,9992	0,9917	0,9955	0,0076
gini	6	auto	400	MaxAbsScaler	0,9967	0,9992	0,9970	0,9981	0,0032
gini	6	auto	200	MaxAbsScaler	0,9967	0,9992	0,9970	0,9981	0,0032
gini	6	auto	600	Quantile Transformer	0,9955	0,9992	0,9955	0,9974	0,0044
gini	6	auto	600	Quantile Transformer	0,9967	0,9992	0,9970	0,9981	0,0032
gini	6	auto	200	Quantile Transformer	0,9961	0,9985	0,9970	0,9977	0,0038
				MinMaxScaler	0,9953	0,9985	0,9966	0,9972	0,0046
				StandardScaler	0,9953	0,9985	0,9966	0,9972	0,0046
				RobustScaler	0,9953	0,9985	0,9966	0,9972	0,0046
				MaxAbsScaler	0,9953	0,9985	0,9966	0,9972	0,0046
				Quantile Transformer	0,9961	0,9985	0,9970	0,9977	0,0038

Tablo 7'de verilen performans sonuçları kapsamında Rastgele Orman sınıflandırıcısı için en iyi hiperparametreler Tablo 8'de verilmektedir.

Destek Vektör sınıflandırıcısında C, gamma, kernel ve ölçekleme fonksiyonları ızgara arama algoritmasına verilerek Tablo 9'daki sonuçlar elde edilmiştir. Bu hiperparametreler ile en yüksek %98,20 doğruluk oranı elde edilmiştir.

Tablo 8. Rastgele Orman Sınıflayıcısında Elde Edilen En Performanslı Sonuçları

criterion	HiperParametreler			Ölçekleme	Doğruluk Oranı
	max_depth	max_features	n_estimators		
gini	6	auto	400	MinMaxScaler	0,9967
gini	6	auto	200	MinMaxScaler	0,9967
gini	6	auto	400	StandardScaler	0,9967
gini	6	auto	200	StandardScaler	0,9967
gini	6	auto	400	RobustScaler	0,9967
gini	6	auto	200	RobustScaler	0,9967
gini	6	auto	400	MaxAbsScaler	0,9967
gini	6	auto	200	MaxAbsScaler	0,9967
gini	6	auto	600	QuantileTransformer	0,9967

Tablo 9. Destek Vektör ile Izgara-Arama Algoritmasının Sonuçları

C	Hiperparametreler			Doğruluk Oranı	Kesinlik	Duyarlılık	F1 Skoru	MSE
	Gamma	kernel	Ölçekleme					
1000	0,1	rbf	MinMaxScaler	0,9571	0,9750	0,9743	0,9747	0,0428
1000	0,1	rbf	MinMaxScaler	0,9647	0,9814	0,9761	0,9787	0,0352
1000	0,1	rbf	MinMaxScaler	0,9558	0,9758	0,9721	0,9740	0,0441
1000	0,01	rbf	StandardScaler	0,9667	0,9796	0,9811	0,9803	0,0332
1000	0,01	rbf	StandardScaler	0,9718	0,9845	0,9814	0,9829	0,0281
1000	0,01	rbf	StandardScaler	0,9718	0,9845	0,9814	0,9829	0,0281
1000	0,01	rbf	RobustScaler	0,9635	0,9788	0,9781	0,9784	0,0364
100	0,1	rbf	RobustScaler	0,9628	0,9852	0,9702	0,9777	0,0371
100	0,1	rbf	RobustScaler	0,9615	0,9803	0,9744	0,9774	0,0384
1000	0,1	rbf	MaxAbsScaler	0,9526	0,9720	0,9720	0,9720	0,0473
1000	0,1	rbf	MaxAbsScaler	0,9603	0,9806	0,9715	0,9760	0,0396
1000	0,1	rbf	MaxAbsScaler	0,9545	0,9743	0,9721	0,9732	0,0454
1000	0,1	rbf	Quantile Transformer	0,9814	0,9871	0,9909	0,9890	0,0185
1000	0,1	rbf	Quantile Transformer	0,9788	0,9891	0,9853	0,9872	0,0211
1000	0,1	rbf	Quantile Transformer	0,9820	0,9894	0,9894	0,9894	0,0179
			MinMaxScaler	0,9592	0,9764	0,9750	0,9751	0,0399
			StandardScaler	0,9701	0,9838	0,9806	0,9821	0,0296
			RobustScaler	0,9626	0,9850	0,9701	0,9776	0,0369
			MaxAbsScaler	0,9558	0,9758	0,9721	0,9740	0,0441
			Quantile Transformer	0,9807	0,9868	0,9905	0,9886	0,0197

Tablo 9’da verilen performans sonuçları kapsamında Destek Vektör sınıflandırıcısı için en iyi hiperparametre: 1000, 0,1, rbf ve Quantile Transformer olarak bulunmaktadır.

K-En Yakın Komşu sınıflandırıcısında n_neighbors, p ve ölçekleme fonksiyonları ızgara arama algoritmasına verilerek Tablo 10’daki sonuçlar elde edilmiştir. Bu hiperparametreler ile en yüksek %92,45 doğruluk oranı elde edilmiştir.

Tablo 10. K-En Yakın Komşu ile Izgara-Arama Algoritmasının Sonuçları

HiperParametreler			Doğruluk Oranı	Kesinlik	Duyarlılık	F1 Skoru	MSE
n_neighbors	p	Ölçekleme					
5	1	Quantile Transformer	0,9245	0,9826	0,9319	0,9566	0,0754
5	1	Quantile Transformer	0,9161	0,9790	0,9238	0,9506	0,0838
5	2	Quantile Transformer	0,9154	0,9782	0,9238	0,9492	0,0845
		MinMaxScaler	0,9056	0,9242	0,9233	0,9286	0,0906
		StandardScaler	0,9068	0,9282	0,9304	0,9324	0,0895
		RobustScaler	0,9067	0,9282	0,9304	0,9326	0,0896
		MaxAbsScaler	0,9048	0,9236	0,9228	0,9267	0,0921
		Quantile Transformer	0,9076	0,9402	0,9388	0,9412	0,0882

Tablo 10’da verilen performans sonuçları kapsamında K-En Yakın Komşu sınıflandırıcısı için en iyi hiperparametreler: 5, 1 ve Quantile Transformer olarak bulunmaktadır.

Adaboost sınıflandırıcısında base_estimator_criterion, base_estimator_splitter, n_estimators ve ölçekleme fonksiyonları ızgara arama algoritmasına verilerek Tablo 11’deki sonuçlar elde edilmiştir. Bu hiperparametreler ile en yüksek %99,80 doğruluk oranı elde edilmiştir.

Tablo 11. Adaboost ile Izgara-Arama Algoritmasının Sonuçları

HiperParametreler				Doğruluk Oranı	Kesinlik	Duyarlılık	F1 Skoru	MSE
base_estimator_criterion	base_estimator_splitter	n_estimators	Ölçekleme					
entropy	best	8	MinMaxScaler	0,9929	0,9970	0,9947	0,9958	0,0071
entropy	best	4	MinMaxScaler	0,9942	0,9977	0,9955	0,9966	0,0057
entropy	best	8	MinMaxScaler	0,9980	0,9992	0,9984	0,9989	0,0019
entropy	best	8	StandardScaler	0,9929	0,9970	0,9947	0,9958	0,0071
entropy	best	4	StandardScaler	0,9942	0,9977	0,9955	0,9966	0,0057
entropy	best	8	StandardScaler	0,9980	0,9992	0,9984	0,9989	0,0019
entropy	best	8	RobustScaler	0,9929	0,9970	0,9947	0,9958	0,0071
entropy	best	4	RobustScaler	0,9942	0,9977	0,9955	0,9966	0,0057
entropy	best	8	RobustScaler	0,9980	0,9992	0,9984	0,9989	0,0019
entropy	best	8	MaxAbsScaler	0,9929	0,9970	0,9947	0,9958	0,0071
entropy	best	4	MaxAbsScaler	0,9942	0,9977	0,9955	0,9966	0,0057
entropy	best	8	MaxAbsScaler	0,9980	0,9992	0,9984	0,9989	0,0019
entropy	best	8	Quantile Transformer	0,9929	0,9970	0,9947	0,9958	0,0071
entropy	best	4	Quantile Transformer	0,9942	0,9977	0,9955	0,9966	0,0057
entropy	best	8	Quantile Transformer	0,9980	0,9992	0,9984	0,9989	0,0019
			MinMaxScaler	0,9950	0,9977	0,9962	0,9970	0,0051
			StandardScaler	0,9950	0,9977	0,9962	0,9970	0,0051
			RobustScaler	0,9950	0,9977	0,9962	0,9970	0,0051
			MaxAbsScaler	0,9950	0,9977	0,9962	0,9970	0,0051
			Quantile Transformer	0,9950	0,9977	0,9962	0,9970	0,0051

Tablo 11’de verilen performans sonuçları kapsamında Adaboost sınıflandırıcısı için en iyi hiperparametreler Tablo 12’de verilmektedir.

Tablo 12. Adaboost Sınıflandırıcısının Izgara-Arama Sonuçları

HiperParametreler				Doğruluk Oranı
base_estimator _criterion	base_estimator_ _splitter	n_estima tors	Ölçekleme	
entropy	best	8	MinMaxScaler	0,9980
entropy	best	8	StandardScaler	0,9980
entropy	best	8	RobustScaler	0,9980
entropy	best	8	MaxAbsScaler	0,9980
entropy	best	8	QuantileTransformer	0,9980

XGBoost sınıflandırıcısında max_depth, n_estimators, learning_rate ve ölçekleme fonksiyonları ızgara arama algoritmasına verilerek Tablo 13’deki sonuçlar elde edilmiştir. Bu hiperparametreler ile en yüksek %99,74 doğruluk oranı elde edilmiştir.

Tablo 13. XGBoost ile Izgara-Arama Algoritmasının Sonuçları

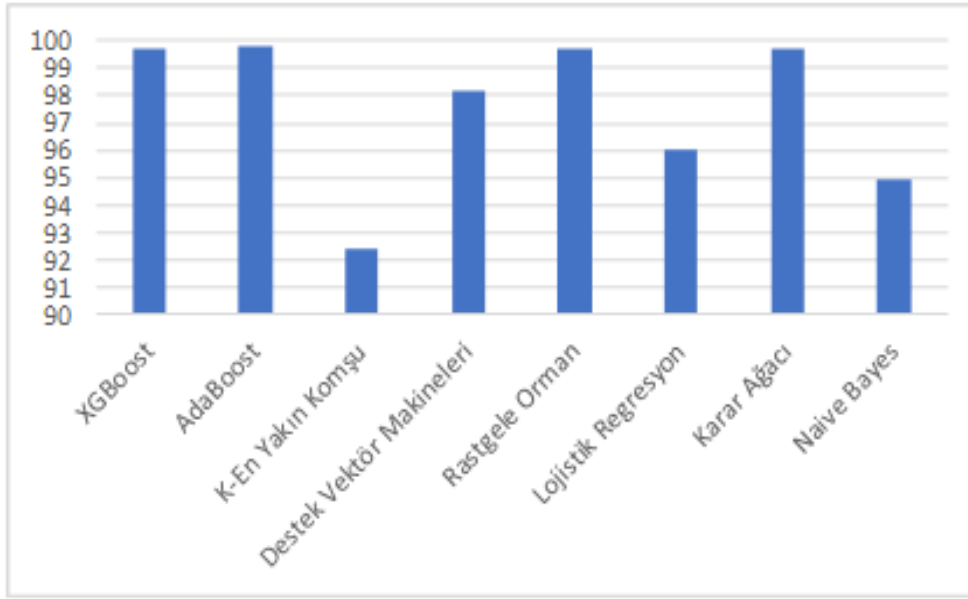
HiperParametreler				Doğruluk Oranı	Kesinlik	Duyarlılık	F1 Skoru	MSE
max_ depth	n_estima tors	learning_ rate	Ölçekleme					
2	60	0,1	MinMaxScaler	0,9926	0,9970	0,9938	0,9942	0,0074
4	120	0,01	MinMaxScaler	0,9945	0,9977	0,9943	0,9950	0,0056
8	180	0,05	MinMaxScaler	0,9974	0,9977	0,9992	0,9985	0,0025
2	60	0,1	StandardScaler	0,9926	0,9970	0,9938	0,9942	0,0074
4	120	0,01	StandardScaler	0,9945	0,9977	0,9943	0,9950	0,0056
8	180	0,05	StandardScaler	0,9974	0,9977	0,9992	0,9985	0,0025
2	60	0,1	RobustScaler	0,9926	0,9970	0,9938	0,9942	0,0074
4	120	0,01	RobustScaler	0,9945	0,9977	0,9943	0,9950	0,0056
8	180	0,05	RobustScaler	0,9974	0,9977	0,9992	0,9985	0,0025
2	60	0,1	MaxAbsScaler	0,9926	0,9970	0,9938	0,9942	0,0074
4	120	0,01	MaxAbsScaler	0,9945	0,9977	0,9943	0,9950	0,0056
8	180	0,05	MaxAbsScaler	0,9974	0,9977	0,9992	0,9985	0,0025
2	60	0,1	Quantile Transformer	0,9926	0,9970	0,9938	0,9942	0,0074
4	120	0,01	Quantile Transformer	0,9945	0,9977	0,9943	0,9950	0,0056
8	180	0,05	Quantile Transformer	0,9974	0,9977	0,9992	0,9985	0,0025
			MinMaxScaler	0,9961	0,9977	0,9977	0,9977	0,0038
			StandardScaler	0,9961	0,9977	0,9977	0,9977	0,0038
			RobustScaler	0,9961	0,9977	0,9977	0,9977	0,0038
			MaxAbsScaler	0,9961	0,9977	0,9977	0,9977	0,0038
			Quantile Transformer	0,9961	0,9977	0,9977	0,9977	0,0038

Tablo 13’de verilen performans sonuçları kapsamında Xgboost sınıflandırıcısı için en uygun hiperparametreler Tablo 14’de verilmektedir.

Tablo 14. XGBoost ile En Performanslı Sonular

HiperParametreler				Dođruluk Oranı
max_depth	n_estimators	learning_rate	Ölekleme	
8	180	0,05	MinMaxScaler	0,9974
8	180	0,05	StandardScaler	0,9974
8	180	0,05	RobustScaler	0,9974
8	180	0,05	MaxAbsScaler	0,9974
8	180	0,05	QuantileTransformer	0,9974

alıřmamızdaki makine ğrenmesi algoritmalarının en yüksek performanslarının karşılaştırılması Őekil 4’de verilmektedir.

**Őekil 4.** Makine ğrenmesi Algoritmalarının Performans Karşılaştırması

Veri setine uygulanan makine ğrenmesi algoritmalarına ızgara-arama yöntemi ile hiperparametre optimizasyonu yapılmıştır. Yapılan optimizasyon neticesinde en yüksek başarıyı %99,80 dođruluk ile Adaboost sınıflayıcısı vermektedir. En başarılı sonucun karışıklık matrisi Tablo 15’de gösterilmektedir.

Tablo 15. En Yüksek Dođruluk Oranına Sahip Modelin Karışıklık Matrisi

	Dođru Pozitif	Dođru Negatif
Tahmini Pozitif	1321	1
Tahmini Negatif	2	238

Karışıklık matrisine gre 1562 test verisinde 1. sınıftan 1323, 2. sınıftan 239 veri bulunmaktadır. Yüksek başarılı sınıflayıcımız 1. sınıftan 1321 ve 2. sınıftan 238 veriyi dođru olarak sınıflamaktadır. 1562 veri iinden 3 test verisini yanlış sınıflandırmaktadır.

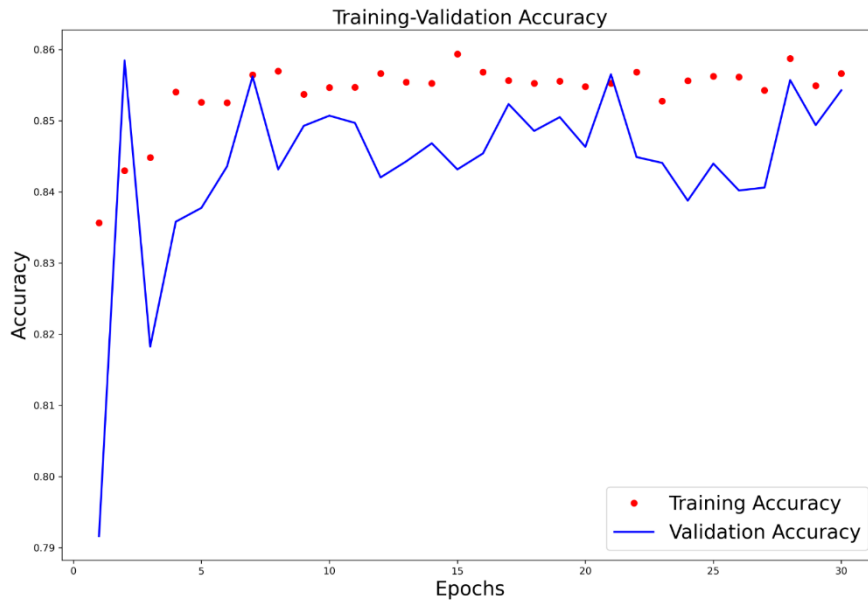
3.2. Derin Sinir Ađının Performans Sonuları

alıřmamızda derin sinir ađı üzerinde ızgara-arama yapılmıştır. ızgara-arama sonuları Tablo 16’da grlmektedir.

Tablo 16. Derin Sinir Ağı ile En Performanslı Sonuçlar

HiperParametreler				Doğruluk Oranı
Epochs	Batch Size	Loss Function	Optimizer	
30	8	Binary_crossentropy	Adam	0,8660
30	16	Binary_crossentropy	Nadam	0,8248
50	8	Binary_crossentropy	SGD	0,8026
50	8	Binary_crossentropy	Adagrad	0,8578
100	16	Binary_crossentropy	Adadelta	0,8468

Derin sinir ağı üzerinde ızgara-arama sonucunda 30 epoch, 8 batch size, binary_crossentropy kayıp fonksiyonu ve Adam optimizasyon algoritmasını kullanarak en yüksek %86,60 doğruluk oranı elde edilmiştir. Şekil 5'te eğitim-validasyon doğruluk grafiği gösterilmektedir.

**Şekil 5.** Derin Sinir Ağının Eğitim Sonucu

Çalışmamızda oluşturulan derin sinir ağı modelinde 4 gizli katmana sahip 50 nöronlu bir ağ kullanılmıştır. Tablo 17'de bu ağın yapısı açıklanmaktadır.

Tablo 17. Derin Sinir Ağının Yapısı

Katmanlar	Nöron Sayısı	Aktivasyon Fonksiyonu	Çekirdek Başlatıcı
Giriş Katmanı	19	relu	uniform
1. Gizli Katman	16	relu	uniform
2. Gizli Katman	8	relu	uniform
3. Gizli Katman	4	relu	uniform
4. Gizli Katman	2	relu	uniform
Çıktı Katmanı	1	sigmoid	uniform

5. Sonuçlar

Çalışmamızda NASA-JPL Asteroit veriseti kullanılarak asteroitlerin tehlikelilik durumuna göre sınıflandırma yapılmıştır. Xgboost, AdaBoost, K-En Yakın Komşu, Destek Vektör, Rastgele

Orman, Lojistik Regresyon, Karar Ağaçları ve Naive Bayes sınıflandırıcıları kullanılmıştır. Çalışmamızda Xgboost algoritması ile %99,74, AdaBoost ile %99,80, K-En Yakın Komşu ile %92,45, Destek Vektör ile %98,20, Rastgele Orman ile %99,67, Lojistik Regresyon ile %96,03, Karar Ağaçları ile %99,74 ve Naive Bayes ile %94,90 doğruluk oranı elde edilmiştir. Ayrıca çalışmamızda Derin Sinir Ağı ile de %86,60 başarımla sağlanmıştır. En yüksek başarımla, %99,80 ile AdaBoost algoritmasıyla elde edilmiştir. Asteroit sınıflandırılmasında en verimli çözümü Robust ölçeklemesiyle entropy kriteriyle 8 tahmin edici kullanarak ve 'best' temel tahmin edicisi parametreleri uygulandığında en yüksek başarımla elde edilmektedir. Çalışmadaki tüm sınıflandırma algoritmaları ve sinir ağları üzerinde ızgara-arama işlemi yapılmış ve en iyi parametreler elde edilmiştir.

Yapılan çalışma ışığında verisetindeki 19 öznelik sayesinde bir sınıflama işlemi yapılmaktadır. Öznelikler, korelasyon matrisi ile belirlenmiştir. Farklı verisetlerinden asteroit numarası bulunarak asteroitin iç yapısı öğrenilebilmektedir. Tehlikelilik durumuna göre çarpma olasılığı hatta çarpışma anında Dünyaya verdiği hasar hesaplanabilmektedir. Asteroitin büyüklüğü, yörüngesi, hızı ve hammaddesi çarpışma için bilinmesi gereken kriterlerdir.

Çalışmamız ile literatürdeki diğer çalışmaların karşılaştırılması Tablo 18'de verilmektedir. Çalışmamızda birçok metot ile çalışılmış olup literatürde bulunan diğer çalışmalardan daha iyi bir sonuç elde edilmiştir.

Tablo 18. Literatürdeki makine öğrenme algoritmalarının karşılaştırılması

Yazarlar	Metotlar	Konular	Sonuçlar
Popescu M. ve arkadaşları [5]	Rastgele Orman S.	Düz alanların tespiti	Doğruluk : %90
Smirnov E. A, Markov AB. [8]	KNN, Karar Ağaçları, Lojistik Regresyon, Gradyan Arttırma	Asteroitleri tanımlamak	Doğruluk : %99
Mommert M. ve arkadaşları [9]	KNN	Kuyruklu yıldız ve asteroit sınıflandırması	Doğruluk : %92,2
McIntyre KJ. [19]	KNN	Bus-DeMeo taksonomik sınıflandırma	Doğruluk : >%80
Rabeendran AC, Denneau L. [12]	ANN	Optik ve elektronik kusurların tespiti	Doğruluk : %99,6
Önerilen Çalışmamız	DNN, KNN, Karar A., Rastgele Orman, SVC, Lojistik Regresyon NB, Adaboost, Xgboost	Asteroit Sınıflandırması	(Adaboost) Doğruluk : %99,8

Yazarın Katkıları

Bu çalışmada yazarların eşit ve ortak katkısı bulunmaktadır.

Çıkar Çatışması

Yazarlar, çıkar çatışması olmadığını beyan eder.

Kaynaklar

- [1]. Furfaro, R., Barocco, R., Linares, R., Topputo, F., Reddy, V., Simo, J., Le Corre, L., "Modeling irregular small bodies gravity field via extreme learning machines and Bayesian

- optimization", *Advances in Space Research*, 2021, 67(1): 617-638.
- [2]. Kabaş, A., Bulut, İ., Doğru, S. S., Akin, T., "Bazı Ana Kuşak Asteroidlerin Işık Eğrileri Ve Işık Eğrilerinden Belirlenen Parametreler", XVI. Ulusal Astronomi Kongresi ve V. Ulusal Öğrenci Astronomi Kongresi, Çanakkale, 1188-1199, 2008.
- [3]. Chapman, C. R., Ramlose, T., "Solar system exploration" NASA STI/Recon Technical Report N, 1989, 89: 25944.
- [4]. Popescu, M., Licandro, J., Carvano, J. M., Stoicescu, R., de León, J., Morate, D., Boacă, I. L., Cristescu, C. P., "Taxonomic classification of asteroids based on MOVIS near-infrared colors", *Astronomy & Astrophysics*, 2018, 617(A12).
- [5]. Cambioni, S., Bennett, C. A., Walsh, K. J., DellaGiustina, D. N., Golish, D. R., Becker, K. J., Lauretta, D., S., "A search for smooth terrains on asteroid (101955) Bennu using machine learning", In EPSC-DPS Joint Meeting 2019, 2019.
- [6]. Erasmus, N., Mommert, M., Trilling, D. E., Sickafoose, A. A., Van Gend, C., Hora, J. L., "Characterization of near-earth asteroids using KMTNET-SAAO", *The Astronomical Journal*, 2017, 154(4): 162.
- [7]. Heinze, A. N., Tonry, J. L., Denneau, L., Flewelling, H., Stalder, B., Rest, A., Smith, K. W., Smartt, S. J., Weiland, H., "A first catalog of variable stars measured by the Asteroid Terrestrial-impact Last Alert System (ATLAS)", *The Astronomical Journal*, 2018, 156(5): 241.
- [8]. Smirnov, E. A., Markov, A. B., "Identification of asteroids trapped inside three-body mean motion resonances: a machine-learning approach", *Monthly Notices of the Royal Astronomical Society*, 2017, 469(2): 2024-2031.
- [9]. Mommert, M., Trilling, D. E., Hora, J. L., Lejoly, C., Gustafsson, A., Knight, M., Moskovitz, N., Smith, H. A., "Systematic characterization of and search for activity in potentially active asteroids", *The Planetary Science Journal*, 2020, 1(1): 10.
- [10]. Beasley, M., Lewicki, C. A., Smith, A., Lintott, C., Christensen, E., "AsteroidZoo: A new zooniverse project to detect asteroids and improve asteroid detection algorithms", In AGU Fall Meeting Abstracts, 2013.
- [11]. Nugent, C. R., Dailey, J., Cutri, R. M., Masci, F. J., Mainzer, A. K., "Machine learning and next-generation asteroid surveys", In AAS/Division for Planetary Sciences Meeting Abstracts, 2017, 49: 103-03.
- [12]. Rabeendran, A. C., Denneau, L. "A two-stage deep learning detection classifier for the atlas asteroid survey", *Publications of the Astronomical Society of the Pacific*, 2021, 133(1021): 034501.
- [13]. Pasko, V., "Prediction of Orbital Parameters for Undiscovered Potentially Hazardous Asteroids Using Machine Learning", In Stardust Final Conference, Springer Netherlands, 52: 45-65.
- [14]. Johnson, C. A., DellaGiustina, D. N. "Hazards on Hazards, Ensuring Spacecraft Safety While Sampling Asteroid Surface Materials", In AGU Fall Meeting Abstracts, 2016, 2016: NH13A-1748.
- [15]. Gustetic, J. L., Friedensen, V., Kessler, J. L., Jackson, S., Parr, J., "NASA's Asteroid Grand Challenge: Strategy, Results, and Lessons Learned", *Space Policy*, 2018, 44: 1-13.
- [16]. Lieu, M., Baines, D., Giordano, F., Merin, B., Arviset, C., Altieri, B., Conversi, B., Carry, B., "Deep Learning of Astronomical Features with Big Data", *Astronomical Data Analysis Software and Systems XXVII*, 2019, 523: 49.
- [17]. Fluke, C. J., Jacobs, C., "Surveying the reach and maturity of machine learning and artificial intelligence in astronomy", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2020, 10(2): e1349.
- [18]. Dotto, E., Della Corte, V., Amoroso, M., Bertini, I., Brucato, J. R., Capannolo, A., ... & Fretz, K., "LICIACube-the Light Italian Cubesat for Imaging of Asteroids in support of the NASA DART mission towards asteroid (65803) Didymos", *Planetary and Space Science*, 2021, 199:

105185.

- [19]. McIntyre, K. J., "Applying Machine Learning To Asteroid Classification Utilizing Spectroscopically Derived Spectrophotometry", Master Thesis, The University of North Dakota, 2019.
- [20]. Tholen, D. J., Barucci, M. A., "Asteroid taxonomy", *Asteroids II*, 1989, 298-315.
- [21]. Tedesco, E. F., Williams, J. G., Matson, D. L., Veeder, G. J., Gradie, J. C., Lebofsky, L. A., "Three-parameter asteroid taxonomy classifications", *Asteroids II*, 1989, 1151-1161.
- [22]. Tholen, D. J., "Asteroid taxonomic classifications", In *Asteroids II*, 1989, 1139–1150.
- [23]. Safavian, S. R., Landgrebe, D., "A survey of decision tree classifier methodology", *IEEE transactions on systems, man, and cybernetics*, 1991, 21(3): 660-674.
- [24]. Rish, I., "An empirical study of the naive Bayes classifier", In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 2001, 3(22): 41-46.
- [25]. Sperandei, S., "Understanding logistic regression analysis"; *Biochemia medica*, 2014 24(1): 12-18.
- [26]. Breiman, L., "Random forests", *Machine learning*, 2001, 45(1): 5-32.
- [27]. Cortes, C., Vapnik, V., "Support-vector networks", *Machine learning*, 1995, 20(3): 273-297.
- [28]. Pal, M., "Random forest classifier for remote sensing classification", *International journal of remote sensing*, 2005, 26(1): 217-222.
- [29]. Guo, G., Wang, H., Bell, D., Bi, Y., Greer, K., "KNN model-based approach in classification", In *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*, Berlin, 2003, 986-996.
- [30]. Chen, T., Guestrin, C., "Xgboost: A scalable tree boosting system", In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, 785-794.
- [31]. Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., "Xgboost: extreme gradient boosting", *R package version 0.4-2*, 2015, 1(4): 1-4.
- [32]. Freund, Y., Schapire, R. E., "Experiments with a new boosting algorithm", *icml*, 1996, 96: 148-156.
- [33]. Schapire, R. E., "Explaining adaboost", In *Empirical inference*, Berlin, 2013, 37-52.
- [34]. Öztemel, E., "Yapay sinir ađlari", *PapatyaYayincilik*, İstanbul, (2003).
- [35]. Şen, Z., "Yapay sinir ađlari", *Su Vakfi*, İstanbul, (2004).