



## Ülkelerin Gelişmişlik Düzeylerinin Karar Ağacı ve Rastgele Orman Yöntemleriyle Tahmin Edilmesi

### Predicting Countries' Development Levels Using the Decision Tree and Random Forest Methods

Batuhan Özkan\* , Coşkun Parim\*\* , Erhan Çene\*\*\*

#### Öz

Ülkelerin gelişmişlik düzeyleri ile ekonomik açıdan kalkınma düzeyleri arasında çok yakın bir ilişki söz konusudur. Ülkeler, çeşitli ölçütlere göre incelenerek, gelişmişlik düzeylerine göre az gelişmişten çok gelişmişe doğru farklı gruplarda değerlendirilebilirler. Ülkelerin gelişmişlik düzeylerinin belirlenmesinde, genellikle sosyo-ekonomik faktörler belirleyici rol oynamaktadır. Gelişmişlik düzeyi her ne kadar sosyo-ekonomik değişkenler yardımıyla belirlense de, ülkelerin sınıflandırılması farklı organizasyonlar (Birleşmiş Milletler, Uluslararası Para Fonu vb.) tarafından farklı yöntemlerle yapılabilmektedir. Bu durum bir ülkenin gelişmişlik düzeyinin yönetime ve organizasyona göre farklı kategoride yer almasına sebep olmaktadır. Bu çalışmanın amacı, 193 ülke için gelişmişlik düzeyini tahmin eden bir makine öğrenmesi modeli geliştirmektir. Gelişmişlik düzeyi, "Yüksek Gelir", "Üst Orta Gelir", "Alt Orta Gelir" ve "Düşük Gelir" kategorilerinden oluşmaktadır. Ülkelerin gelişmişlik seviyesini etkileyen 26 değişken ise, Dünya Gelişmişlik İndeksi (World Development Indicators - WDI) veri tabanından elde edilmiştir. İlk olarak özellik seçimi olarak gelişmişlik düzeyini etkileyen en önemli değişkenlerin belirlenmesinde, rastgele orman metodu yardımıyla değişken önemi kullanılmıştır. Önemli bulunan bağımsız değişkenler yardımıyla, karar ağaçları ve rastgele orman algoritmaları kullanılarak gelişmişlik düzeyleri sınıflandırılmıştır. Rastgele orman algoritmasıyla oluşturulan modelin ülkelerin gelişmişliklerini %70 oranında doğru sınıflandırdığı belirlenmiştir. Ayrıca, bulgular Ergen Doğurganlık Hızı, Toplam Doğurganlık Oranı ve Tarım, Orman ve Balıkçılık'ın GSYİH (Gayri Safi Yurtiçi Hasıla) daki payının ülkelerin gelişmişliklerini etkileyen en önemli değişkenler olduğunu göstermektedir.

#### Anahtar Kelimeler

Gelişmişlik Seviyesi, Karar Ağacı, Rastgele Orman, Doğurganlık Hızı, Makine Öğrenmesi

#### Abstract

A very close relationship exists between countries' development levels and economic level. Countries can be examined according to various criteria and evaluated under different groups based on their level of development, from underdeveloped to highly developed. Socioeconomic factors generally play a decisive role in determining countries' levels of development. Although the level of development is determined with the help of socioeconomic variables, different organizations (e.g., United Nations [UN], International Monetary Fund [IMF]) may make country classifications with different methods. This situation causes a country's development level to occur in different categories based on the method used and the organization that performed it. The aim of this study is to propose a machine learning model

\* **Sorumlu Yazar:** Batuhan Özkan (Arş. Gör.), Yıldız Teknik Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, İstanbul, Türkiye.  
E-Posta: baozkan@yildiz.edu.tr ORCID: 0000-0002-1773-9317

\*\* Coşkun Parim (Arş. Gör. Dr.), Yıldız Teknik Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, İstanbul, Türkiye.  
E-Posta: cparim@yildiz.edu.tr ORCID: 0000-0002-6412-1325

\*\*\* Erhan Çene (Dr. Öğr. Üyesi), Yıldız Teknik Üniversitesi, Fen Edebiyat Fakültesi, İstatistik Bölümü, İstanbul, Türkiye.  
E-Posta: ecene@yildiz.edu.tr ORCID: 0000-0001-5336-6004

**Atf:** Ozkan, B., Parim, C., & Cene, E. (2023). Ülkelerin gelişmişlik düzeylerinin karar ağacı ve rastgele orman yöntemleriyle tahmin edilmesi. *EKOIST Journal of Econometrics and Statistics*, 38, 87-104. <https://doi.org/10.26650/ekoist.2023.38.1172190>



that predicts the development level for 193 countries. Development level consists of the categories of high income, upper middle income, lower middle income, and low income. The 26 variables that affect countries' development levels were obtained from the World Development Indicators (WDI) database. Firstly, random forest based variable importance was used to determine the variables which have the most important effects on countries' development levels. Afterwards, countries' development levels were classified using decision trees and random forest algorithms with the most important variables selected through variable importance. The model composed with the random forest algorithm was determined to have correctly classified countries' development levels at an accuracy of 70%. In addition, the findings show the variables of adolescent fertility rate, total fertility rate, and the share of agriculture, forestry, and fisheries in gross domestic product (GDP) to be the most important variables affecting countries' development levels.

**Keywords**

Development Level, Decision Tree, Random Forest, Fertility Rate, Machine Learning

***Extended Summary***

Although countries' development levels are often associated with economic indicators, many different factors are found to affect their level of development. The methods used to determine countries' development levels differ by organization (e.g., United Nations, World Bank, International Monetary Fund), and thus countries may show different levels of development, with the literature showing statistical methods such as machine learning, artificial neural networks, multiple discriminant analysis, and regression analysis to have been used to predict countries' development levels.

Machine learning is the art of making future predictions from data using statistical algorithms. Machine learning methods train a model using part of the data as training data and then tests the performance of the proposed model using the remaining data (i.e., test data). In this way, the accuracy of a produced model is tested on test data the model has never seen before. Studies are found to have used machine learning and econometric methods to classify countries' development levels, just as has been done in many areas.

This study developed a machine learning model that predicts the development levels for 193 countries using the World Development Indicator's 2019 dataset. The dependent variable of the model has been classified as high income (\$12,535 or more a year per capita), upper middle income (from \$4,046-\$12,535), lower middle income (\$1,035-\$4,046), and low income (Less than \$1,035). The groups in this classification were determined according to the World Bank's Atlas Method, which they use to estimate the size of economies in terms of gross national income a year per capita in US dollars. The World Bank uses the Atlas conversion factor instead of simple exchange rates when calculating gross national income (GNI—formerly GNP) in US dollars for certain operational and analytical purposes. The purpose of the Atlas conversion factor is to reduce the effect of exchange rate fluctuations caused by inflation when comparing national incomes between countries. The Atlas conversion factor for any given year is the average of a country's exchange rate for that year and the exchange rates for the previous two years, adjusted for the difference between the

inflation rate in the country and international inflation. The independent variables are made up of the 26 variables obtained from the World Development Indicators (WDI) database that can be associated with countries' development levels. These variables are related to health, industry, economy, agriculture, and technology. The study uses the decision tree and random forest machine learning methods. The results this study obtains will reveal the characteristics of countries that are at different development levels, as well as which variables affect them the most.

The data set used in the machine learning algorithms has been divided into 80% training and 20% test data. The decision tree and random forest algorithms were then applied to the data set. When considering the accuracy rates of both methods, the decision tree is seen to have a 65% accuracy rate, while the random forest has 70%. This shows the countries' development levels to have been classified correctly at a rate of 70%. When examining the correct classification rates for each income group, all the countries in the high income group are seen to have been correctly classified, while the accuracy of the classifications in the low, lower middle, and high middle groups can be said to range between approximately 65%-97%. This may be due to the effects changes have on countries in the lower income groups. In addition, the study has included values for sensitivity, specificity, positive and negative predictive values, prevalence, detection rate, detection prevalence, and balanced accuracy in order to show the classification of the categories of developmental levels.

When examining the results, the decision tree algorithm revealed three variables to be present that affect countries' development levels namely agriculture, forestry and fisheries (AFF), total fertility rate (TFR) and adolescent fertility rate (AFR). Countries with low AFF ( $<4.4$ ) and high AFR ( $\geq 37$ ) result a upper middle income country whereas countries with low AFF ( $<4.4$ ) but also low AFR ( $<37$ ) yield a high income country. When the AFF has moderate or high values ( $\geq 4.4$ ) the structure for predicting the countries development level is a bit more complex which requires considering different value ranges for the TFR. When moderate or high AFF ( $\geq 4.4$ ) combined with high TFR ( $\geq 4$ ) produce a low income country. On the other hand, lower middle income countries appear in two situations. 1) High AFF ( $\geq 12$ ) and low or moderate TFR ( $<4$ ) and 2) Moderate AFF ( $4.4 < \text{AFF} < 12$ ) and moderate TFR ( $2.8 < \text{TFR} < 4$ ). The last branch of the decision tree produces upper middle income countries with moderate AFF ( $4.4 < \text{AFF} < 12$ ) and low TFR ( $<2.8$ ).

## Giriş

Ülkelerin gelişmişlik düzeyleri sıklıkla ekonomik göstergelerle ilişkilendirilse de, gelişmişlik düzeyini etkileyen pek çok farklı etmen de bulunmaktadır. Teknoloji, nüfus, eğitim (Bloom ve Canning, 2000) ve sağlık (Guisan ve Aguayo, 2007) ülkelerin gelişmişlik düzeyi üzerinde etkili olduğu düşünülen diğer etmenlerdir (Demirci ve Karaatlı, 2019).

Ülkelerin gelişmişlik düzeyi ile en çok ilişkilendirilen değişken Gayri Safi Yurt İçi Hasıla'dır (GSYİH). Genellikle GSYİH ile ülkelerin gelişmişliği arasında pozitif bir ilişki olduğu düşünülmektedir. Buna karşın GSYİH' nın tek başına göz önüne alındığı durumlar ülkelerin hatalı sınıflandırılmasına da yol açabilmektedir (Öztürk, 2007; Erden Özsoy ve Tosunoğlu, 2017). Ülkelerin gelişmişlik düzeyi incelenirken diğer kategorilerdeki etmenlerin de göz önüne alınması gereklidir.

Makine öğrenimi (ML), geleneksel programlama yöntemlerini kullanarak programlanması zor olan karmaşık problemlerin çözümlerini otomatikleştirmek için algoritmaları ve teknikleri inceleyen bir bilgisayar bilimi alanıdır (Rebala, Ravi ve Churiwala, 2019). Geleneksel programlama yönteminde ilk adım, program için ayrıntılı bir tasarım, yani sorunu çözmek için sabit bir dizi adım veya kural oluşturmaktır. İkinci adım ise, detaylı tasarımın bilgisayar dilinde bir program olarak uygulanmasıdır. Bu yaklaşım, yapılacak iş çok net olduğunda dahi birçok gerçek dünya problemi için zorlayıcı olabilir. Makine öğrenimi algoritmaları, bu zor sorunların çoğunu daha genel bir yaklaşım uygulayarak çözebilir. Bu algoritmalar, açık ve ayrıntılı bir tasarım kullanmak yerine, ayrıntılı tasarımı bir dizi etiketli veriden öğrenirler. Bir makine öğrenimi algoritmasının amacı, veri kümesinde olmayan veri noktalarının etiketlerini doğru bir şekilde tahmin edebilmesi için etiketli bir veri kümesinden bir model veya bir dizi kural öğrenmektir. Makine öğrenimi algoritmaları, önce veri kümesini işlemeye dayalı bir model oluşturarak ve ardından bu modeli çalıştırarak yeni bir veri noktasının etiketini tahmin ederek sorunları çözer (Rebala, Ravi ve Churiwala, 2019). Bu yaklaşım denetimli makine öğrenimi olarak bilinir. Bu çalışmada denetimli makine öğrenmesi algoritmalarından olan karar ağacı ve rastgele orman yöntemleri kullanılarak ülkelerin gelişmişlik düzeyi tahmin edilecektir. Pek çok alanda olduğu gibi, ülkelerin gelişmişlik düzeyinin sınıflandırılmasında makine öğrenmesi ve ekonometrik yöntemler kullanan çalışmalar mevcuttur.

Ahmad ve Saleem (2012), Birleşmiş Milletler ve Dünya Bankası'ndan 186 ülke için topladıkları 2010 yılı verileri ile ülkelerin gelişmişlik düzeyini tahmin eden bir model geliştirmişlerdir. Yazarlar bu çalışmalarında veriyi eğitim ve test verisine bölerek yapay sinir ağları kullanmışlar ve test verilerinde %95 lik bir başarı elde etmişlerdir. Yaptıkları önem analizi, gelişmişlik üzerinde en etkili değişkenlerin kişi başına sağlık harcaması ve beklenen ortalama eğitim süresi olarak bulunmuştur.

Gelişmiş ve gelişmekte olan ülkelerin sağlık ve eğitim alanlarına yoğunlaşmalarının ülkelerin gelişmişlik düzeyini yukarı çekeceği sonucuna varmışlardır.

Ozturk, S.G. (2007), 77 ülke için Birleşmiş Milletler'den elde ettiği 2005 yılı İnsani Gelişmişlik İndeks verilerini çoklu diskriminant analizi ile uygulayarak ülkelerin gelişmişlik seviyesini tahmin etmiştir. Çalışma, sağlık, eğitim ve kadınların politikaya katılımı gibi ekonomik göstergeler dışında kalan göstergeleri kullanmıştır. Çalışmada veri eğitim ve test verisine ayrılmamıştır. Modelin başarısı %94.8 olarak bulunmuştur. Çalışma sonucunda sağlıktaki gelişmişlik seviyesi, gelişmiş ve gelişmemiş ülkeler arasındaki farklılığı oluştururken, gelişmekte olan ülkeleri belirlemede en faydalı olan göstergeler kadınların politikaya katılımı, ekonomik büyüme ve ticari gelişimdir.

Upreti (2015), gelişmekte olan 76 ülkenin ekonomik büyümesi üzerinde etkili olan etmenleri 1995, 2000, 2005 ve 2010 yılları için regresyon analizi ile incelemiştir. Sonuç olarak, yüksek ihracat, doğal kaynak bolluğu, yüksek yaşam beklentisi ve yüksek yatırım oranı ekonomik büyüme üzerinde olumlu etkide bulunduğu görülmüştür. Çalışma doğrudan yabancı sermayenin farklı zaman dilimlerinde farklı sonuçlar verdiğini ve genel kanının aksine büyük ekonomilerin her zaman yavaş büyümediği sonucuna varmıştır. Yazarlar bunu gelişmekte olan pek çok ülkenin başlangıçta benzer kişi başına düşen GSYİH'a sahip olmasıyla açıklamıştır.

Demirci ve Karaatlı (2019), 104 ülke için Uluslararası Para Fonu ve Dünya Bankası'ndan topladıkları 2016 yılı verileri ile ülkelerin gelişmişlik düzeylerini tahmin etmişlerdir. Çalışmada 17 farklı makine öğrenmesi yöntemiyle model kurulmuş ve en başarılı modelin %90'ın üzerinde başarıya sahip olduğu görülmüştür. Bu çalışmada verinin eğitim ve test verisine bölünüp bölünmediğiyle ilgili bir bilgiye yer verilmemiştir. Değişken önemi analizi yapılmadığından hangi değişkenlerin ülkelerin sınıflandırılması üzerinde etkili olduğu da belirtilmemiştir. Çalışmada herhangi bir ekonometrik çıkarım yapılmamıştır.

Jemna (2015), Romanya'daki doğurganlık oranı ve ekonomik büyüme arasındaki nedensellik ilişkisini incelemiştir. Çalışmada durağanlık testleri, vektör otoregresyon (VAR) modeli, nedensellik analizi ve etki-tepki analizi kullanılmıştır. Hem doğurganlık oranı hem de ekonomik büyüme durağan bulunmuş, her iki VAR modeli anlamlı çıkmış ve her iki değişkenin uzun dönemde birbirini üzerinde etkili olduğu sonucuna varmıştır. Çalışma doğurganlık oranının ekonomik büyümeden olumlu yönde etkilendiğini belirtmiştir. Benzer başka bir çalışmada, Lacalle-Calderon, Perez-Trujillo, & Neira (2017), 1970-2010 zaman aralığında 151 ülkeye ait doğurganlık ve GSYİH arasındaki ilişki araştırılmıştır. Bu amaçla, hem doğrusal regresyon hem de panel kartil regresyonu ve koşullu panel kartil regresyonu uygulanmıştır. Çalışma doğurganlık ve ekonomik büyüme arasında J-şeklinde bir yapı olduğunu ortaya koymuştur. Çalışma doğurganlık oranının fazla olduğu yerlerde, doğurganlık

oranını azaltmak için daha fazla GSYİH'ı ihtiyaç duyulduğu sonucuna varmıştır. Kadın başına doğum sayısının fazla olduğu ülkelerde, düşük doğurganlığa sahip olan ülkelere göre negatif ve pozitif eğim katsayılarının daha yüksek olduğu görülmüştür.

Kubar (2016), toplam 53 düşük ve orta gelir grubundaki ülkelere ait 1995-2010 dönemindeki ekonomik büyüme, tarım sektörü katma değeri, sabit sermaye tüketimi, tüketici fiyat endeksi, mal ve hizmet ihracatı, nihai tüketim harcaması, mal ve hizmet ithalatı ve sanayi sektörü katma değeri verilerini OECD ve WDI veri tabanlarından elde etmiştir. Bu değişkenlerin ekonomik büyüme üzerindeki etkisini yedi basit panel regresyon modeli, birim kök testi, panel eşbütünleşme ve panel nedensellik analizleriyle incelemiştir. Çalışmanın sonucunda, hem düşük hem orta gelir grubundaki ülkeler için ekonomik büyüme tarım sektörü katma değeri ve sabit sermaye tüketiminden olumsuz olarak etkilenirken, mal ve hizmet ihracatı, nihai tüketim harcaması, mal ve hizmet ithalatı ve sanayi sektörü katma değerinde olumlu olarak etkilenmektedir. Tüketici fiyat endeksi ise düşük gelir grubundaki ülkeleri olumlu, orta gelir grubundaki ülkeleri olumsuz etkilemektedir.

Stec, Filip, Grzebyk, ve Pierscieniak (2014), 28 Avrupa Birliği Üyesi ülkenin 2005-2010 yılı arasında verisini topladığı 27 ölçüt ile bir sosyo-ekonomik gelişmişlik indeksi geliştirmiştir. Gelişmişlik indeksinde yer alan ölçütler demografik potansiyel ve çalışan pazarı, ekonomik potansiyel, sosyal altyapı ve teknik altyapı ana başlıklarında yer almaktadır. Yüksek, orta-yüksek, orta-düşük ve düşük isimli dört gelişmişlik seviyesinden oluşan bu indekse göre, Lüksemburg, Finlandiya, İsveç ve Danimarka en yüksek grupta yer alırken, Slovakya, Bulgaristan, Romanya ve Polonya en düşük grupta yer almaktadır. Çalışma söz konusu dönemde AB ülkelerinin ortalama düzeyde bir gelişmişlik artışı gösterdiğini, 2010 krizinin de etkisiyle çoğu AB ülkesinin görece düşük gelişmişlik göstergelerine sahip olduğunu vurgulamıştır.

Bulut, Babacan ve Ertekin (2021), dört farklı insani gelişmişlik düzeyinden 16 ülkeye ait 2005-2017 yılı verilerini kullanarak ekonomik büyümeye etki eden etmenleri panel veri analizi ile incelemiştir. Bu amaçla, yatay kesit bağımlılığı, homojenite analizi, birim kök testleri ve eşbütünleşme testleri kullanılmıştır. Çalışmada nüfusun, kamu harcamalarının, enflasyonun, ticaret hacminin ve işsizliğin gelişme düzeyinden bağımsız olarak ekonomik büyümeyi pozitif etkilediği ifade edilmiştir.

Demiray Erol (2013), Türkiye ve AB üyesi 28 ülkenin sosyo-ekonomik seviyesini temel bileşenler analizi ile karşılaştırmıştır. Bu amaçla, yaşam beklentisi, sağlık, göç, ekonomik göstergeler, istihdam ve suç göstergelerinden oluşan 19 değişkeni 5 temel bileşene indirgemiş ve bu 5 bileşenden tek bir skor elde ederek ülkeleri sıralamıştır. 5 temel bileşen toplam değişkenliğin %82 sini açıklamıştır. Oluşturulan sıralamada Lüksemburg, İsveç ve Danimarka ilk üç sırada, Letonya, Romanya ve Türkiye ise son üç sırada yer almıştır. Türkiye'nin sosyal gelişmişlik seviyesini artırabilmesi için eğitim, sağlık harcamalarını arttırması ve gelir dağılımında adaleti sağlamaya

yardımcı etkin genişletici maliye politikası ve para politikası uygulaması gerektiğini belirtmiştir.

Benzer bir çalışmada, Koşar Taş ve Örk Özel (2017), Türkiye ve AB üyesi 28 ülkenin sosyo-ekonomik seviyesini faktör analizi kullanarak karşılaştırmıştır. Bu amaçla, 28 değişkeni araştırma ve eğitim, gelir ve refah, sağlık, işsizlik, yoksulluk ve nüfus isimli altı faktöre indirgemiş ve ülkeleri her bir faktör skoruna göre sıralamıştır. Bu çalışmada 6 faktör toplam değişimin yaklaşık %82'sini açıklamıştır. Oluşturulan gelişmişlik faktörüne göre Almanya, Fransa ve Birleşik Krallık ilk üç sırada yer alırken, Kıbrıs, Lüksemburg ve Malta son üç sırada yer almıştır. Türkiye bu çalışmada kendisine yedinci sırada yer bulmuştur.

Bir diğer çalışmada, Parim, Özkan ve Çene (2019) 54 ülkenin son 20 yıldaki ekonomik değişimlerini ortaya koymak amacıyla sosyo-ekonomik faktörler yardımıyla kümeleme analizi uygulayarak ülkeleri gelişmişlik düzeylerine göre sınıflandırmışlardır. Bu amaçla doğrudan yabancı sermaye, GSYİH, döviz kuru, enflasyon, işsizlik, ticarete açıklık, nüfus, işçi verimliliği, politik hak ve sivil özgürlük indeksi gibi değişkenler kullanılmıştır. K-ortalama kümeleme analizi yöntemi 1995 ve 2015 yılı verisi için iki kez tekrarlanmış ve sonuçlar karşılaştırılmıştır.

Çeştepe, Vergil ve Ergun (2012), hizmet ticareti ve ekonomik büyüme arasındaki ilişkiyi geliştirmiş ve geliştirmekte olan ülkelerde araştırmışlardır. Bu amaçla 1980-2007 döneminde 35 ülkeye ait, GSYİH ile yatırım harcamaları, kamu harcamaları ve hizmet ticareti verileri arasında panel veri analizi kullanılmıştır. Tüm ülkeler, gelişmiş ülkeler ve geliştirmekte olan ülkeler için ayrı ayrı gerçekleştirilen panel regresyon analizi sonucunda hizmet ticareti geliştirmekte olan ülkeleri olumlu etkilerken, gelişmiş ülkelerde herhangi bir anlamlı etkiye rastlanamamıştır.

Uçan ve Koçak (2018), Pedroni eşbütünleşme testi ile insani gelişme endeksi ve büyüme arasındaki ilişkiyi incelemiştir. Bu amaçla, 1990-2015 dönemi için Türkiye, Almanya, ABD, Norveç ve İtalya'ya ait GSYİH, insani gelişme endeksi, eğitim endeksi ve sağlık endeksi verileri kullanılmıştır. Pedroni panel eşbütünleşme testi sonucuna göre, uzun dönemde sağlık, eğitim ve insani gelişme arasında ekonomik büyüme arasında anlamlı bir ilişki vardır. Panel regresyon analizi sonucuna göre insani gelişme endeksi ve eğitim endeksi ekonomik büyümeyi pozitif etkilerken, sağlık endeksi ise büyümeyi negatif etkilemektedir.

Zhang ve Danish (2019) geliştirmekte olan 29 Asya ülkesi için 1990-2016 yıllarını kapsayan, GSYİH, internet kullanımı, mobil telefon kullanımı, insani gelişmişlik indeksi, ticarete açıklık ve doğrudan yabancı sermaye verisini panel veri analizinde kullanmışlardır. Çalışmada ekonomik büyüme bağımlı, diğer değişkenler bağımsız değişkenlerdir. Yazarlar, veri setleri heterojen ve otokorelasyon gösterdiği için Driscoll-Kraay ve dayanıklı tahminçiler kullanmışlardır. Çalışma sonucunda yüksek



insani gelişmişlik ve mobil telefon kullanımı, ekonomik büyümeyi desteklerken, internet kullanımı için bu yönde bir sonuç bulunamamıştır.

Bu çalışmada da, 193 ülke için gelişmişlik düzeyini tahmin eden bir makine öğrenmesi modeli geliştirilmiştir. Modelin bağımlı değişkeni, “Yüksek Gelir”, “Üst Orta Gelir”, “Alt Orta Gelir” ve “Düşük Gelir” olmak üzere 4 sığtan oluşan gelişmişlik düzeyidir. Bağımsız değişkenler ise, Dünya Gelişmişlik İndeksi (World Development Indicators - WDI) veri tabanından elde edilen ve ülkelerin gelişmişlik seviyesiyle ilişkilendirilebilecek 26 değişkenden oluşmaktadır. Bu değişkenlere ve kısaltmalarına Tablo 1’den ulaşılabilir. Bu tablo incelendiğinde, bu değişkenlerin, sağlık, sanayi, ekonomi, tarım ve teknoloji ile ilgili olduğu görülmektedir. Tüm değişkenler 2019 yılı için toplanmıştır. Çalışmada karar ağacı ve rastgele orman makine öğrenmesi metodları kullanılmıştır. Bu çalışmadan elde edilecek sonuçlar, farklı gelişmişlik seviyesindeki ülkelerin karakteristiklerini ve hangi değişkenlerden en çok etkilendiğini ortaya koyacaktır.

## Yöntem

### Veri Seti

Çalışmada, Dünya Bankası’nın veri bankasından alınan Dünya Gelişmişlik Göstergeleri (World Development Indicators) veri seti kullanılmıştır. Veri seti, 193 ülkeye ait 2019 yılı verilerini içermektedir. Ülkelerin gelişmişlik seviyeleri; Yüksek Gelir (12.535\$ veya daha fazla), Üst Orta Gelir (4.046\$ ile 12.535\$ arası), Alt Orta Gelir (1.035\$ ile 4.046\$ arası) ve Düşük Gelir (1.035\$ veya daha az) olarak sınıflandırılmıştır. Bu sınıflama için Dünya Bankası Atlas Yöntemi’ni kullanmıştır. Atlas yöntemi, Dünya Bankası tarafından ekonomilerin büyüklüğünü ABD doları cinsinden gayri safi milli gelire göre tahmin etmek için kullanılan bir yöntemdir. Dünya Bankası, belirli operasyonel ve analitik amaçlar için ABD doları cinsinden gayri safi milli geliri (GNI-eski adıyla GSMH) hesaplarırken, basit döviz kurları yerine Atlas dönüştürme faktörünü kullanmaktadır. Atlas dönüşüm faktörünün amacı, ulusal gelirlerin ülkeler arası karşılaştırmasında enflasyonun neden olduğu döviz kuru dalgalanmalarının etkisini azaltmaktır. Herhangi bir yıl için Atlas dönüşüm faktörü, o yıl için bir ülkenin döviz kurunun ve önceki iki yıl için döviz kurlarının ortalamasıdır ve ülkedeki enflasyon oranı ile uluslararası enflasyon arasındaki farka göre düzeltilmektedir (World Bank, 2022).

### Karar ağacı

Karar ağacı, hem regresyon ağacını hem de sınıflama ağacını bir arada sunan bir algoritmaya sahiptir. Bu algoritma, veri setinin sahip olduğu özellikleri kullanarak oluşturduğu kurallar yardımıyla veri setini alt örneklemelere ayırmaktadır. Düğümler ve



dallardan meydana gelen karar ağacı diyagramları görsel olarak da sunulabilmektedir (Çene, 2022).

Veri madenciliği, makine öğrenmesi, yapay zekâ uygulamaları gibi pek çok alanda karar ağaçları yaygın olarak kullanılmaktadır. Yaygın kullanımının başlıca sebepleri arasında kullanım kolaylığı, veriyi görselleştirmesi neticesinde yorumlanmasının kolay olması, hem kesikli hem de sürekli veriler için uygulanabilir olması vb. yer almaktadır.

Karar ağacı elde etmek için pek çok farklı istatistiksel algoritma kullanılmaktadır. Sınıflama ve Regresyon Ağaçları (Classification and Regression Trees, CART), Otomatik Ki-Kare Etkileşim Belirleme (Chi-Squared Automatic Interaction Detection, CHAID), Hızlı Yansız Etkili İstatistiksel Ağaç (Quick, Unbiased, Efficient, Statistical Tree, QUEST), ID3, C4.5 ve C5.0 bu algoritalardan bazılarıdır (Özkan, Çene ve Parim, 2018).

### **Rastgele Orman**

İlk olarak Breiman (2001) tarafından tanıtilen rastgele orman (RO) algoritması, yanıt değişkeni ile ilişkilerinin biçimine ilişkin önceden herhangi bir varsayımda bulunmadan çeşitli türdeki tahmin değişkenlerine dayalı tahmin kuralları oluşturmaktadır (Probst vd. (2019)).

Rastgele orman, popüler makine öğrenimi yöntemleri arasında birçok veri türü için benzersiz bir tahmin doğruluğu ve kolay yorumlanabilen parametrik olmayan bir yöntemdir. RO'da kullanılan rastgele örnekleme ve topluluk stratejileri, daha iyi genellemelerin yanı sıra doğru tahminler elde etmesini sağlamaktadır (Yang vd. (2010)).

Yöntem; her bir ağacın yapısını, ormanın yapısını ve boyutunu ve ayrıca rastgeleliğini kontrol eden birkaç hiperparametreyi içermektedir (Probst, Wright ve Boulesteix, 2019). Farklı bileşenlerin kombinasyonu olan bu yöntemle rastgele ormanları düzgün bir şekilde analiz etmek zordur. Ormanların temel bileşenleri arasında hem torbalama (Bagging) hem de Sınıflandırma ve Regresyon Ağaçları (CART) kritik öneme sahiptir. Torbalama, orijinal veri setinden önyükleme örnekleri oluşturan, her örnekten bir tahmin edici oluşturan ve ortalama olarak karar veren genel bir toplama şemasıdır (Biau ve Scornet, 2016). Problemin karmaşıklığı ve ölçeği nedeniyle tek adımda iyi bir model bulmanın imkânsız olduğu, özellikle büyük, yüksek boyutlu veri kümeleri için, kararsız tahminleri iyileştirmek için en etkili hesaplama prosedürlerden biridir (Bühlmann ve Yu, 2002; Kleiner, Talwalkar, Sarkar ve Jordan, 2014; Wager, Hastie ve Efron, 2014). CART bölünme kriteri ise eksenlere dik en iyi kesimleri seçmek için bireysel ağaçların oluşturulmasında kullanılmaktadır. Sınıflandırma için Gini safsızlığı, regresyon için ise hata tahmin kareler baz alınarak CART bölünme kriteri optimize edilir ve her ağacın her düğümünde en iyi kesim seçilir (Biau ve Scornet, 2016).

## Özellik Seçimi

Özellik seçimi (ÖS), veri ön işlemede en sık kullanılan ve önemli tekniklerden biridir ve makine öğrenme sürecinin vazgeçilmez bir bileşeni haline gelmiştir (Kalousis, Prados ve Hilario, 2007). Makine öğrenmesi ve istatistikte değişken seçimi, öznelik seçimi veya değişken alt küme seçimi olarak da bilinmektedir. ÖS ile ilgili özellikler tespit edilerek ilgisiz, gereksiz veya gürültülü veriler kaldırılmaktadır. Bu süreç tahmin doğruluğunu ve anlaşılabilirliği arttırmaktadır (Kumar ve Minz, 2014). Genel olarak, büyüklük ve değerlendirme ölçütü arasında en iyi bağlılığa sahip alt küme yaklaşımı kullanılarak bir dizi aday özellik elde edilmektedir. Denetimli tümevarımsal öğrenme açısından özellik seçiminde kullanılan diğer yaklaşımlar ise; bir değerlendirme ölçüsünü optimize eden özelliklerin alt kümesinin belirtilen boyutu ve değerlendirme ölçütlerinde belirli bir kısıtlamayı karşılayan alt kümenin daha küçük boyutu yaklaşımlarıdır (Molina, Belanche ve Nebot, 2002).

Özellik seçimi için iki ana yaklaşım vardır. Birincisi bireysel değerlendirme, ikincisi ise alt küme değerlendirmesidir. Özelliklerin sıralanması bireysel değerlendirme olarak bilinmektedir (Guyon ve Elisseeff, 2003). Bireysel değerlendirmede, bireysel bir özelliğin ağırlığı, ilgililik derecesine göre atanmaktadır. Alt küme değerlendirmesinde ise aday özellik alt kümeleri, arama stratejisi kullanılarak oluşturulmaktadır (Kumar ve Minz, 2014). Özellik seçimi için genel prosedürde dört temel adım vardır. Bunlar; alt küme oluşturma, alt kümenin değerlendirilmesi, durdurma kriterleri ve sonuç doğrulamadır.

Bu çalışmada değişken öneminin belirlenmesinde alt küme değerlendirmesi yöntemlerinden biri olan Özyinelemeli Değişken Eleme (Recursive Feature Elimination) yöntemi kullanılmıştır (Guyon, Weston ve Barnhill, 2002). Bu yöntem tüm değişkenleri kullanarak bir model oluşturarak başlar ve ardından her bir değişken için bir önem skoru hesaplar, en az önemli olan değişkeni modelden çıkarır ve modeli tekrar kurar. Değişken azaltmanın artık model başarısını etkilemediği ya da belirlenen değişken sayısına ulaşıldığında algoritma durur.

Çalışmada Özyinelemeli Değişken Eleme, karar ağacı ve rastgele orman analizleri R programlama dili (R Core Team, 2021) içerisindeki caret paketiyle yapılmıştır (Max, Wing, Weston, Williams, Keifer, Engelhardt, Cooper, Mayer, Ziem, Scrucca, Hunt ve Kuhn, 2021).

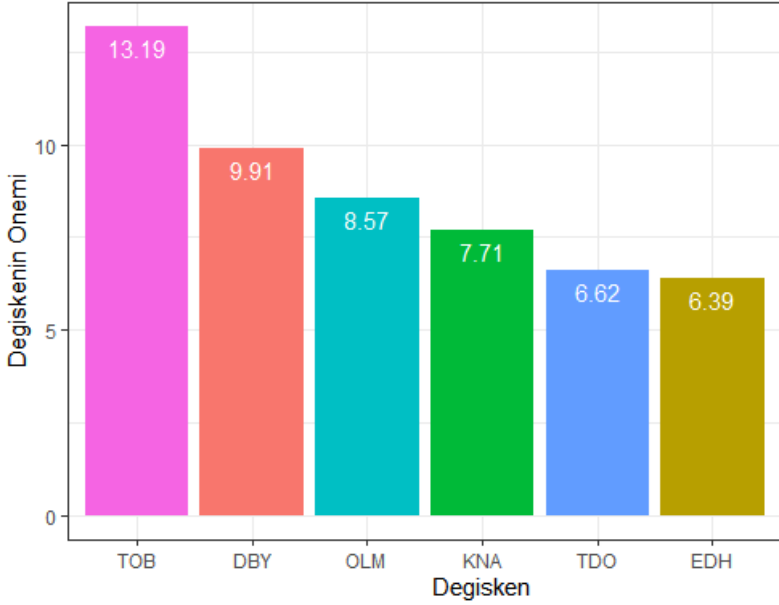
Tablo 1

*Değişken Kısaltmaları, İsimleri ve Tanımları*

<b>Değişkenler</b>	<b>Açıklaması</b>
EDH	Ergen Doğurganlık Hızı (15-19 yaş arası 1000 kadın başına doğum)
TOB	Tarım, Ormancılık ve Balıkçılığın katma değeri (GSYİH yüzdesi)
MHI	Mal ve Hizmet İhracatı (GSYİH yüzdesi)
TDO	Toplam Doğurganlık Oranı (Kişi başı)
DYY	Doğrudan Yabancı Yatırım Net Girişleri (Amerikan Doları)
OAL	Ormanlık Alan (Kilometre kare)
GSYIH	Gayri Safi Yurtiçi Hasıla (Amerikan Doları)
GSYIHB	Gayri Safi Yurtiçi Hasıla Büyümesi (Yıllık yüzde olarak)
GSMH	Kişi Başına Gayri Safi Milli Hasıla (Amerikan Doları)
BSO	Bürüt Sermaye oluşumu (GSYİH yüzdesi)
YTI	Yüksek Teknoloji İhracatı (üretim ihracatının yüzdesi)
KAO	Kızamık Aşısı Olma (12-23 aylık çocukların yüzdesi)
MHIT	Mal ve Hizmet İthalatı (GSYİH yüzdesi)
SAN	Sanayi (inşaat dahil) katma değeri (GSYİH yüzdesi)
ENF	Enflasyon (Yıllık yüzde)
DBY	Doğumda Beklenen Yaşam Süresi (Yıl)
TIC	Mal Ticareti (GSYİH yüzdesi)
ASH	Askeri Harcamalar (GSYİH yüzdesi)
MOB	Mobil Abonelikler (100 kişi başına)
OLM	5 Yaş Altı Ölüm Oranı (Her 1000 doğum için)
NTSE	Net Ticaret Sınırları Endeksi (2000 yılı baz alınarak ölçülen ihracat birim değer endekslerinin ithalat birim değer endekslerine yüzde oranı)
PREM	Alınan Kişisel Havaleler (Amerikan Doları)
NYO	Nüfus Yoğunluğu (Kilometre kare kare başına kişi)
NAR	Nüfus Artışı (Yıllık yüzde)
NUF	Toplam Nüfus
KNA	Kentsel Nüfus Artışı (Yıllık yüzde)

## Bulgular

Bu çalışmada da, 193 ülke için gelişmişlik düzeyini tahmin etmek için kullanılan Bağımsız değişkenler Dünya Gelişmişlik İndeksi (World Development Indicators - WDI) veri tabanından elde edilen 26 değişkenden oluşmaktadır. Gelişmişlik seviyesini etkileyen en önemli değişkenlerin belirlenmesinde, değişken önemi kullanılmıştır. Değişken önemi, Özyinelemeli Değişken Eleme yöntemi ile hesaplanmıştır. Değişken önemi hesabı 5 tekrarlı 10 lu çapraz doğrulama ile yapılmıştır. Sonuçlar incelendiğinde en yüksek doğruluk (0,7375) ve Kappa (0,6375) değerine sahip olan 6 değişkenin gelişmişlik düzeyini tahmin etmede en çok etkili olduğu görülmüştür. Değişkenlerin önem sırası grafik yardımıyla görselleştirilmiş ve Şekil 1'de verilmiştir. Değişken önemi, özellik seçimi için kullanılan sıralama yöntemi kullanılarak hesaplanır. Herhangi bir birime sahip değildir. Bulunan ideal küme boyutu olan 6 değişken için önemler, tüm yeniden örneklemelerdeki modellerin önemlerinin ortalaması alınarak elde edilir. Değişkenler önem sırasına göre TOB, DBY, OLM, KNA, TDO ve EDH olarak sıralanabilir. Diğer 20 değişkenin bunlara göre daha önemsiz olduğu söylenebilir.



Şekil 1. Gelişmişlik seviyesini en çok etkileyen değişkenler

Makine öğrenmesi algoritmalarında kullanılacak olan veri seti 10-katlamalı çapraz doğrulama yöntemi ile bölümlenmiştir. Bölümlenmiş veri seti için değişkenlerin ilk beş gözlem değeri Tablo 2’de verilmiştir. Doğruluk değerleri hesaplanırken kesim noktası 0,5 olarak belirlenmiştir. Karar ağacı yönteminde ağaç budaması için karmaşıklık parametresi 0,115 için doğruluk en yüksek çıkmıştır. Rastgele Orman yöntemi için 500 ağaç kullanılmış ve optimum rastgele alt değişken sayısı 5 için en yüksek doğruluk değeri elde edilmiştir. Tablo 3’te Karar ağacı ve rastgele orman algoritmalarının sonuçları yer almaktadır. Her iki yöntemin de doğruluk oranlarına bakıldığında karar ağacının %65 ve rastgele ormanın ise %70 olduğu görülmektedir. Bu da ülkelerin gelişmişlik seviyelerinin %70 oranında doğru sınıflandırıldığını göstermektedir. Her bir gelir grubu için doğru sınıflandırma oranlarına bakıldığında yüksek gelir grubundaki ülkelerin tamamının doğru sınıflandırıldığı görülmektedir. Düşük, alt ve üst orta gruplarındaki sınıflandırmaların yaklaşık olarak %65 ile %97 arasında olduğu söylenebilir. Bu da alt gelir gruplarındaki ülkelerdeki değişimlerin etkisinden kaynaklanıyor olabilir. Ayrıca, analiz bulguları hakkında daha detaylı bilgi oluşturması açısından duyarlılık, özgüllük, pozitif ve negatif tahmin değerleri, yaygınlık, algılama oranı, algılama yaygınlığı ve dengeli doğruluk değerleri de gelişmişlik düzeylerinin kategorilerinin sınıflandırılmasını göstermesi açısından Tablo 3’te karışıklık matrisleri ise Tablo 4 ve Tablo 5’te verilmiştir.

Tablo 2

*Eğitim ve test verisi için değişkenlerin ilk beş gözlem değeri*

	<b>TOB</b>	<b>DBY</b>	<b>OLM</b>	<b>KNA</b>	<b>TDO</b>
<b>EĞİTİM</b>	0.219	81.866	8.8	-0.009	1.590
	0.630	83.704	4.0	0.783	1.480
	0.316	81.687	2.0	0.891	0.918
	1.844	82.907	2.3	1.121	1.530
	0.218	82.446	2.8	2.228	1.340
	0.916	78.787	6.4	0.701	1.705
<b>TEST</b>	1.392	82.958	2.7	1.330	1.700
	1.071	81.792	3.6	0.817	1.460
	1.007	84.356	2.5	-0.119	1.360
	1.667	83.226	3.1	0.162	0.918

Tablo 3

*Makine Öğrenmesi Algoritmalarının Performansı*

<b>Sınıf</b>	<b>Karar Ağacı</b>				<b>Rastgele Orman</b>			
	<b>Yüksek</b>	<b>Düşük</b>	<b>Alt Orta</b>	<b>Üst Orta</b>	<b>Yüksek</b>	<b>Düşük</b>	<b>Alt Orta</b>	<b>Üst Orta</b>
Duyarlılık (Sensitivity)	1	0.0	0.5000	0.6	0.9231	1	0.5000	0.5000
Özgüllük (Specificity)	0.8333	1	0.8148	0.8519	0.7917	0.9394	1	0.8519
Pozitif Tahmin Değeri (Pos Pred Value)	0.7647	NaN	0.5000	0.6	0.7059	0.6667	1	0.5556
Negatif Tahmin Değeri (Neg Pred Value)	1	0.8919	0.8148	0.8519	0.9500	1	0.8437	0.8214
Yaygınlık (Prevalence)	0.3514	0.1081	0.2703	0.2703	0.3514	0.1081	0.2703	0.2703
Algılama Oranı (Detection Rate)	0.3514	0.0	0.1351	0.1622	0.3243	0.1081	0.1351	0.1351
Algılama Yaygınlığı (Detection Prevalence)	0.4595	0.0	0.2703	0.2703	0.4595	0.1622	0.1351	0.2432
Dengeli Doğruluk (Balanced Accuracy)	0.9167	0.5000	0.6574	0.7259	0.8574	0.9697	0.7500	0.6759
<b>Doğruluk</b>		0.6486				0.7027		

Tablo 4

*Karar ağacı için karışıklık matrisi*

<b>Tahmin</b>	<b>Gerçek</b>			
	<b>Alt Orta Gelir</b>	<b>Düşük Gelir</b>	<b>Üst Orta Gelir</b>	<b>Yüksek Gelir</b>
Alt Orta Gelir	5	4	1	0
Düşük Gelir	0	0	0	0
Üst Orta Gelir	4	0	6	0
Yüksek Gelir	1	0	3	13

Tablo 5

*Rastgele orman için karışıklık matrisi*

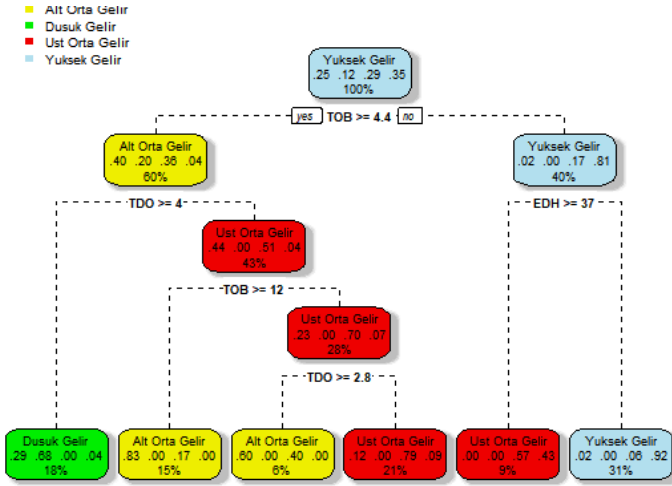
<b>Tahmin</b>	<b>Gerçek</b>			
	<b>Alt Orta Gelir</b>	<b>Düşük Gelir</b>	<b>Üst Orta Gelir</b>	<b>Yüksek Gelir</b>
Alt Orta Gelir	5	0	0	0
Düşük Gelir	2	4	0	0
Üst Orta Gelir	3	0	5	1
Yüksek Gelir	0	0	5	12

Karar ağacı analizinde düğümler, düşük gelir, alt orta gelir, üst orta gelir ve yüksek gelir olarak isimlendirilmiştir. Bunlar sırasıyla, yeşil, sarı, kırmızı ve mavi renkler ile renklendirilmiştir. Olasılık olarak hangi gelir grubu olasılığı daha fazla ise onun rengi ile düğüm renklendirilmektedir. Düğümlerin altında yazan koşul sağlandığında evet yazan sol tarafa, koşul sağlanmıyorsa da hayır yazan sağ tarafa devam edilir. Ayrıca, her bir düğümün altında yüzdesi yer almaktadır. Burada belirtilmesi gereken diğer bir durum ise TOB, TDO ve EDH değişkenlerinin diğer değişkenlere göre daha baskın olduğudur. Karar ağacının kök düğüm ve dallarında geriye kalan 23 değişken bu değişkenlerin baskın olması nedeniyle yer almamıştır.

Ülkelerin gelişmişlik seviyeleri için uygulanan karar ağacı analizi sonuçları Şekil 1’de verilmektedir. Kök düğümden de görülebileceği gibi test veri setindeki ülkelerin %25’i düşük, %12’si alt orta, %29’u üst orta ve %35’i yüksek gelir grubuna dahildir. Kök düğümde yer alan ve gelir gruplarının belirlenmesindeki en önemli değişkenin TOB (Tarım, Ormancılık ve Balıkçılığın katma değeri (GSYİH yüzdesi)) olduğu söylenebilir. TOB değişkeni 4.4’e eşit veya daha fazla olduğunda ülkelerin yüksek gelir grubuna dahil olma olasılığı %4’e gerilemektedir. Bu koşulda, ülkelerin yüzde %40 olasılıkla alt orta gelir grubuna dahil olacağı söylenebilir. TOB değişkeni 4.4’e eşit veya daha fazla olduğunda bir de TDO (Toplam Doğurganlık Oranı (Kişi başı)) değişkeni de 4’ten fazla ise %68 oranında düşük gelir grubuna sahip bir ülke olduğu hatta %96 olasılıkla da düşük ve alt gelir grubuna sahip bir ülke olduğu söylenebilir.

TDO değişkeni 4’ten az ve TOB değişkeni de 12’ye eşit ya da fazla olduğunda %83 olasılıkla alt gelir grubuna sahip bir ülke olduğu görülebilir. Ayrıca sonuçlar göstermektedir ki TOB değişkeni 4.4 ile 12 arasında olduğunda, TDO değişkeni 2.8 ile 4 arasında ise %60 oranında alt gelir grubuna sahip bir ülke, TDO değişkeni 2.8 ‘den az ise de %79 oranında üst orta gelir grubuna sahip bir ülke olacağı söylenebilir.

TOB değişkeni 4.4’ten az bir orana sahip ise %81 olasılıkla yüksek gelir grubuna sahip bir ülke olacağı, buna ek olarak EDH (Ergen Doğurganlık Hızı) değişkeninin de 37 ‘den az olması durumunda da yüksek gelir grubuna sahip bir ülke olması olasılığının %92’ye yükseleceği söylenebilir. Bu koşullar altında EDH 37’den fazla olursa %57 olasılıkla üst orta gelir grubu bir ülke olacağı aşıkardır.



Şekil 2. Ülkelerin Gelişmişlik Seviyelerinin Belirlenmesi için Karar Ağacı Analizi Sonuçları

## Sonuç ve Tartışma

Karar ağacı algoritması ülkelerin gelişmişlik düzeyi üzerinde etkili üç değişkenin olduğunu göstermiştir. Bu değişkenlerden Ergen Doğurganlık Hızı, Yüksek Gelire Sahip Ülkelerle, Üst Orta Gelir Düzeyi'ne sahip ülkeleri birbirinden ayırmaktadır. Ergen Doğurganlık Hızı daha düşük olan ülkeler Yüksek Gelir Düzeyi'nde yer almaktadır. Bu sonuç Jemna (2015) ve Lacalle Calderon (2017)'in çalışmalarıyla örtüşmektedir. Benzer şekilde Toplam Doğurganlık Oranı arttıkça gelir düzeyinin azaldığını söyleyen çalışmalara karşın bu çalışma, Toplam Doğurganlık Oranı ile Tarım, Orman ve Balıkçılık'ın GSYİH daki payının bir ülkenin gelir düzeyini belirlerken birlikte değerlendirilmesi gerektiğini önermektedir. TOB orta ve yüksek düzeyde ( $\geq 4,4$ ) ve TDO yüksek olduğunda ( $\geq 4$ ) ülkeler düşük gelişmişlik düzeyindeyken; TOB orta ( $\geq 4,4$  ve  $< 12$ ) ve TDO düşük ( $< 2,8$ ) iken ülkeler üst orta gelir grubunda yer almaktadır. Alt orta ülkeler ise iki durumda karşımıza çıkmaktadır. Bunlardan birincisi TOB'un yüksek ( $\geq 12$ ) ve TDO'nun orta ve düşük olduğu ( $< 4$ ) durum, ikincisi ise TOB'un orta ( $\geq 4,4$  ve  $< 12$ ) ve TDO'nun orta düzeyde olduğu durumdur.

Bu çalışmada, ülkelerin gelir seviyelerinin tahmini için karar ağacı ve rastgele orman yöntemleriyle makine öğrenmesi modelleri kurulmuştur. Aday değişkenler öncelikle özellik seçimi yöntemiyle 6 değişkene indirgenmiş ve makine öğrenmesi modelleri bu 6 değişkenle oluşturulmuştur. Karar ağacı modeli %65 doğruluğa sahipken, rastgele orman modeli %70 doğruluğa sahiptir. Karar ağacı modeli, ülkelerin ekonomik gelişmişlik seviyesi üzerinde en etkili değişkenlerin Ergen Doğurganlık Hızı, Toplam Doğurganlık Oranı ve, Tarım, Orman ve Balıkçılık'ın GSYİH 'daki payı olduğunu göstermiştir.



**Hakem Değerlendirmesi:** Dış bağımsız.

**Çıkar Çatışması:** Yazarlar çıkar çatışması bildirmemiştir.

**Finansal Destek:** Yazarlar bu çalışma için finansal destek almadığını beyan etmiştir.

**Yazar Katkısı:** Çalışma Konsepti/Tasarımı: B.Ö., C.P., E.Ç.; Veri Toplama: B.Ö., C.P., E.Ç.; Veri Analizi /Yorumlama: B.Ö., C.P., E.Ç.; Yazı Taslağı: B.Ö., C.P., E.Ç.; İçeriğin Eleştirel İncelemesi: B.Ö., C.P., E.Ç.; Son Onay ve Sorumluluk: B.Ö., C.P., E.Ç.

**Peer-review:** Externally peer-reviewed.

**Conflict of Interest:** The authors has no conflict of interest to declare.

**Grant Support:** The authors declared that this study has received no financial support.

**Author Contributions:** Conception/Design of study: B.Ö., C.P., E.Ç.; Data Acquisition: B.Ö., C.P., E.Ç.; Data Analysis/ Interpretation: B.Ö., C.P., E.Ç.; Drafting Manuscript: B.Ö., C.P., E.Ç.; Critical Revision of Manuscript: B.Ö., C.P., E.Ç.; Final Approval and Accountability: B.Ö., C.P., E.Ç.

---

## Kaynakça/References

- Ahmad Z ve Saleem A. (2012). Predicting Level of Development for Different Countries. *Journal of Sustainable Development*, 5(11). doi:10.5539/jsd.v5n11p15
- Biau G ve Scornet E. (2016). A random forest guided tour. *Test*, 25(2), 197-227.
- Bloom DE ve Canning D. (2000). The health and wealth of nations. *Science*, 287(5456), 1207-1209.
- Breiman L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Bulut Ş, Babacan A, ve Ertekin Ş. (2021). Ekonomik Büyümenin Belirleyicilerinin Farklı İnsani Gelişmişlik Düzeyindeki Ülkelere Göre Analizi. *Sayıştay Dergisi*, (120), 89-114.
- Bühlmann P ve Yu B. (2002). Analyzing bagging. *The Annals of Statistics*, 30(4), 927-961.
- Çene E. (2022). Makine Öğrenmesi Yöntemleriyle Euroleague Basketbol Maç Sonuçlarının Tahmin Edilmesi Ve Maç Sonuçları Üzerinde En Etkili Değişkenlerin Bulunması. *Spor ve Performans Araştırmaları Dergisi*, 13(1), 29-52.
- Çeştepe H, Vergil H, ve Ergun H. (2012). Hizmet Ticaretinin Büyümeye Etkisi: Gelişmiş ve Gelişmekte Olan Ülkeler Üzerine Bir Panel Veri Analizi. *Business and Economics Research Journal*, 3(4), 91-105.
- Demiray Erol E. (2013). Türkiye ve Avrupa Birliği Üyesi Ülkelerin Sosyo-Ekonomik Gelişmişlik Düzeylerinin Karşılaştırmalı Analizi. *Sosyal ve Beşeri Bilimler Dergisi*, 5(1), 198-208.
- Demirci E ve Karaatlı M. (2019). Ülkelerin Gelişmişlik Seviyelerinin Tahmininde Kullanılan Sınıflandırma Algoritmalarının Karşılaştırılması. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 24(3), 703-714.
- Erden Özsoy C ve Tosunoğlu BT. (2017). GSYH'nin ötesi: ekonomik gelişmenin ölçümünde alternatif metrikler. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26(1), 285-301.
- Guisan M-C ve Aguayo E. (2007). Health expenditure, poverty and economic development in Latin America 2000-2005. *International journal of Applied Econometrics and Quantitative studies*, 4(1), 5-24.
- Guyon I ve Elisseeff A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar), 1157-1182.
- Guyon I, Weston J, ve Barnhill S. (2002). Gene selection for cancer classification using DCA. *Machine Learning*, 46, 389-422.
- Jemna D-V. (2015). Causality Relationship between Economic Development and Fertility in Romania on Regional Level. *Procedia Economics and Finance*, 20(15), 334-341. doi:10.1016/s2212-5671(15)00081-7

- Kalousis A, Prados J, ve Hilario M. (2007). Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowledge and information systems*, 12(1), 95-116.
- Kleiner A, Talwalkar A, Sarkar P, ve Jordan MI. (2014). A scalable bootstrap for massive data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(4), 795-816.
- Koçak E ve Uçan O. (2018). İnsani gelişme endeksi ile büyüme ilişkisi: Pedroni eşbütünlük testi. *Journal of Politics Economy and Management*, 1(2), 55-61.
- Koşar Taş Ç ve Örk Özel S. (2017). Faktör Analizi Yöntemi İle Türkiye ve Avrupa Birliği Üyesi Ülkelerin Sosyo-Ekonomik Göstergeler Bakımından Gelişmişlik Düzeylerinin Karşılaştırılması. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 26(3), 60-77.
- Kubar Y. (2016). Az gelişmiş ve gelişmekte olan ülkelerin kalkınma göstergeleri ile ekonomik büyüme arasındaki ilişki: Bir panel veri analizi (1995-2010). *Ardahan Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 2(4), 65-99.
- Kumar V ve Minz S. (2014). Feature selection: a literature review. *SmartCR*, 4(3), 211-229.
- Lacalle-Calderon M, Perez-Trujillo M, ve Neira I. (2017). Fertility and Economic Development: Quantile Regression Evidence on the Inverse J-shaped Pattern. *European Journal of Population*, 33(1), 1-31. doi:10.1007/s10680-016-9382-4
- Max A, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, ... Kuhn MM. (2021). Package 'caret' R topics documented :
- Molina LC, Belanche L, ve Nebot À. (2002). Feature selection algorithms: A survey and experimental evaluation. *IEEE International Conference on Data Mining Proceedings*. içinde (ss. 306-313). IEEE.
- Özkan B, Çene E, ve Parim C. (2018). İstanbul'daki Üniversite Öğrencilerinin Memnuniyet Düzeylerinin Çok Değişkenli İstatistiksel Yöntemler ve Karar Ağacıyla İncelenmesi. *International Conference on Data Science and Applications* içinde (ss. 489-505).
- Öztürk SG. (2007). Classifying and predicting country types through development factors that influence economic, social, educational and health environments of countries. *SWDI Proceedings papers S*, 759, 665-674.
- Parim C, Özkan B, ve Cene E. (2019). Clustering of Countries by the Factors Affecting Levels of Development and It's Comparison by Years. *Data Science and Applications*, 2(1), 4-7.
- Probst P, Wright MN, ve Boulesteix A. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery*, 9(3), e1301.
- R Core Team. (2021). R: A language and environment for statistical computing, R Foundation for Statistical Computing., Vienna, Austria. <https://www.r-project.org/>. adresinden erişildi.
- Rebala G, Ravi A, ve Churiwala S. (2019). *An introduction to machine learning*. Cham: Springer.
- Stec M, Filip P, Grzebyk M, ve Pierscieniak A. (2014). Socio-economic development in the eu member states – Concept and classification. *Engineering Economics*, 25(5), 504-512. doi:10.5755/j01.ee.25.5.6413
- Upreti P. (2015). Factors affecting economic growth in developing countries. *Major Themes in Economics*, 17(1), 37-54.
- Wager S, Hastie T, ve Efron B. (2014). Confidence intervals for random forests: The jackknife and the infinitesimal jackknife. *The Journal of Machine Learning Research*, 15(1), 1625-1651.
- World Bank. (2022). The World Bank Atlas method - detailed methodology. 6 Temmuz 2022 tarihinde <https://datahelpdesk.worldbank.org/knowledgebase/articles/378832-what-is-the-world-bank-atlas-method> adresinden erişildi.

Yang P, Hwa Yang Y, B Zhou B, ve Y Zomaya A. (2010). A review of ensemble methods in bioinformatics. *Current Bioinformatics*, 5(4), 296-308.

Zhang J ve Danish. (2019). The dynamic linkage between information and communication technology, human development index, and economic growth: evidence from Asian economies. *Environmental Science and Pollution Research*, 26(26), 26982-26990.