

Sosyal Siyaset Konferansları Dergisi Journal of Social Policy Conferences

Araştırma Makalesi | Research Article

 Açık Erişim | Open Access

Türkiye’de İllere Göre İç Göçe İlişkin Makine Öğrenmesi Yöntemleri ile Değişken Önemi ve Göç Tahmini

Machine Learning Methods for Feature Importance and Migration Prediction Related to
Internal Migration by Province in Turkey



Meryem Pulat Öztürk¹  

¹ Fırat Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, Elâzığ, Türkiye.

Öz

Bu çalışmada, 2018-2022 yılları arasında Türkiye’de il bazında iç göçü etkileyen değişkenler değişken önemi ile belirlenmiş ve makine öğrenmesi algoritmaları ile göç tahmini yapılmıştır. Alınan göç ve verilen göç ayrı ayrı ele alınmıştır. Çalışmanın temel amacı, alınan ve verilen göçü etkileyen faktörleri ayrıntılı biçimde belirlemek ve tahminleme sürecinde hangi değişkenlerin daha açıklayıcı olduğunu ortaya koymaktır. Bu yönüyle çalışma, Türkiye’de iç göçe yönelik makine öğrenmesi yaklaşımıyla gerçekleştirilen nadir örneklerden biri olup, alana özgün katkılar sunmaktadır. Yapılan çalışmada 2018-2022 yılları arasında il bazında erişilebilen ve literatürde sıklıkla kullanılan: akademisyen sayısı, hastane sayısı, özel hastane sayısı, GSYİH, öğretmen sayısı, okul sayısı, tarım alanı, nüfus yoğunluğu, nüfus artış hızı, muhtemel eğitim süresi, ortalama eğitim süresi, trafik kaza sayısı, otomobil sayısı, traktör sayısı, bitirilen eğitim durumu, ithalat, ihracat, hekim sayısı, hastane yatak sayısı ve konut satış sayısı değişkenleri alınmıştır. R programında; Caret paketi ile çeşitli modelleme teknikleri, Boruta algoritması, Özyinelemeli Özellik Eleme algoritması ve Ortak önem skoru kullanılarak değişken önemi belirleme işlemi gerçekleştirilmiştir. Daha sonra tüm değişkenler ve en önemli değişkenler dikkate alınarak; basit doğrusal regresyon, Destek Vektör Regresyonu, Karar Ağaçları, Rassal Ormanlar, Güçlendirme, Gradyan Artırma Makinesi, Çantalama ve Yığınlama algoritmaları kullanılarak tahmin yapılmıştır. Elde edilen sonuçlara göre, özellikle trafik kaza sayısı, öğretmen sayısı, akademisyen sayısı, hastane sayısı ve konut satış sayısı gibi değişkenler hem alınan hem de verilen göçte öne çıkan belirleyiciler arasında yer almıştır. Ayrıca, değişken seçimi sonrası yapılan modellemelerin hata düzeylerini düşürdüğü ve özellikle topluluk öğrenmesi algoritmalarının (örneğin stacking) yüksek başarı sağladığı görülmüştür.

Abstract

In this study, the variables affecting internal migration on a provincial basis in Turkey between 2018 and 2022 were determined using feature importance, and migration estimates were made using machine learning algorithms. Incoming and outgoing migrations were considered separately. The main objective of the study is to determine in detail the factors that influence incoming and outgoing migration and to reveal which variables are more explanatory in the prediction process. In this regard, this study is one of the rare examples of a machine learning approach on internal migration in Turkey and provides original contributions to the field. In the study, the following features were taken that can be accessed on a provincial basis between 2018 and 2022 and are frequently used in the literature: number of academicians, number of hospitals, number of private hospitals, GDP, number of teachers, number of schools, agricultural areas, population density, population growth rate, probable duration of education, average duration of education, number of traffic accidents, number of automobiles, number of tractors, completed education status, import, export, number of physicians, number of hospital beds, and number of housing sales. In the R programme, the feature importance determination process was performed using the Caret package with various modelling techniques, the Boruta algorithm, the recursive feature elimination algorithm, and the overall importance



Atıf | Citation: Öztürk, M. P. (2025). Türkiye’de illere göre iç göçe ilişkin makine öğrenmesi yöntemleri ile değişken önemi ve göç tahmini. *Sosyal Siyaset Konferansları Dergisi–Journal of Social Policy Conferences*, (89), 129-160. <https://doi.org/10.26650/jspc.2025.89.1604614>

 This work is licensed under Creative Commons Attribution-NonCommercial 4.0 International License.  

 2025. Öztürk, M. P.

 Sorumlu Yazar | Corresponding author: Meryem Pulat Öztürk mpulat@firat.edu.tr



score. Then, considering all and the most important features, estimation was made using simple linear regression, support vector regression, decision trees, random forests, boosting, gradient boosting machine, and bagging and stacking algorithms. The findings indicate that variables such as the number of traffic accidents, number of teachers, number of academics, number of hospitals, and number of housing sales are among the prominent determinants of both incoming and outgoing migration. Moreover, the application of variable selection significantly reduced prediction errors, and ensemble learning algorithms—particularly stacking—achieved notably high performance.

Anahtar Kelimeler İç göç · makine öğrenmesi · değişken önemi · tahmin

Keywords Internal migration · machine learning · feature importance · prediction

Yazar Notu Bu çalışma, "22. Uluslararası Bilimsel Araştırmalar Kongresi, (UBAK) 08-09 Mart 2025" özet bildirisi olarak sunulmuştur ve özet bildirisi kitabında yayınlanmıştır.

Author Note This study was presented as an abstract at the "22nd International Scientific Research Congress (UBAK) 08-09 March 2025" and was published in the abstract proceedings book.

Extended Summary

This study aims to analyse the factors affecting internal migration in Turkey on a provincial basis between 2018 and 2022. The main objectives of the research are to identify the most significant variables influencing internal migration and to predict migration patterns using machine learning algorithms. Predictions for immigration and emigration were made using all available variables and, alternatively, only the most significant variables identified through feature importance methods. The performance of these predictions was then compared.

The dataset used in this research includes variables commonly employed in the literature and accessible at the provincial level. These variables encompass the number of academics, hospitals, private hospitals, gross domestic product (GDP), number of teachers, number of schools, agricultural area, population density, population growth rate, expected years of schooling, average years of education, number of traffic accidents, number of automobiles, number of tractors, educational attainment, imports, exports, number of physicians, hospital bed capacity, and number of housing sales.

For feature importance determination, several methods were employed in R programming, including the Boruta algorithm, recursive feature elimination (RFE), and the joint importance score approach. Modelling and predictions were performed using a variety of algorithms, including simple linear regression, support vector regression, decision trees, random forests, boosting, gradient boosting machines, bagging, and stacking. According to the findings obtained for the dependent variable of migration, the variables in the top 10 are the number of traffic accidents, number of academicians, number of hospitals, number of private hospitals, number of schools, number of housing sales, number of teachers, number of automobiles, number of hospital beds, population growth rate, number of physicians, population density, import, GDP, completed education status, agricultural area, average education period, and export. The algorithms that give the best results are the common importance score method and the stacking (XGBOOST, GBM, and BAGGING) algorithm after variable importance (for RMSE and MSE). The best result for the MAE value is the stacking (SVM, XGBOOST, and GBM) algorithm with all variables. When the results are examined, the common importance score method and the stacking (RF, XGBOOST, GBM) algorithm after variable importance can be preferred since there is not much difference between them.

We observed that the algorithms’ performance generally improved after feature selection, resulting in lower error values. In addition, for the migration-dependent variable taken, most algorithms were seen to have better performance values with the common importance score approach compared to the first 10 feature selections. For the given migration-dependent variable, most algorithms were seen to have better performance values with the first ten feature selection approaches compared to the common importance score approach. When the algorithms that gave the best results were examined, it was seen that ensemble methods mostly had good performance.

If we talk about the contribution of the study, it is seen that econometric methods are mostly used in the literature review of studies conducted on internal migration in Turkey. No study was found on variable importance and migration estimation with the machine learning approach. In addition, a comprehensive model test was conducted in the study. In this study, important variables in incoming and outgoing migration were determined, and migration estimation was made with a low error value using the machine learning approach. Policies can be developed by

focusing on important variables, and migration estimation can be made using models that perform well. Further studies are aimed at increasing the comprehensibility of the models and making more detailed interpretations using explainable AI analyses.

Türkiye’de İllere Göre İç Göçe İlişkin Makine Öğrenmesi Yöntemleri ile Değişken Önemi ve Göç Tahmini

Göç, insanların geçici veya kalıcı olarak yeni bir yerde yaşamak için bir yerden başka bir yere hareket etmesi olarak adlandırılır. İş bulmak, geçimini sağlamak veya daha iyi yaşam koşullarına sahip olmak için daha elverişli yerlere taşınma isteği göç olgusunun temelini oluşturur. Bu bakış açısından, insanların zevkten ziyade zorunluluktan göç ettiği söylenebilir. Göç, bir toplumun sosyal, kültürel, finansal ve politik bileşenleriyle yakından ilişkilidir ve bunlar üzerinde oldukça etkilidir. Göç, iç ve dış göçler olmak üzere iki grupta sınıflandırılır ve ele alınır (Doğan, 2017, s.49). İç göç, bir bireyin kendi ülkesi içinde bir bölgeden diğerine daimî veya geçici ikamet amacıyla hareket etmesidir. İç göç, insanların yaşadıkları şehirdeki sosyal, kültürel ve ekonomik sorunlar nedeniyle ülkenin başka bir şehrinde yaşamayı tercih etmeleri durumunu ifade eder. Yönüne bakıldığında, “kırsal-kırsal”, “kırsal-kentsel”, “kentsel-kırsal”, “kentsel-kentsel” olmak üzere dört tür iç göç vardır. Ayrıca, nedenlere bağlı olarak gönüllü ve zorunlu göçler de vardır. Birincisi daha iyi yaşam koşulları arayışı için yapılan göçü ifade ederken, ikincisi bir hükümet politikası veya bir bölgedeki yaşam koşullarındaki hızlı bir değişiklik nedeniyle gerçekleşen göçü ifade eder. Ayrıca, geçici ve kalıcı göçler de vardır (Ozsari ve Food, 2016, s.36). Türkiye’de akış çoğunlukla kırsaldan kentsel ve kentselden kentsel doğruya. Kentleşme, Türkiye’deki endüstriyel gelişmeler sayesinde ivme kazanmış ve bu da karşılığında göç oranlarını artırmıştır. 1950’lerden itibaren göç giderek artmaktadır. Daha sonra endüstrileşme ve kentleşmenin etkisiyle akışın yönü kırsaldan kentsel doğru olmuşken, 1980’lerden sonra çoğunlukla kentselden kentsel doğru gerçekleşmiştir. 1950’lerden sonra endüstriyel faaliyetlerin neden olduğu nüfus hareketleri, kırsal kesimden insan nüfusunun hızla boşalmasına ve kentsel alanlarda nüfus patlamasına yol açmış, bu da kırsal ve kentsel alanlarda eşitsiz dağılımlara yol açmış ve heterojen demografik örüntülere neden olmuştur. Ancak kırsal koşullar iyileştirilir ve yatırımlar yoluyla daha müreffeh bir yaşam kurulursa bu düzensiz dağılım çözülebilir (Doğan, 2017, s.49). Başlangıçta kırsaldan kentsel göçler bir sorun olarak görülmemiştir. Ayrıca devlet tarafından teşvik edilmiştir. Ancak sonraki yıllarda özellikle kentsel alanlardaki ekonomik ve sosyal sorunlar, insanların kırsaldan kente göçü bir sorun olarak görmesine neden olmuştur. Bu nedenle köylerden göçün önlenmesi için çok sayıda ulusal ve yerel plan ve program üretilmiştir. Güneydoğu Anadolu Projesi ve Doğu Anadolu Projesi bu tür girişimlere güzel birer örnektir (Doğan, 2017, s.50).

Türkiye’de iç göç, kırsal alanlardan kentsel alanlara, küçük bir şehirden daha büyük bir şehre veya başka bir küçük şehre veya büyük bir şehirden başka bir şehre yapılan hareketleri ifade eder. İç göçler, geçici (mevsimlik) ve daimî olmak üzere ikiye ayrılır. İç göç, bir ülkenin nüfusunu değiştirmez, ancak bölgelerin veya şehirlerin nüfus oranlarını değiştirir. İç göçün nedenleri arasında 1950’lerden itibaren sanayileşmeyle birlikte iş gücünün işten çıkarılması, hızlı tarımsal makineleşme, olumsuz yönde değişen arazi-kişi oranı ve daha iyi geçim kaynağı elde etme isteği yer alır. Sosyal, ekonomik ve siyasal nedenlerle ortaya çıkan iç göçler, bir ülkenin nispeten daha gelişmiş bölge ve şehirlerinden az gelişmiş olanlarına doğru gerçekleşir (Doğan, 2017, s.51). Kırsal ve kentsel nüfus gelişmiş ülkeler arasında homojen bir şekilde dağılmışken, gelişmekte olan ülkelerde heterojen bir dağılım vardır. Yani bir şehirde yoğun iç göç hareketleri varsa, gelecekte bazı sorunlar yaşaması muhtemeldir.

İç göç Türkiye için önemli bir sorundur ve iç göçle ilgili çalışmaların çok önemlidir. Bu durum nedeniyle, bu çalışmanın amacı Türkiye’de iç göç için alınan göç ve verilen göç ayrı ayrı ele alınarak değişken önemleri bulunup; verilen göç ve alınan göçün tahmini tüm değişkenler ve değişken önemi sonrası en önemli olan değişkenler dikkate alınarak analizler yapılarak performansları karşılaştırmaktır. Çalışmada yazarın teorik çerçevesi, mevcut literatürden elde edilen ekonometrik yöntemlerin ötesine geçerek makine öğrenmesi yaklaşımlarını kullanarak iç göçün dinamiklerini daha derinlemesine anlamayı hedeflemektedir. Bu bağlamda daha önceki çalışmaların sınırlamaları ve metodolojik eksiklikleri göz önünde bulundurularak, göç tahminine yönelik daha doğru ve yorumlanabilir sonuçlar üretmeyi amaçlamaktadır. Bu kapsamda yapılan çalışmada 2018-2022 yılları arasında il bazında verileri analiz edilmiştir.

LİTERATÜR İNCELEMESİ

Gezici ve Keskin (2005), Türkiye’deki bölgesel eşitsizlikler ve iç göç arasındaki etkileşimi regresyon analizi ile incelemiştir. Çalışma, iç göç hareketlerinin Türkiye’deki bölgesel kalkınma farklılıkları üzerinde önemli etkiler yarattığını göstermektedir. Özellikle, ekonomik fırsatların daha fazla olduğu büyük kentlere ve gelişmiş bölgelere yönelen göç, geri kalmış bölgelerde nüfus azalmasına ve ekonomik durgunluğa yol açmaktadır. Yazarlar, bölgesel kalkınma stratejilerinin göç hareketleriyle olan ilişkisine dikkat çekerek, iç göçün Türkiye’deki bölgesel eşitsizlikleri artırıcı bir rol oynayabileceğini vurgulamaktadır. Çalışma, iç göç ve bölgesel eşitsizliklerin karşılıklı olarak birbirini etkilediğini öne sürmekte ve bu dinamiklerin sosyoekonomik politikalar kapsamında ele alınması gerektiğini belirtmektedir.

Bülbül ve Kose (2010), Türkiye’deki bölgelerarası iç göç hareketlerini çok boyutlu ölçekleme yöntemi kullanarak analiz etmiştir. Bu yöntemle, Türkiye’deki farklı bölgeler arasındaki göç hareketlerinin benzerlik ve farklılıkları demografik özellikleri, sosyo-ekonomik göstergeleri ve göç verileri bakımından incelenmiştir. Çalışma, Türkiye’deki göç desenlerinin bölgesel kalkınmışlık seviyelerine göre nasıl farklılık gösterdiğini ortaya koymaktadır. Özellikle, göç hareketlerinin ekonomik fırsatlar ve yaşam standartları yüksek olan bölgelere doğru yoğunlaştığı belirtilmiştir.

Ercilasun, Gencer ve Ersin (2011), Türkiye’deki iç göçleri belirleyen faktörleri ekonometrik analizle modellemişlerdir. Türkiye’nin iç göçünü, bulundukları il dışında doğan nüfus, üniversite öğrencisi sayısı ve ortalama ücretler için geliştirilen bir vekil değişken gibi değişkenlere dayanarak açıklamaya çalışmışlardır. Ekonomik fırsatlar, işsizlik oranları, gelir düzeyleri, eğitim ve sağlık hizmetleri gibi faktörlerin iç göçü etkileyen ana etmenler olduğu tespit edilmiştir. Ayrıca, iç göçün, özellikle gelişmiş bölgelere doğru yönelme eğiliminde olduğu vurgulanmıştır.

Çatalbaş ve Yazar (2015), Türkiye’deki bölgeler arası iç göçü etkileyen faktörleri panel veri analizi kullanarak incelemiştir ve bu faktörler arasındaki ilişkileri gösteren ekonometrik model panel veri analizi yöntemiyle tahmin edilerek bir sonraki dönem için öngörülmesi amaçlanmıştır. Panel veri analizi sonuçları, bölgeler arası göç üzerinde bölgenin zenginlik düzeyinin olumlu, enflasyon oranı, istihdam oranı ve terör sorununun ise olumsuz etkiye sahip olduğunu göstermektedir. Özellikle, iş ve gelir fırsatlarının daha fazla olduğu bölgelere göç eğiliminin arttığı görülmüştür.

Ozsari ve Food (2016), Türkiye’de iç göçün nedenlerini logit yöntemi ile analiz etmiştir. Çalışmada kullanılan logit analiz, göç üzerinde etkili olan faktörlerin ağırlıklarını ve önem derecelerini belirleyerek, hangi değişkenlerin göç kararını en çok etkilediğini tespit etmeye yönelik veriler sunmaktadır. Analiz sonuçlarına göre, yüksek boşanma oranının Türkiye’de iç göçün önemli bir faktörü olduğu ve bu değişken ile iç göç arasında pozitif bir ilişki olduğu da tespit edilmiştir. Bu değişkenlere ek olarak, Türk halkı intihar oranının yüksek olduğu şehirlerden göç etmeyi tercih etmektedir.

Manavgat ve Saygılı (2016), Türkiye’de iç göçü etkileyen faktörleri mekânsal panel veri analizi yöntemiyle incelemiştir. Çalışmada, Türkiye’deki bölgeler arası ekonomik, sosyal ve çevresel değişkenlerin göç üzerindeki etkileri ele alınmıştır. Mekânsal panel veri analizinin kullanımı, farklı bölgeler arasındaki göç eğilimlerini ve bu eğilimlerdeki mekânsal bağımlılıkları anlamaya olanak tanımıştır. Bulgular, ekonomik fırsatların artmasıyla belirli bölgelere olan göçün yoğunlaştığını ve bu faktörlerin göçün yönü üzerinde önemli bir etkisi olduğunu göstermektedir.

Dücan (2016), Türkiye’de iç göçün sosyo-ekonomik nedenlerini bölgesel bir analizle incelemiştir. 21 bölge için panel regresyon analizi yapılmıştır. Bulgular, okullaşma oranındaki artış göçü artırırken, istihdam artışı göçü azaltmaktadır. Terör olaylarındaki artışın da göçe sebep olan önemli bir değişken olarak bulunmuştur. Özellikle gelişmişlik düzeyi düşük bölgelerden ekonomik fırsatların daha fazla olduğu bölgelere yoğun göç gözlemlenmiştir.

Atalay ve Öztürk (2016), Türkiye’deki illeri göç ve işsizlik istatistiklerine göre kümeleme analiziyle incelemiştir. Çalışma, Türkiye’nin illerin göç, işsizlik ve işgücü istatistiklerine göre kümelenerek karşılaştırmalı analizini yapmayı amaçlamaktadır. Kümelenme analizi sonuçları, işsizlik ve göç oranlarının birbirleriyle bağlantılı olarak iller arasında belirgin farklılıklar gösterdiğini ortaya koymuştur. Çalışma, belirli kümeler üzerinden işsizlik ve göç oranlarının nasıl şekillendiğini vurgulamakta ve salt işsizlik ve göç oranlarının birbirlerinin tek başına belirleyicisi olmadığını göstermektedir.

Anavatan (2017), Türkiye’deki iç göç hareketlerini ve bu hareketlerin belirleyicilerini mekânsal veri analizi kullanarak incelemiştir. Ekilen tarım arazi miktarı ve yüksek suç oranı net göçü negatif olarak etkilediği; eğitim düzeyi, servet, sanayi, turizm ve ticaretin artması ise net göçü artırdığı tespit edilmiştir. Ayrıca, mekânsal veri analizi ile bölgeler arasındaki farklılıklar daha net bir şekilde ortaya konmuş ve göçün mekânsal dağılımındaki dengesizlikler tespit edilmiştir.

Çelik ve Arslan (2018), göç ile işsizlik arasındaki ilişkiyi ampirik bir uygulama üzerinden incelemişlerdir. Genel işsizlik ve genç işsizlik ile alınan ve verilen göç ilişkisine ilişkin yapılan korelasyon analizine göre göçün genel işsizlik ve genç işsizlik arasında güçlü ve pozitif bir ilişki olduğunu ve göçün artması durumunda işsizliğin de artacağı görülmüştür. Çalışma, işsizlik oranlarının yüksek olduğu bölgelerden daha düşük işsizlik oranlarına sahip bölgelere göçün artış gösterdiğini ortaya koymaktadır. Yazarlar, Türkiye’deki işsizlik oranlarının göç hareketlerini nasıl etkilediğini ve göçün iş gücü piyasası üzerindeki etkilerini inceleyerek, bu ilişkinin sosyal politikalar ve iş gücü planlaması için önemini vurgulamaktadırlar.

Özdemir (2018), Türkiye’deki bölgelerarası iç göç hareketlerinin belirleyicilerini incelemiştir. Çalışmada panel veri analizi ile Türkiye’deki 26 Alt Bölge için net göç hızının belirleyicileri tespit edilmesi amaçlanmıştır. Bölgeler arası göç üzerinde etkili olan en önemli değişkenlerden eğitim, bölgenin zenginlik düzeyi, kentleşme olumlu yönde terör sorunu ise olumsuz etkiye sahip olduğu bulunmuştur. Özellikle düşük gelişmişlik düzeyine sahip bölgelerden daha yüksek gelişmişlik seviyesine sahip bölgelere göçün arttığı gözlemlenmiştir.

Emeç, Birecikli ve Güler (2019), İstanbul’a yönelik iç göç hareketlerini ekonometrik analiz yoluyla incelemiştir. Çalışmada, İstanbul’a göç eden bireylerin sosyo-ekonomik ve demografik özellikleri ile göçün nedenleri analiz edilmiştir. Model olarak, iki durumlu (birey göç etmiş/etmemiş) ardından çok durumlu logit model uygulanmıştır. Genç bireylerin İstanbul’a göç etme olasılığı daha yüksekken, yaşlı bireylerin bunu yapma olasılığı daha düşük olduğu; inşaat sektöründe çalışanların diğer sektörlerle kıyasla İstanbul’a göç etme eğilimlerinin daha yüksek olduğu ve daha düşük eğitim niteliklerine sahip olanların göç etme olasılığı daha yüksek olduğu tespit edilmiştir.

Özdemir ve Doğan (2022), Türkiye’deki bölgelerarası iç göç hareketlerinin kentleşme üzerindeki etkilerini İstanbul örneği üzerinden incelemiştir. Çalışma, İstanbul’un hızlı kentleşme sürecinde iç göçün rolünü ve bu

göçlerin kentsel yapıya olan etkilerini ele almaktadır. Araştırma, iç göçle birlikte artan nüfusun İstanbul’un altyapı, konut, ulaşım gibi kentsel sorunlarını nasıl etkilediğini ortaya koymaktadır. Sonuçlar, göçün yoğun olduğu bölgelerde düzensiz kentleşmenin arttığını ve şehirdeki ekonomik dengesizlikleri derinleştirdiğini göstermektedir.

Şahin ve Kılıç (2024), Türkiye’de iller arası alınan göç hareketlerini çok boyutlu panel veri analizi kullanarak incelemiştir. Çalışma, Türkiye’de göç alan illerin işsizlik, istihdam, hekim sayısı ve hastane sayısı değişkenleri dikkate alınarak analizler gerçekleştirilmiştir. Bulgularda, istihdam, hekim sayısı ve hastane sayısı, iller arası göç politikalarında önem arz eden unsurlar olduğu görülmüştür.

Bahçıvan ve Kılıç (2024), Türkiye’de iç göçün belirleyici faktörlerini panel veri analizi kullanarak incelemiştir. Çalışmada, ekonomik, sosyal ve demografik değişkenlerin göç üzerindeki etkileri değerlendirilmiştir. Panel veri analizine dayalı bulgular, istihdam oranı, eğitim değişkeni ve nüfus artış oranı net göç üzerinde anlamlı bir etkisi olduğunu ve pozitif yönde etkilediği gösterilmiştir. Çalışma, Türkiye’deki göç hareketlerinin farklı bölgelerdeki ekonomik dengesizlikler ve yaşam standartlarındaki farklılıklardan etkilendiğini vurgulamaktadır.

Gözen ve Çelikay (2024), Türkiye’de kamu harcamaları ve yoksulluğun iç göç üzerindeki mekânsal etkilerini ampirik bir yaklaşımla incelemiştir. Çalışmada, kamu harcamalarının dağılımı ve yoksulluk seviyelerinin göç eğilimleri üzerindeki rolü ele alınmaktadır. Bulgular, yüksek yoksulluk oranlarının iç göçü teşvik edici bir faktör olarak öne çıktığını, kamu harcamalarının ise bölgesel gelişmişlik düzeyini etkileyerek göç kararlarında belirleyici olabildiğini göstermektedir. Yazarlar, kamu harcamalarının dengeli dağılımının göç hareketlerini dengeleyici bir unsur olabileceğini vurgulamaktadır.

Yapılan literatür incelemesinde incelenen çalışmaların bir kısmına yukarıda değinilmiştir. İncelemeler sonucunda, Türkiye’de iç göçe ilişkin yapılan çalışmalarda çoğunlukla ekonometrik yöntemler kullanıldığı görülmüştür. Makine öğrenmesi yaklaşımı ile değişken önemi ve göç tahminine ilişkin çalışmaya rastlanmamıştır.

YÖNTEM

Bu çalışmada, 2018-2022 yılları arasında Türkiye’de il bazında iç göçü etkileyen değişkenler değişken önemi ile belirlenmiş ve makine öğrenmesi algoritmaları ile göç tahmini yapılmıştır. İç göç verisine ilişkin analizde alınan göç ve verilen göç ayrı ayrı ele alınmıştır. Alınan göçte ve verilen göçte en önemli olan değişkenlerin bulunması amaçlanmıştır. Ayrıca verilen göç ve alınan göçün tahmini tüm değişkenler ve değişken önemi sonrası en önemli olan değişkenler dikkate alınarak yapılmış ve performansları karşılaştırılmıştır. Çalışmada; Caret paketi ile çeşitli modelleme teknikleri, Boruta algoritması, Özyinelemeli Özellik Eleme (Recursive Feature Elimination) algoritması ve Ortak önem skoru kullanılarak değişken önemi belirleme işlemi gerçekleştirilmiştir. Daha sonra tüm değişkenler ve en önemli değişkenler dikkate alınarak; basit doğrusal regresyon (linear regression), Destek Vektör Regresyonu (Support Vector Regression), Karar Ağaçları (Decision Trees), Rassal Ormanlar (Random Forests), Güçlendirme (Boosting), Gradyan Artırma Makinesi (Gradient Boosting Machine), Çantalama (Bagging) ve Yığınlama (Stacking) algoritmaları ile tahmin yapılmıştır.

Tüm değişkenleri kullanarak oluşturulan model, maksimum bilgiye sahip olduğundan genellikle en yüksek doğruluğa sahip olabilir. Ancak, aşırı öğrenme (overfitting) riski de taşıyabilir ve bu model, yeni verilerde daha düşük performans gösterebilir. Sadece en önemli değişkenlerle oluşturulan model, daha basit ve yorumlanabilir olmaktadır. Bu model, genellikle daha az karmaşıktır ve overfitting riski daha düşüktür. Önemli değişkenlerle oluşturulan model, özellikle iş dünyası ve akademik analizlerde daha yorumlanabilir olmaktadır. Böylece karar vericiler veya paydaşlar, modelin nasıl çalıştığını ve hangi değişkenlerin sonucu

nasıl etkilediğini daha iyi anlayabilirler. Tüm değişkenleri kullanmak yerine sadece önemli değişkenleri kullanmak, veri toplama ve işleme maliyetini düşürmektedir. Özellikle büyük veri setlerinde bu, işlem süresini ve hesaplama maliyetlerini önemli ölçüde azaltmaktadır. Önemli değişkenleri kullanarak model kurmak, bazen genelleme yeteneğini ve modelin performansını artırabilir. Eğer önemli değişkenlerle oluşturulan modelin performansı kabul edilebilir seviyedeysse, bu modeli tercih etmek genellikle daha pratik olur.

DEĞİŞKEN ÖNEMİ

Değişken önemi (öznitelik önemi-feature importance), her bir açıklayıcı değişkeninin bir makine öğrenimi modelinin çıktısı üzerindeki etkisini değerlendirmek için kullanılan bir ölçüttür. Çıktı (hedef-bağımlı) değişkeni tahmin etmede hangi özniteliklerin en etkili olduğunu gösterir, araştırmacıların değişken ilişkilerini anlamalarına, model doğruluğunu iyileştirmelerine ve yorumlanabilirliği geliştirmelerine yardımcı olur (Breiman, 2001, s.11; Chen ve Guestrin, 2016, s.785).

Öznitelik önemi, hangi değişkenlerin tahminleri en güçlü şekilde etkilediğini ortaya çıkararak model yorumlanabilirliğini iyileştirmeye yardımcı olur ve araştırmacıların ve uygulayıcıların en etkili özelliklere odaklanmasını sağlar (Guyon ve Elisseeff, 2003, s.1157). Öznitelik önem puanları, özellikle finans, sağlık ve çevre bilimi gibi özellik etkilerini anlamanın önemli olduğu alanlarda alan anlayışını ve karar vermeyi destekleyebilecek içgörüler sağlar (Cheng ve diğerleri, 2011).

Ancak, öznitelik öneminin sınırlamaları vardır. Derin sinir ağları gibi karmaşık modellerde, öznitelik önem puanlarını yorumlamak zor olabilir, çünkü öznitelikler arasındaki doğrusal olmayan etkileşimler yeterince yakalanamayabilmektedir (Ribeiro ve diğerleri, 2016, s.1135). Ek olarak, ağaç tabanlı modeller gibi belirli yöntemlerden elde edilen önem puanları, daha fazla seviyeye veya daha yüksek varyansa sahip özniteliklere doğru önyargılı olabilir ve bu da potansiyel olarak yanıltıcı sonuçlara yol açabilmektedir (Strobl vd., 2007, s.2-3).

Çalışmada; literatürde sıklıkla kullanılan aynı zamanda iyi performansa sahip olan Caret paketi ile çeşitli modelleme teknikleri, Boruta algoritması, Özyinelemeli Özellik Eleme (Recursive Feature Elimination) algoritması ve Ortak önem skoru kullanılarak değişken önemi belirleme işlemi gerçekleştirilmiştir.

Boruta Algoritması

Boruta algoritması şu adımlardan oluşur:

1. Tüm değişkenlerin kopyalarını ekleyerek bilgi sistemini genişletin (bilgi sistemi, orijinal kümedeki öznitelik sayısı 5'ten az olsa bile her zaman en az 5 gölge öznitelik kadar genişletilir).
2. Eklenen öznitelikleri, yanıtla olan korelasyonlarını kaldırmak için karıştırın.
3. Genişletilmiş bilgi sisteminde rastgele orman algoritmasını çalıştırın ve hesaplanan Z puanlarını (Her bir değişken (orijinal ve gölge) için doğruluk kaybının Z puanı hesaplanır. Z puanı, doğruluk kaybının ortalamasının, doğruluk kaybının standart sapmasına bölünmesi ile elde edilir.) toplayın.
4. Gölge öznitelikler arasındaki maksimum Z puanını (MZSA) bulun ve ardından MZSA' dan daha iyi puan alan her öznitelige bir isabet atayın.
5. Belirsiz öneme sahip her öznitelik için MZSA ile iki taraflı bir eşitlik testi gerçekleştirin.
6. MZSA' dan önemli ölçüde daha düşük öneme sahip öznitelikleri 'önemsiz' olarak kabul edin ve bunları bilgi sisteminden kalıcı olarak kaldırın.
7. MZSA' dan önemli ölçüde daha yüksek öneme sahip öznitelikleri 'önemli' olarak kabul edin.
8. Tüm gölge öznitelikleri kaldırın.

9. Tüm özniteliklere önem atanana veya algoritma daha önce belirlenen rastgele orman çalıştırma sınırına ulaşana kadar işlemleri tekrarlayın (Kursa ve Rudnicki, 2010, s.3-4)

Özyinelemeli Öznitelik Eleme (Recursive Feature Elimination-RFE)

RFE algoritması, sınıflandırıcı performansını yinelemeli bir şekilde öznitelikleri eleyerek ölçmektedir. RFE, sınıflandırıcıya tüm öznitelikleri vererek modelin performansını hesaplamaktadır. Sınıflandırıcıda kullanılan her bir özneneliğin önemi veya sıralaması hesaplanır. Daha sonra, özniteliklerin yinelemeli bir şekilde elendiği alt kümeler oluşturulur. Her biriyle birlikte sınıflandırıcı modeli yeniden eğitilir ve performansı hesaplanır. Ayrıca, kümede kullanılan özniteliklerin önem veya sıralama değerleri de hesaplanır. Önem veya sıralama değerleri, kullanılan sınıflandırıcı modeline bağlı olarak bilgi kazanımı yöntemi veya Gini indeksi ile hesaplanabilir. Algoritma son adımda en iyi performansı gösteren alt kümeyi belirler (Ustebay, Turgut & Aydın, 2018, s.73).

Ortak Önem Skoru

Ortak önem skoru hesaplamak, farklı algoritmalarından elde edilen değişken önem skorlarının ortalamasını veya birleşimini kullanarak yapılabilmektedir. İlk olarak çeşitli makine öğrenmesi modelleri eğitilir. Her model, değişkenler için bir önem skoru üretir. Birden fazla makine öğrenmesi modelinden elde edilen değişken önem skorlarının ortalaması alınarak, ortak önem skoru hesaplanmış olur. Elde edilen ortak skorlar kullanılarak değişkenler sıralanır. En yüksek skoru alan değişkenler en önemli olarak değerlendirilir. Ortak önem skoru sonuçları, daha az önemli değişkenleri çıkararak modelin karmaşıklığını azaltmak veya daha anlamlı analizler yapmak için kullanılabilmektedir. Yapılan çalışmada caret paketindeki rpart, lm, rf, gbm ve xgbtree algoritmaları kullanılmış ayrıca 10 ve üzeri ortak önem skoruna sahip değişkenler dikkate alınmıştır.

TAHMİN ALGORİTMALARI

Makine öğrenmesi, sistemin geçmişteki deneyimlerinden elde edilen öğrenmelerini kullanarak bir model oluşturmasına ve gelecekteki karşılaşacağı durumlar karşısında bir tahminde bulunmasını sağlar. Örneklerden öğrenen sistemlerde değişik öğrenme stratejileri kullanılmaktadır. Öğrenmeyi gerçekleştirecek olan sistem ve kullanılan öğrenme algoritması bu stratejilere bağlı olarak değişmektedir (Öztemel, 2003, s.25). Genel olarak öğrenme stratejileri; denetimli, denetimsiz ve pekiştirmeli öğrenme olmak üzere üç grupta incelenmektedir.

Denetimli Öğrenme (Supervised Learning): Denetimli öğrenme, belirli bir çıktı için etiketlenmiş girdi verileri üzerinde bir bilgisayar algoritmasının eğitildiği yapay zekâ oluşturmaya yönelik bir yaklaşımdır. Model, girdi verileri ile çıktı etiketleri arasındaki temel örüntüleri ve ilişkileri algılayana kadar eğitilir ve daha önce hiç görülmemiş verilerle sunulduğunda doğru etiketleme sonuçları vermesine olanak tanır. Denetimli makine öğrenmesi regresyon ve sınıflandırma olarak iki alt grupta ele alınmaktadır. Regresyon, nicel değişkenlerin tahmin ve kestirimi için kullanılır. Sınıflandırma ise nitel değişkenlerin modellenmesi ve tahmin edilmesi için nitel kategorilere gözlemler atama işlemidir (Akay, 2018). Yapılan çalışma regresyon problemidir.

Yapılan çalışmada denetimli öğrenme algoritmalarından; Basit doğrusal regresyon (Linear Regression), Destek Vektör Regresyonu (Support Vector Regression), Karar Ağaçları (Decision Trees), Rassal Ormanlar (Random Forests), Güçlendirme (Boosting), Gradyan Artırma Makinesi (Gradient Boosting Machine), Çanta-lama (Bagging) ve Yığınlama (Stacking) algoritmaları ile tahmin yapılmıştır.

Basit Doğrusal Regresyon (Linear Regression)

Regresyon, değişkenler arasındaki ilişkiyi belirlemek için kullanılan bir istatistiksel analiz yöntemidir. İlişki, bağımlı ve bağımsız değişkenler arasında belirlenmektedir Doğrusal regresyon basit doğrusal veya çoklu doğrusal regresyon olabilmektedir (Kavitha, Varuna ve Ramya, 2016, s.3).

Çoklu Doğrusal Regresyon: birden fazla bağımsız veya öngörücü değişkenle yapılan tahmin sürecidir (Kavitha, Varuna ve Ramya, 2016, s.3). Çoklu doğrusal regresyon denklemi aşağıda verilmiştir (Maulud ve Abdulazeez, 2020, s.141).

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + \varepsilon \quad (1)$$

y (Bağımlı Değişken): Modelin açıklamaya veya tahmin etmeye çalıştığı değişkendir.

$x_1, x_2, \dots, x_m$ (Bağımsız Değişkenler): y 'yi açıklamak veya tahmin etmek için kullanılan birden fazla açıklayıcı değişkenlerdir.

β_0 (Intercept/Kesim Noktası): Tüm x_i 'ler sıfır olduğunda y 'nin model tarafından tahmin edilen değeridir.

$\beta_1, \beta_2, \dots, \beta_m$ (Katsayılar): Her bir bağımsız değişkenin ($x_1, x_2, \dots, x_m$) y üzerindeki etkisini ifade eder.

ε (Hata Terimi): Gerçek y değerleri ile modelin tahmin ettiği y değerleri arasındaki farkı temsil eder. Hata terimi modele dahil edilemeyen etkileri, ölçüm hatalarını ve rastgele sapmaları içerir.

Destek Vektör Makinesi (Support Vector Machine-SVM)

Vapnik (Vapnik, 1995) tarafından geliştirilen SVM sınıflandırmasının birincil amacı, iki sınıfı ayıracak çizgiyi (hiper düzlemi) belirlemektir. Destek vektörleri, sınıfları ayıran çizgiyi (hiper düzlemi) belirlemek için en yakın mesafede bulunan veri noktalarıdır. Bu noktalar, sınıflandırma kararını veren ve modelin başarısını etkileyen kritik öneme sahiptir. Amaç genelleme hatasını en aza indirmek ve böylece marjı (sınıflar arasındaki en yakın veri noktalarına (destek vektörlerine) olan mesafeyi ifade etmektedir) en üst düzeye çıkarmak için iki sınıf arasında optimum bir ayırıcı hiper düzlem oluşturmaktır. Herhangi iki sınıf sonsuz sayıda hiper düzlemlle ayrılabilir, SVM genelleme hatasını (yani, görülmeyen test desenleri için hatayı) en aza indiren hiper düzlemi belirler. Verilerin doğrusal olarak ayrılamadığı durumlarda, doğrusal olanlar yerine doğrusal olmayan sınıflandırıcılar kullanılabilir. Verilerin SVM kullanılarak doğrusal olarak ayırt edilemediği durumlarda, verileri daha büyük boyutlu uzaylara taşıyarak analiz etmek için çekirdek fonksiyonları kullanılır. Çekirdek fonksiyonunun seçimi çok önemlidir ve farklı çekirdek fonksiyonları seçilerek farklı performans sonuçları elde edilebilir (Deka, 2014: 374). Sık kullanılan çekirdek (kernel) fonksiyonları aşağıda sunulmuştur (Hussain ve ark., 2011, s.147):

$$\text{Doğrusal çekirdek fonksiyon} : K(x_i, x_j) = (x_i^T x_j) \quad (2)$$

$$d - \text{dereceli polinom çekirdek fonksiyon} : K(x_i, x_j) = (\gamma x_i^T x_j + 1)^d \quad (3)$$

$$\text{Radyal tabanlı çekirdek fonksiyon} : K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \quad (4)$$

$$\text{Sigmoid çekirdek fonksiyon} : K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r) \quad (5)$$

SVM regresyonu (SVR), SVM algoritması ile benzer bir prensibe dayanır ancak sınıflandırmadan farklı olarak sürekli bir hedef değişkenin tahmin edilmesini amacıyla uyarlanmış bir versiyonudur. Drucker vd. (1997) tarafından önerilen SVR, temel olarak, hedef değişkenin tahminini yaparken belirli bir hata payı içinde kalmayı amaçlar. Bu amaçla, epsilon-tube kayıp fonksiyonu kullanılır; yani, modelin tahmin ettiği değer ile gerçek değer arasındaki farkın epsilon değerini aşmadığı sürece hata sayılmaz. Bu yaklaşım, modelin aşırı uyum yapmasını engeller ve genelleme hatasını minimize eder. SVR, çekirdek (kernel) fonksiyonları kulla-

arak verileri daha yüksek boyutlu bir uzaya taşıyarak burada doğrusal regresyon yapar. Bu sayede, doğrusal olmayan ilişkileri de modelleyebilir. SVR'nin başarısı, doğru epsilon değeri ve uygun çekirdek fonksiyonunun seçilmesine bağlıdır ve bu seçimler modelin performansını doğrudan etkiler.

Karar Ağaçları (Decision Tree)

En basit haliyle, karar ağacı analizi "böl ve yönet" yaklaşımına dayalı bir sınıflandırma ve regresyon yaklaşımıdır. Mitchell (1997), karar ağacı öğrenimini, öğrenilen işlevin bir karar ağacıyla temsil edildiği ayrık değerli hedef işlevler için bir tahmin yöntemi olarak tanımlamıştır. Karar ağacı öğrenimi, tümevarımsal çıkarım için en yaygın kullanılan ve pratik yöntemlerden biridir. Bir karar ağacı, eğitim setinin özellik alanını yinelemeli olarak bölerek oluşturulur. Amaç, sağlam ve bilgilendirici bir hiyerarşik sınıflandırma modeli geliştirmek için özellik alanını bölen bir dizi karar kuralı belirlemektir. Bir karar kuralının seçilmesinin ardından, özellik alanı iki ayrı alt alana bölünür. Daha sonra, bölümleri temsil eden bir karar ağacı, oluşturulan alt alanların her birine, hepsi tek bir sınıfın örneklerini içerece kadar yinelemeli olarak bölümlere uygulanarak üretilir (Myles ve diğerleri, 2004, s.276). Regresyon problemlerinde ise, benzer bir şekilde, her karar kuralı, veri kümesindeki örnekleri belirli bir değeri tahmin etmeye yönelik olarak iki alt küme halinde bölmektedir. Bu durumda, hedef değişkenin sürekli bir değeri olduğundan, her bir alt kümede, örneklerin ortalama hedef değerine yakınlık sağlanmaktadır. Yineleme süreci, her bir alt kümenin daha homojen hale gelene kadar devam etmektedir ve her bir yaprak düğümde, o alt kümedeki örneklerin ortalama hedef değeri tahmin edilmektedir.

Topluluk Öğrenmesi (Ensemble Learning)

Klasik makine öğrenmesi algoritmaları tek bir algoritmayı kullanarak tahmin yapmaya çalışırken topluluk öğrenmesi algoritmaları ise birden fazla algoritmayı kullanarak tahmin yapmaya çalışır. Birden fazla algortmadan alınan değerler birleştirilip sonuç elde edilmektedir (Bilgin, 2018, s.135). Çalışmada topluluk öğrenmesi yöntemlerinden iyi performansa sahip sıklıkla kullanılan Rassal Ormanlar, Güçlendirme, Gradyan Artırma Makinesi, Çantalama ve Yığınlama algoritmaları kullanılmıştır.

Rasgele Orman (Random Forest-RF): Rastgele orman, budanmamış sınıflandırma veya regresyon ağaçlarının bir topluluğudur. Bagging'in (Breiman 2001) gelişmiş bir versiyonudur ve buna rastgelelik eklenmiştir. Rastgele orman birçok sınıflandırma ağacı üretir ve her ağaç, bir ağaç sınıflandırma algoritması kullanılarak orijinal verilerden farklı bir önyükleme (bootstrap) örneği ile oluşturulur. Orman oluşturulduktan sonra, sınıflandırılması gereken yeni bir nesne, sınıflandırma için ormandaki her ağaca yerleştirilir. Her ağaç, ağacın nesnenin sınıfı hakkındaki kararını gösteren bir oy verir. Modellerin sonuçları sınıflandırma probleminde oylama, regresyon probleminde ortalaması alınarak birleştirilip tek bir çıktı elde edilmektedir (Hasan, Nasser, Ahmad ve Molla, 2016, s.133).

GBM (Gradyan Artırma Makinesi- Gradient Boosting Machines): GBM, Friedman (2001) tarafından önerilen birden fazla zayıf öğrenicinin (genellikle karar ağaçlarının) ardışık olarak eğitilmesiyle güçlü bir model oluşturan bir topluluk öğrenme yöntemidir. Bu algortmada her yeni model, önceki modelin yaptığı hataları azaltmaya odaklanır. İlk olarak model veri setine uygulanır ve tahmin hataları hesaplanır. Daha sonra, bu hataları minimize etmek için kayıp fonksiyonunun negatif gradyanı yönünde yeni bir model oluşturulur ve süreç belirli bir iterasyon boyunca devam eder. Gradyan artırma yöntemi, tahmin hatalarını aşamalı olarak azaltarak modelin doğruluğunu artırır ve özellikle yüksek boyutlu, karmaşık veri setlerinde etkili sonuçlar vermektedir (Friedman, 2001, s.1214-1215).

Güçlendirme (Boosting): Boosting, Schapire (1990) tarafından önerilen algortmada ilk olarak, her veri noktasına eşit bir ağırlık atanır ve bu bilgi modele beslenerek model eğitilir ve yanlış sınıflandırılmış veri noktaları belirlenir. Daha sonra model, yanlış sınıflandırılmış veri noktalarının ağırlığı artırılarak eğitilir.

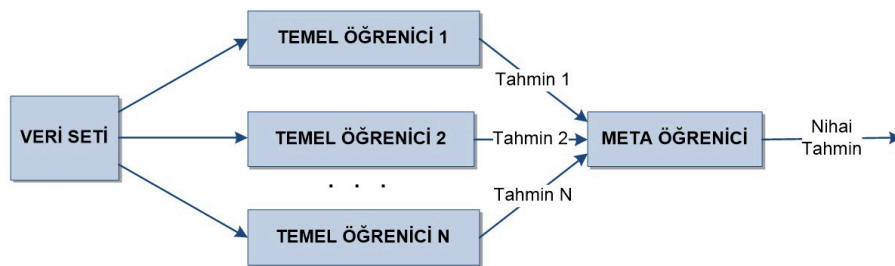
İkinci model ağırlıklandırılmış veriler üzerinden işlem yapmaktadır. Yanlış sınıflandırılmış veriler, ağırlıkları artırılmış olarak sonraki modele iletilmektedir. Modelin performansı, önceki modelde yanlış sınıflandırılmış örnekler daha yüksek bir ağırlık atayarak iyileştirilebilir (James ve diğerleri, 2013, s.321). Bu algorithmada, her bir model genellikle zayıf öğrenci olarak kabul edilir. Boosting yöntemi, zayıf öğrencileri (weak learner) birleştirerek daha güçlü, daha doğru bir model (yani güçlü öğrenci (strong learner)) oluşturmaktadır. Bu algorithmada ağaçlar sırayla büyütülür: her ağaç, daha önce büyütülmüş ağaçlardan gelen bilgiler kullanılarak büyütülür. Güçlendirme önyüklemeye örneklemesini içermez; bunun yerine her ağaç, orijinal veri kümesinin değiştirilmiş bir sürümüne uydurulur (James ve diğerleri, 2013, s.321).

Çantalama (Bagging): Bagging, Breiman (1996) tarafından önerilen sınıflandırma ve regresyon problemlerinde tahminlerin doğruluğunu artırmak için geliştirilmiş bir tekniktir. Belirli bir modelin, orijinal veri kümesinden rastgele seçilen alt kümelerine uygulanması ve sonuçların birleştirilmesine dayanır (Bilgin, 2018). Bagging algoritması, Bootstrap Aggregating olarak da bilinir, n örnek içeren bir eğitim setinden, bir eğitim setini n örnekle rastgele değiştirerek türetilir. Bazı örnekler yeni eğitim setinden hariç tutulurken, diğerleri birden fazla kez gerçekleşebilir. Daha sonra, modeller çeşitli örneklerden oluşan eğitim setleriyle eğitilir ve modellerin sonuçları, sınıflandırma problemindeki oylama ve regresyon problemindeki ortalama hesaplanarak birleştirilir ve tekil bir çıktı elde edilir (Liaw & Wiener, 2002, s.18).

Yığınlama (Stacking): Stacking algoritması, farklı sınıflandırıcı modeller kullanılarak oluşturulan heterojen bir topluluk öğrenme yöntemidir. Son zamanlarda geliştirilmiş olmasına rağmen, daha iyi performans elde etmek için hangi modellerin kullanılacağına dair kabul görmüş bir yol olmadığından, yani kullanılacak modeller deneme yanılma yoluyla belirlendiğinden, Bagging ve Boosting yöntemlerinden daha az kullanılır (Witten ve diğerleri, 2005, s.369).

Figür 1 incelendiğinde, yığınlama algoritmasının iki seviyeli bir algoritma olduğu görülmektedir. İlk seviyede ikiden fazla temel öğrenen eğitilir, ardından ilk seviye öğrenenlerin tahminleri ikinci seviye, yani meta seviye sınıflandırıcı tarafından birleştirilir. Bu yöntem, aşırı uyum olasılığını azaltmaktadır (Salunkhe ve Mali, 2016, s.727-728). En iyi performansı sağlayan algoritmalar konusunda bir fikir birliği yoktur, bu nedenle herhangi bir öğrenme planı teoride uygulanabilir. Ancak, stacking algoritmasını tanıtan Wolpert (1992), basit doğrusal modellerin veya yapraklarında doğrusal modellerin yer aldığı ağaçların ikinci seviye bir sınıflandırıcı olarak iyi performans gösterdiğini öne sürmüştür (Witten vd., 2005, s.372). Breiman (1996), Wolpert (1992) tarafından tanıtılan stacking genelleme kavramını sayısal tahmine uyarladı. Ting ve Witten (1997) çeşitli seviye 2 (meta-öğrenen) modellerini deneysel olarak karşılaştırdılar ve basit doğrusal modelin en iyi performansı gösterdiğini buldular (Witten vd., 2005, s.372). Yapılan çalışmada, stacking algoritmasında kullanılan modeller Tablo 1’de verilmiştir.

Figür 1
Stacking Algoritması



Tablo 1*Stacking Algoritmasında Kullanılan Modeller*

LM, RPART, SVM	RPART, RF, XGBOOST
LM, RPART, RF	RPART, RF, GBM
LM, RPART, XGBTREE	RPART, RF, BAGGING
LM, RPART, GBM	RPART, XGBOOST, GBM
LM, RPART, BAGGING	RPART, XGBOOST, BAGGING
LM, SVM, RF	SVM, RF, XGBOOST
LM, SVM, XGBOOST	SVM, RF, GBM
LM, SVM, GBM	SVM, RF, BAGGING
LM, SVM, BAGGING	SVM, XGBOOST, GBM
LM, RF, XGBOOST	SVM, XGBOOST, BAGGING
LM, RF, GBM	SVM, GBM, BAGGING
LM, RF, BAGGING	RF, XGBOOST, GBM
RPART, SVM, RF	RF, XGBOOST, BAGGING
RPART, SVM, XGBOOST	RF, GBM, BAGGING
RPART, SVM, GBM	XGBOOST, GBM, BAGGING
RPART, SVM, BAGGING	

PERFORMANS METRİKLERİ

Model performansını değerlendirmek için farklı ölçütler kullanılmaktadır. Bu ölçütler ile farklı modellerin performanslarını kıyaslanabilmektedir. Böylece veri setine en uygun modelinin hangisi olduğu belirlenir. Regresyon modellerinde model performansını değerlendirmek için Ortalama karesel hata (Mean Squared Error- MSE), Ortalama Mutlak Hata (Mean Absolute Error- MAE) ve Hata kareler ortalamasının karekökü (Root Mean square error- RMSE) yaygın olarak kullanılan üç değerlendirme metriğidir. Performans metriklerinin denklemleri aşağıda verilmiştir.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

n : Gözlem sayısını ifade eder.

y_i : Gerçek değerleri ifade eder.

$\hat{y}_i$: Tahmin edilen değerleri ifade eder.

YAZILIM

Bu çalışmada R 3.6.1 tabanlı çalışan RStudio 1.4 versiyonu ile uyumlu paketler arasından seçim yapılarak algoritmalar uygulanmıştır. R programında; Boruta, rpart, randomForest, gbm, plyr, xgboost ve caret paketleri kullanılarak değişken önemi belirleme işlemi gerçekleştirilmiştir. Boruta paketinde boruta algoritması ile değişken önemi belirlenmiş. Caret paketi, çeşitli modelleme teknikleri ile değişken önemini belirlemek ve rfe algoritması ile değişken önemini belirlemek için kullanılır. Caret paketinde çeşitli modelleme teknikleri



ile değişken önemi “varImp” fonksiyonu kullanılarak hesaplanmıştır. Caret paketinde kullanılan modelleme teknikleri: rf (Random Forest) (gerekli paketler: randomForest), lm (Linear Regression), rpart (CART) (gerekli paketler: rpart), gbm (Stochastic Gradient Boosting) (gerekli paketler: gbm, plyr) ve xgbTree (eXtreme Gradient Boosting) (gerekli paketler: xgboost, plyr). Ayrıca; caret paketindeki farklı modellerin değişken önem skorlarının ortalaması alınarak da en önemli değişkenler belirlenmiştir.

Tahmin aşamasında; rpart, randomForest, e1071, xgboost, gbm, ipred ve caretEnsemble paketleri kullanılmıştır. Rpart paketinde karar ağaçları ile, randomForest paketinde rasgele orman algoritması ile, e1071 paketinde destek vektör regresyonu ile, xgboost paketinde xgboost algoritması ile, gbm paketinde gbm algoritması ile, ipred paketi ile Bagging algoritması ile ve caretEnsemble paketi ile Stacking algoritması ile tahmin yapılmıştır. Stacking algoritmasında kullanılan algoritmalar Stacking kısmında değinilmiştir. Tahmin aşamasında hiper parametrelerin varsayılan değerleri kullanılmıştır.

VERİ SETİ

Bu çalışmada, 2018-2022 yılları arasında Türkiye’de il bazında erişilebilen ve literatürde sıklıkla kullanılan değişkenler kullanılarak analizler gerçekleştirilmiştir. Yapılan çalışmada, alınan ve verilen göçe ilişkin en önemli değişkenleri belirlemek ve onlara ilişkin göç tahmini yapabilmek için ayrı ayrı bağımlı değişken olarak alınmıştır. Her iki bağımlı değişken için bağımsız değişkenler aynıdır. Kullanılan bağımsız değişkenler: akademisyen sayısı, hastane sayısı, özel hastane sayısı, GSYİH, öğretmen sayısı, okul sayısı, tarım alanı, nüfus yoğunluğu, nüfus artış hızı, muhtemel eğitim süresi, ortalama eğitim süresi, trafik kaza sayısı, otomobil sayısı, traktör sayısı, bitirilen eğitim durumu, ithalat, ihracat, hekim sayısı, hastane yatak sayısı ve konut satış sayısı değişkenidir. İstihdam il bazında 2018-2022 yılları arasında bulunamadığından, terör verisine de sadece 2020 yılına kadar erişilebildiğinden ve ceza infaz kurumuna ilişkin verilere 2020 yılından sonra il bazında erişilemediğinden bu değişkenler analize dahil edilmek istendiği halde dahil edilememiştir. Değişkenlere ilişkin tanımlama ve veri kaynakları **Tablo 2**’de verilmiştir. Veri setindeki değişkenler farklı ölçeklerde olduğundan standardizasyon işlemi yapılmıştır. R programında scale fonksiyonu ile standardizasyon işlemi gerçekleştirilmiştir.

Tablo 2

Kullanılan Değişkenler, Tanımları ve Veri Kaynakları

Değişken	Tanımı	Veri Kaynağı
Akademisyen sayısı	Toplam akademisyen sayısı	Yükseköğretim Bilgi Yönetim Sistemi
Hastane sayısı	Toplam hastane sayısı	TUİK
Özel hastane sayısı	Toplam özel hastane sayısı	TUİK
GSYİH	Gayrisafi yurt içi hasıla	TUİK
Öğretmen sayısı	Toplam öğretmen sayısı	MEB
Okul sayısı	Toplam okul sayısı	MEB
Tarım alanı	Tahıllar ve diğer bitkisel ürünlerin alanı/Sebze bahçeleri alanı/ Meyveler, içecek ve baharat bitkileri alanı/Süs bitkileri alanı toplamı	TUİK
Nüfus yoğunluğu	Nüfus Yoğunluğu (Kişi/km ²)	TUİK
Nüfus artış hızı	Nüfus Artış Hızı (Binde)	TUİK
Muhtemel eğitim süresi	Muhtemel eğitim süresi, ISCED 1-8 (ilkokul- yükseköğretim)	TUİK
Ortalama eğitim süresi	Ortalama eğitim süresi	TUİK
Trafik kaza sayısı	Toplam trafik kaza sayısı	TUİK
Otomobil sayısı	Toplam otomobil sayısı	TUİK

Değişken	Tanımı	Veri Kaynağı
Traktör sayısı	Toplam traktör sayısı	TÜİK
Bitirilen eğitim durumu	Eğitim durumuna göre nüfus (15 yaş ve üzeri): Yüksekokul veya fakülte mezunu + Yüksek lisans mezunu + Doktora mezunu	TÜİK
İthalat	İthalat (Değer: Bin ABD \$)	TÜİK
İhracat	İhracat (Değer: Bin ABD \$)	TÜİK
Hekim sayısı	Toplam hekim sayısı	TÜİK
Hastane yatak sayısı	Toplam hastane yatak sayısı	TÜİK
Konut satış sayısı	Toplam konut satış sayısı	TÜİK
Alınan göç	Toplam alınan göç (Bağımlı Değişken)	TÜİK
Verilen göç	Toplam verilen göç (Bağımlı Değişken)	TÜİK

BULGULAR

Çalışmada, alınan göç ve verilen göç için ayrı ayrı farklı metotlarla değişken önemi hesaplanmış ardından tüm değişkenler ve değişken önemi sonrası seçilen değişkenlerle göç tahmini yapıp performansları kıyaslanmıştır.

Değişken Öneme İlişkin Sonuçlar

Çalışmada; Caret paketi ile çeşitli modelleme teknikleri, Boruta algoritması, özinelemeli özellik eleme algoritması ve ortak önem skoru kullanılarak elde edilen değişken önem sıralamaları bu kısımda verilmiştir.

Caret Paketine İlişkin Sonuçlar

Caret paketinde rf, lm, rpart, gbm ve xgbTree algoritmaları ile elde edilen değişken önemleri alınan ve verilen göç bağımlı değişkeni için [Tablo 3](#) ve [Tablo 4](#)’te verilmiştir.

[Tablo 3](#) incelendiğinde, alınan göç bağımlı değişkeni için farklı modellerle elde edilen değişken önemleri sırasıyla verilmektedir.

Tablo 3

Alınan Göç Bağımlı Değişkeni için Caret Paketinde Farklı Modeller ile Değişken Önemi

Yöntem	Değişkenler
rf değişken önemi	trafik_kaza_sayisi, ogretmen_sayisi, konut_satis_sayisi, ozel_hastane_sayisi, hastane_sayisi, hekim_sayisi, otomobil_sayisi, akademisyen_sayisi, okul_sayisi, hastane_yatak_sayisi, ithalat, ihracat, nufus_yogunlugu, tarim_alani, nufus_artis_hizi, traktor_sayisi, bitirilen_egitim_durumu, muhtemel_egitim_suresi, ortalama_egitim_suresi, GSYİH
lm değişken önemi	trafik_kaza_sayisi, nufus_artis_hizi, ithalat, hastane_sayisi, nufus_yogunlugu, okul_sayisi, akademisyen_sayisi, GSYİH, konut_satis_sayisi, hastane_yatak_sayisi, ihracat, tarim_alani, ortalama_egitim_suresi, hekim_sayisi, traktor_sayisi, bitirilen_egitim_durumu, muhtemel_egitim_suresi, otomobil_sayisi, ogretmen_sayisi, ozel_hastane_sayisi
rpart değişken önemi	trafik_kaza_sayisi, konut_satis_sayisi, ozel_hastane_sayisi, otomobil_sayisi, ogretmen_sayisi, hekim_sayisi, akademisyen_sayisi, bitirilen_egitim_durumu, ihracat, hastane_yatak_sayisi, GSYİH, tarim_alani, nufus_yogunlugu, nufus_artis_hizi, muhtemel_egitim_suresi,

Yöntem	Değişkenler
	ortalama_egitim_suresi, ithalat, hastane_sayisi, okul_sayisi, traktor_sayisi
gbm değişken önemi	akademisyen_sayisi, ithalat, trafik_kaza_sayisi, hastane_sayisi, özel_hastane_sayisi, otomobil_sayisi, bitirilen_egitim_durumu, konut_satis_sayisi, nufus_yogunlugu, okul_sayisi, ihracat, ogretmen_sayisi, hastane_yatak_sayisi, traktor_sayisi, nufus_artis_hizi, hekim_sayisi muhtemel_egitim_suresi, ortalama_egitim_suresi, tarim_alani, GSYİH
xgbTree değişken önemi	hastane_sayisi, trafik_kaza_sayisi, ogretmen_sayisi, akademisyen_sayisi, konut_satis_sayisi, özel_hastane_sayisi, okul_sayisi, hastane_yatak_sayisi, otomobil_sayisi, nufus_artis_hizi, tarim_alani, ithalat, hekim_sayisi, muhtemel_egitim_suresi, GSYİH, traktor_sayisi, ihracat, ortalama_egitim_suresi, nufus_yogunlugu, bitirilen_egitim_durumu

Tablo 4 incelendiğinde, verilen göç bağımlı değişkeni için farklı modellerle elde edilen değişken önemleri sırasıyla verilmektedir.

Tablo 4

Verilen Göç Bağımlı Değişkeni için Caret Paketinde Farklı Modeller ile Değişken Önemi

Yöntem	Değişkenler
rf değişken önemi	trafik_kaza_sayisi, ogretmen_sayisi, otomobil_sayisi, konut_satis_sayisi, hastane_sayisi, okul_sayisi, hastane_yatak_sayisi, hekim_sayisi, özel_hastane_sayisi, ithalat, akademisyen_sayisi, nufus_yogunlugu, ihracat, muhtemel_egitim_suresi, tarim_alani, bitirilen_egitim_durumu, traktor_sayisi, ortalama_egitim_suresi, GSYİH, nufus_artis_hizi
lm değişken önemi	trafik_kaza_sayisi, ithalat, ihracat, hekim_sayisi, otomobil_sayisi, nufus_artis_hizi, ogretmen_sayisi, akademisyen_sayisi, konut_satis_sayisi, hastane_sayisi, okul_sayisi, nufus_yogunlugu, traktor_sayisi, özel_hastane_sayisi, muhtemel_egitim_suresi, tarim_alani, ortalama_egitim_suresi, hastane_yatak_sayisi, GSYİH, bitirilen_egitim_durumu
rpart değişken önemi	ogretmen_sayisi, trafik_kaza_sayisi, hastane_sayisi, özel_hastane_sayisi, konut_satis_sayisi, okul_sayisi, hastane_yatak_sayisi, hekim_sayisi, bitirilen_egitim_durumu, GSYİH, nufus_artis_hizi, traktor_sayisi nufus_yogunlugu, akademisyen_sayisi, muhtemel_egitim_suresi, ithalat, ortalama_egitim_suresi, tarim_alani, ihracat, otomobil_sayisi
gbm değişken önemi	ithalat, okul_sayisi, trafik_kaza_sayisi, özel_hastane_sayisi, otomobil_sayisi, akademisyen_sayisi, nufus_yogunlugu, ogretmen_sayisi, bitirilen_egitim_durumu, ihracat, traktor_sayisi, hastane_yatak_sayisi, hastane_sayisi, konut_satis_sayisi, muhtemel_egitim_suresi, ortalama_egitim_suresi, hekim_sayisi, GSYİH, tarim_alani, nufus_artis_hizi
xgbTree değişken önemi	hastane_sayisi, akademisyen_sayisi, trafik_kaza_sayisi, okul_sayisi, ogretmen_sayisi, özel_hastane_sayisi, hastane_yatak_sayisi, konut_satis_sayisi, otomobil_sayisi, ithalat, nufus_artis_hizi muhtemel_egitim_suresi, tarim_alani, GSYİH, traktor_sayisi, nufus_yogunlugu, bitirilen_egitim_durumu, ortalama_egitim_suresi, ihracat, hekim_sayisi

Ortak Önem Skoruna İlişkin Sonuçlar

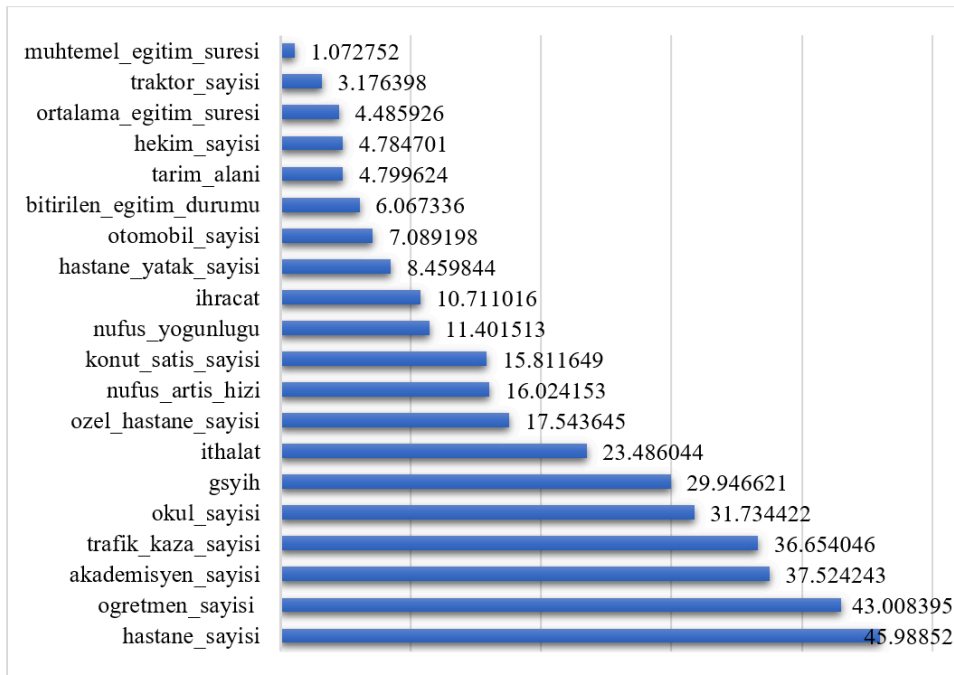
Yapılan çalışmada caret paketindeki rpart, lm, rf, gbm ve xgbtree algoritmaları kullanılarak elde edilen ortak önem skoruna ilişkin değişken önemi değerlerinin sıralaması **Figür 2** ve **Figür 3**’te verilmiştir.

Figür 2 incelendiğinde, alınan göç için on ve üzeri ortak önem skoru değerine sahip değişkenler sırasıyla *hastane sayısı*, *öğretmen sayısı*, *akademisyen sayısı*, *trafik kaza sayısı*, *okul sayısı*, *GSYİH*, *ithalat*, *özel hastane sayısı*, *nüfus artış hızı*, *konut satış sayısı*, *nüfus yoğunluğu* ve *ihracat* olduğu görülmektedir.

Figür 3 incelendiğinde ise verilen göç on ve üzeri ortak önem skoru değerine sahip değişkenler sırasıyla *akademisyen sayısı*, *ithalat*, *trafik kaza sayısı*, *öğretmen sayısı*, *nüfus yoğunluğu*, *hastane sayısı*, *okul sayısı*, *ihracat*, *otomobil sayısı*, *özel hastane sayısı*, *GSYİH*, *hastane yatak sayısı*, *hekim sayısı* ve *tarım alanı* olduğu görülmektedir.

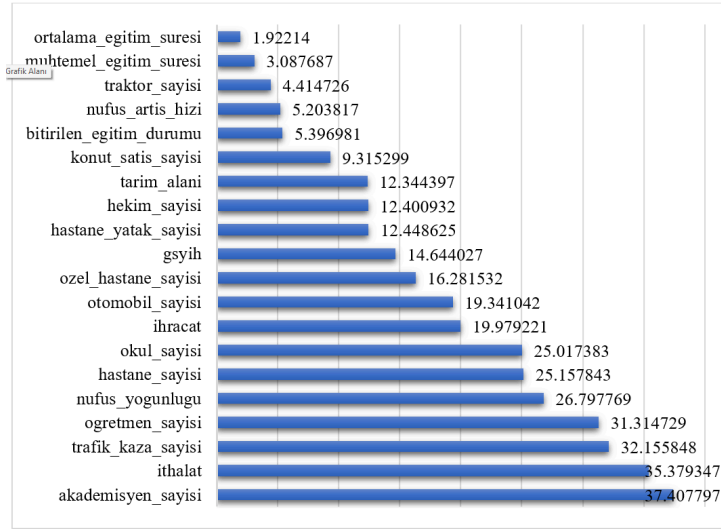
Figür 2

Alınan Göç Bağımlı Değişkeni için Ortak Önem Skoru Değerleri



Figür 3

Verilen Göç Bağımlı Değişkeni için Ortak Önem Skoru Değerler



Boruta Algoritmasına İlişkin Sonuçlar

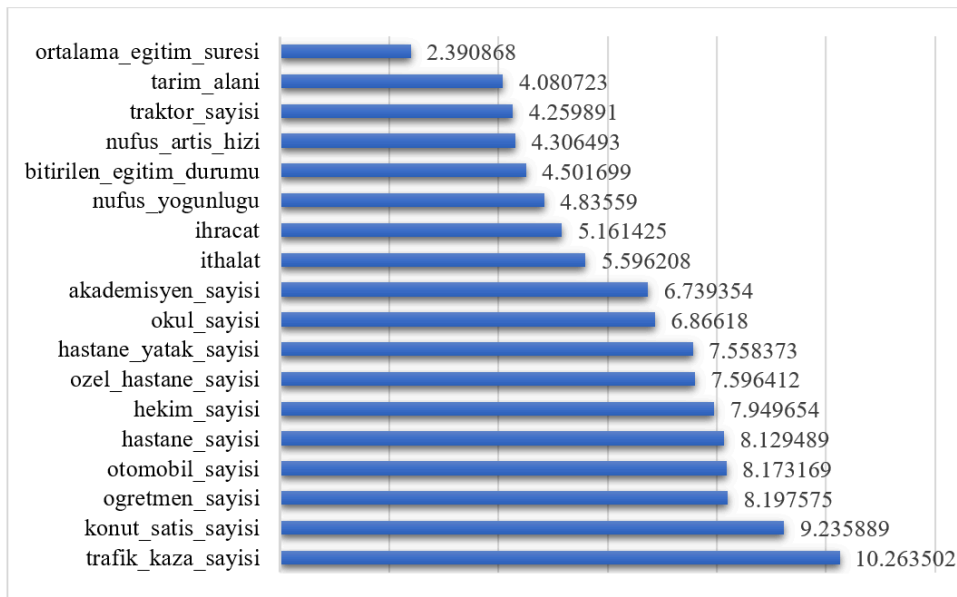
R programında; Boruta paketi kullanılarak Boruta algoritması ile elde edilen değişken önem sıralaması **Figür 4** ve **Figür 5**'te verilmiştir. Yirmi açıklayıcı değişkenden alınan göçte on sekiz tanesi, verilen göçte on yedi tanesi önemli bulunmuştur.

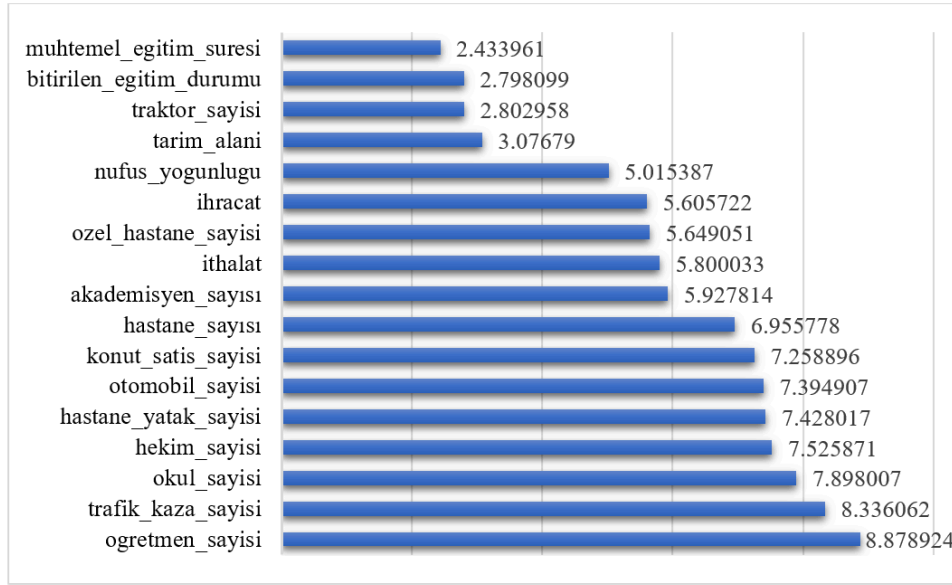
Figür 4 incelendiğinde, alınan göç için en önemli değişkenlerin sırasıyla *trafik kaza sayısı*, *konut satış sayısı*, *öğretmen sayısı*, *otomobil sayısı*, *hastane sayısı*, *hekim sayısı*, *özel hastane sayısı*, *hastane yatak sayısı*, *okul sayısı*, *akademisyen sayısı*, *ithalat*, *ihracat*, *nüfus yoğunluğu*, *bitirilen eğitim durumu*, *nüfus artış hızı*, *traktör sayısı*, *tarım alanı* ve *ortalama öğretim süresi* olduğu görülmektedir.

Figür 5 incelendiğinde ise verilen göç için en önemli değişkenlerin sırasıyla *öğretmen sayısı*, *trafik kaza sayısı*, *okul sayısı*, *hekim sayısı*, *hastane yatak sayısı*, *otomobil sayısı*, *konut satış sayısı*, *hastane sayısı*, *akademisyen sayısı*, *ithalat*, *özel hastane sayısı*, *ihracat*, *nüfus yoğunluğu*, *tarım alanı*, *traktör sayısı*, *bitirilen eğitim durumu* ve *muhtemel eğitim süresi* olduğu görülmektedir.

Figür 4

Alınan Göç Bağımlı Değişkeni için Boruta Değerleri



Figür 5*Verilen Göç Bağımlı Değişkeni için Boruta Değerleri*

Özyinelemeli Özellik Eleme (Recursive Feature Elimination) İlişkin Sonuçlar

RFE algoritması ile elde edilen değişken önem sıralaması **Tablo 5** ve **Tablo 6**’da verilmiştir. Tablodaki sütunlar, modelin farklı sayıda öznitelik ile nasıl performans gösterdiğini karşılaştırmak için kullanılır. Her bir model için RMSE, R^2 , MAE ve bunların standart sapmaları hesaplanarak, hangi özniteliklerin ve hangi sayıda öznitelik seçildiğinde en iyi sonuçları verdiği belirlenmektedir. Yirmi açıklayıcı değişkenden on sekiz tanesi, alınan ve verilen göçte önemli bulunmuştur.

Tablo 5 incelendiğinde sırasıyla en önemli değişkenler: *trafik kaza sayısı, konut satış sayısı, öğretmen sayısı, otomobil sayısı, nüfus artış hızı, hekim sayısı, hastane sayısı, özel hastane sayısı, hastane yatak sayısı, okul sayısı, akademisyen sayısı, GSYİH, traktör sayısı, ithalat, tarım alanı, bitirilen eğitim durumu, nüfus yoğunluğu ve ihracat*’tır.

Tablo 6 incelendiğinde sırasıyla en önemli değişkenler: *öğretmen sayısı, okul sayısı, hastane sayısı, trafik kaza sayısı, hekim sayısı, hastane yatak sayısı, konut satış sayısı, özel hastane sayısı, akademisyen sayısı, otomobil sayısı, nüfus yoğunluğu, nüfus artış hızı, ithalat, ihracat, muhtemel eğitim süresi, tarım alanı, traktör sayısı ve ortalama eğitim süresi*’dir.

Tablo 5*Alınan Göç Bağımlı Değişkeni için RFE ile Önemli Değişkenler ve Performans Değerleri*

Değişken Önem Sıralaması			RFE Performans Değerleri					
Sıralama	Değişken Adı	Öznitelik Sayısı	RMSE	R^2	MAE	RMSESD	R^2SD	MAESD
1	trafik_kaza_sayisi	2	0.1906	0.9561	0.10976	0.1191	0.02907	0.03217
2	konut_satis_sayisi	3	0.1816	0.9613	0.10003	0.1207	0.02752	0.03263
3	ogretmen_sayisi	4	0.1756	0.9628	0.09601	0.1211	0.02822	0.03217
4	otomobil_sayisi	5	0.1738	0.9643	0.09396	0.1234	0.02695	0.03305
5	nufus_artis_hizi	6	0.1755	0.9651	0.09213	0.1217	0.02632	0.03232
6	hekim_sayisi	7	0.1759	0.9651	0.0916	0.1233	0.02612	0.03305
7	hastane_sayisi	8	0.1768	0.9648	0.0917	0.1237	0.02626	0.03247

Değişken Önem Sıralaması			RFE Performans Değerleri					
Sıralama	Değişken Adı	Öznitelik Sayısı	RMSE	R ²	MAE	RMSESD	R ² SD	MAESD
8	ozel_hastane_sayisi	9	0.1761	0.9661	0.09043	0.1272	0.02527	0.03362
9	hastane_yatak_sayisi	10	0.1736	0.9671	0.08893	0.1287	0.02485	0.03382
10	okul_sayisi	12	0.1701	0.9694	0.08574	0.1301	0.02325	0.0337
11	akademisyen_sayisi	14	0.1687	0.9702	0.08445	0.1304	0.02246	0.03411
12	GSYİH	16	0.1699	0.9705	0.08384	0.1336	0.02269	0.0338
13	traktor_sayisi	18	0.167	0.9726	0.081	0.1346	0.01892	0.03318 *
14	ithalat	20	0.1679	0.9725	0.08119	0.1374	0.01931	0.03458
15	tarim_alani							
16	bitirilen_egitim_durumu							
17	nufus_yogunlugu							
18	ihracat							

Tablo 6

Verilen Göç Bağımlı Değişkeni için RFE ile Önemli Değişkenler ve Performans Değerleri

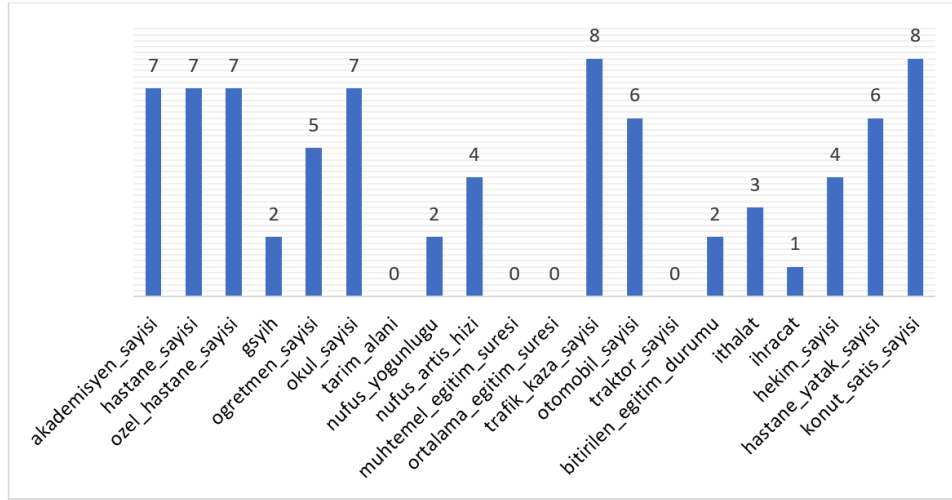
Değişken Önem Sıralaması			RFE Performans Değerleri					
Sıralama	Değişken Adı	Öznitelik Sayısı	RMSE	R ²	MAE	RMSESD	R ² SD	MAESD
1	ogretmen_sayisi	2	0.2181	0.9348	0.10481	0.1745	0.08446	0.04544
2	okul_sayisi	3	0.2069	0.9509	0.09531	0.1705	0.03983	0.03601
3	hastane_sayisi	4	0.2016	0.9556	0.091	0.1701	0.03571	0.03406
4	trafik_kaza_sayisi	5	0.1998	0.9564	0.08972	0.1769	0.03539	0.03429
5	hekim_sayisi	6	0.1942	0.9582	0.08779	0.1717	0.03307	0.03372
6	hastane_yatak_sayisi	7	0.1938	0.9589	0.08688	0.1727	0.03258	0.03357
7	konut_satis_sayisi	8	0.1929	0.9591	0.08662	0.1711	0.03217	0.03328
8	ozel_hastane_sayisi	9	0.193	0.9591	0.08673	0.173	0.03206	0.03348
9	akademisyen_sayisi	10	0.1924	0.9595	0.08642	0.1724	0.032	0.03353
10	otomobil_sayisi	12	0.1932	0.959	0.08668	0.1727	0.03241	0.03354
11	nufus_yogunlugu	14	0.1923	0.9593	0.08667	0.171	0.03241	0.03345
12	nufus_artis_hizi	16	0.1914	0.96	0.0861	0.1728	0.03163	0.03372
13	ithalat	18	0.1902	0.9603	0.08579	0.1735	0.03114	0.03379*
14	ihracat	20	0.1905	0.961	0.08535	0.1754	0.03051	0.03412
15	muhtemel_egitim_suresi							
16	tarim_alani							
17	traktor_sayisi							
18	ortalama_egitim_suresi							

Tüm Yöntemlere İlişkin Genel Sonuçlar

Analizlerde Boruta algoritması, özyinelemeli özellik eleme algoritması, caret paketinde çeşitli modelleme teknikleri ve ortak önem skoru kullanılarak her bir yöntemle elde edilen değişken önemleri Excel programına aktarılmıştır. Daha sonra Figür 6, 7, 8, 9, 10 ve 11’deki grafikler elde edilmiştir.

Figür 6

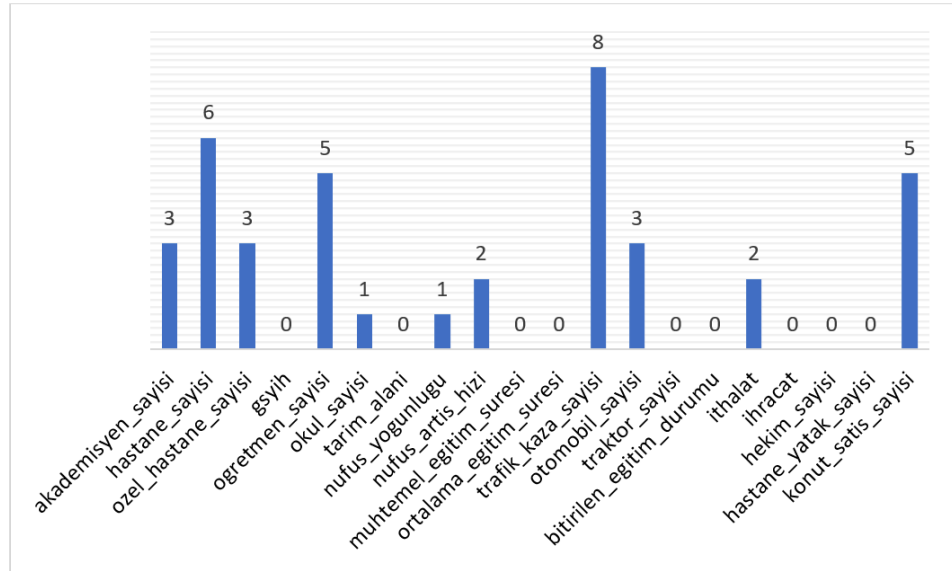
Alınan Göç Bağımlı Değişkeni için Tüm Algoritmalar Dikkate Alınarak İlk Onda Yer Alan Değişkenlerin Sıklıkları



Figür 6’da alınan göç için tüm algoritmalar dikkate alınarak ilk onda yer alan değişkenlerin sıklıkları verilmiştir. Tüm algoritmalar tarafından en önemli olarak seçilen değişkenler (en çok seçilen değişken) (8: Sekiz algoritma tarafından önemli olarak seçilen değişken): *trafik kaza sayısı* ve *konut satış sayısı*. Sırasıyla ilk onda yer alan diğer en önemli değişkenler ise şu şekildedir: *akademisyen sayısı*, *hastane sayısı*, *özel hastane sayısı*, *okul sayısı*, *otomobil sayısı*, *hastane yatak sayısı*, *öğretmen sayısı*, *nüfus artış hızı*, *hekim sayısı*, *ithalat*, *nüfus yoğunluğu*, *GSYİH*, *bitirilen eğitim durumu* ve *ihracat*. İlk onda yer almayan değişkenler ise *tarım alanı*, *ortalama eğitim süresi*, *ihracat*, *muhtemel eğitim süresi* ve *traktör sayısı*dır.

Figür 7

Alınan Göç Bağımlı Değişkeni için Tüm Algoritmalar Dikkate Alınarak İlk Beşte Yer Alan Değişkenlerin Sıklıkları

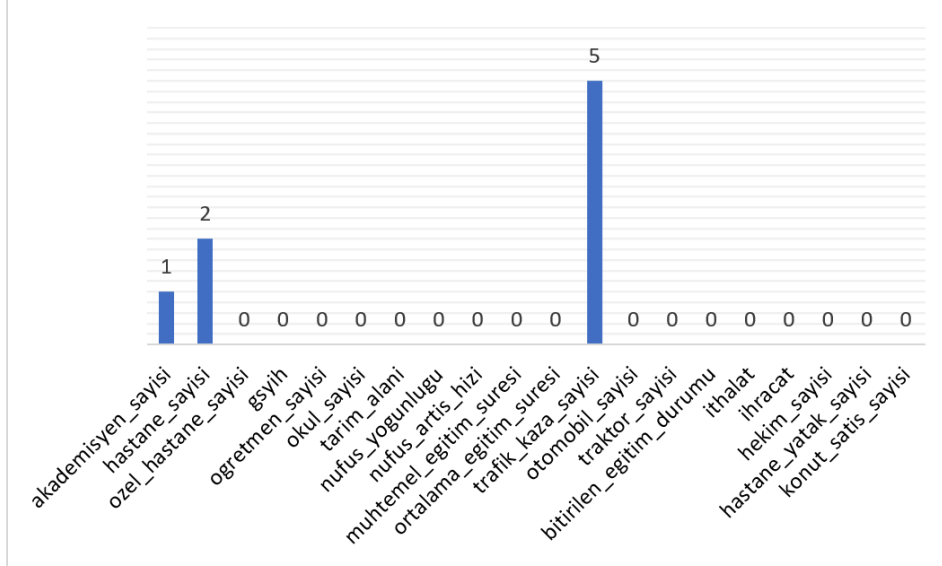


Figür 7’de alınan göç için tüm algoritmalar dikkate alınarak ilk beşte yer alan değişkenlerin sıklıkları verilmiştir. Tüm algoritmalar tarafından en önemli olarak seçilen değişken (en çok seçilen değişken) (8: Sekiz algoritma tarafından önemli olarak seçilen değişken): *trafik kaza sayısı*. Sırasıyla ilk beşte yer alan diğer en önemli değişkenler ise şu şekildedir: *hastane sayısı*, *konut satış sayısı*, *öğretmen sayısı*, *akademisyen sayısı*, *özel hastane sayısı*, *otomobil sayısı*, *nüfus artış hızı*, *ithalat*, *okul sayısı* ve *nüfus yoğunluğu* değişkenleridir.

İlk beşte yer almayan değişkenler ise GSYİH, tarım alanı, muhtemel eğitim süresi, ortalama eğitim süresi, traktör sayısı, bitirilen eğitim durumu, ihracat, hekim sayısı ve hastane yatak sayısıdır.

Figür 8

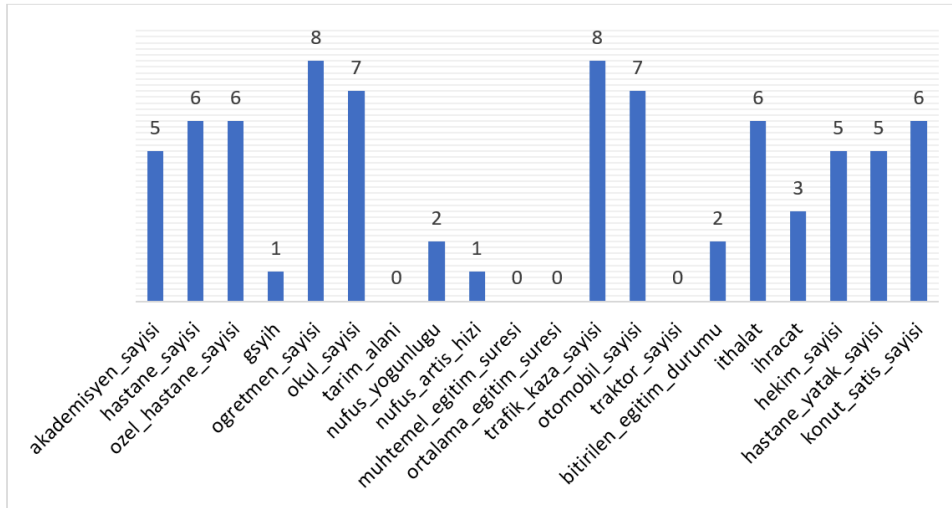
Alınan Göç Bağımlı Değişkeni için Tüm Algoritmalar Dikkate Alınarak İlk Sırada Yer Alan Değişkenlerin Sıklıkları



Figür 8’de alınan göç için tüm algoritmalar dikkate alınarak ilk sırada yer alan değişkenlerin sıklıkları verilmiştir. Algoritmaların beş tanesinde ilk sırada trafik kaza sayısı değişkeni, 2 tanesi hastane sayısı ve 1 tanesinde ise akademisyen sayısı değişkeni ilk sırada yer almaktadır yani en önemli değişkenlerdir.

Figür 9

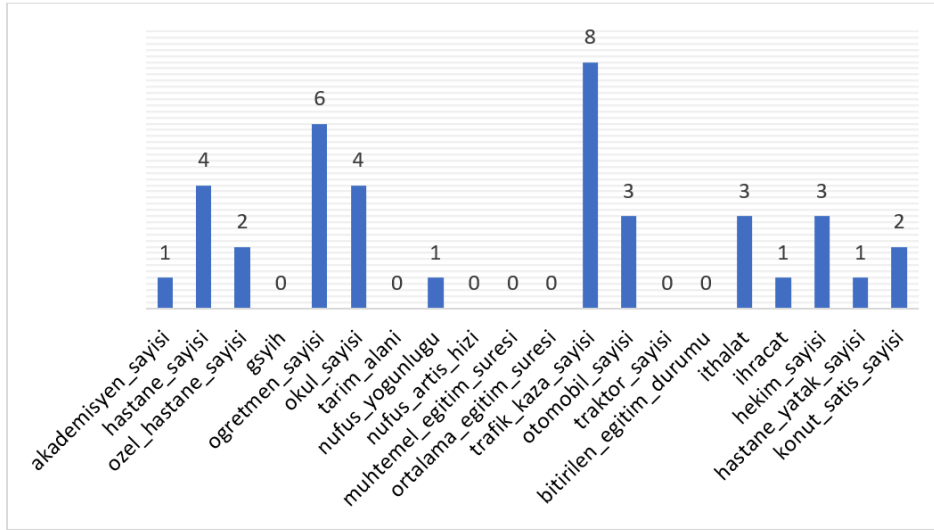
Verilen Göç Bağımlı Değişkeni için Tüm Algoritmalar Dikkate Alınarak İlk Onda Yer Alan Değişkenlerin Sıklıkları



Figür 9’da verilen göç için tüm algoritmalar dikkate alınarak ilk onda yer alan değişkenlerin sıklıkları verilmiştir. Tüm algoritmalar tarafından en önemli olarak seçilen değişkenler (en çok seçilen değişken) (8: Sekiz algoritma tarafından önemli olarak seçilen değişkenler): *öğretmen sayısı* ve *trafik kaza sayısı* değişkenleridir. Sırasıyla ilk onda yer alan diğer en önemli değişkenler ise şu şekildedir: *okul sayısı*, *otomobil sayısı*, *hastane sayısı*, *özel hastane sayısı*, *ithalat*, *konut satış sayısı*, *akademisyen sayısı*, *hekim sayısı*, *hastane yatak sayısı*, *ihracat*, *nüfus yoğunluğu*, *bitirilen eğitim durumu*, *GSYİH* ve *nüfus artış hızı*dır. İlk onda yer almayan değişkenler ise *tarım alanı*, *muhtemel eğitim süresi*, *ortalama eğitim süresi* ve *traktör sayısı*dır.

Figür 10

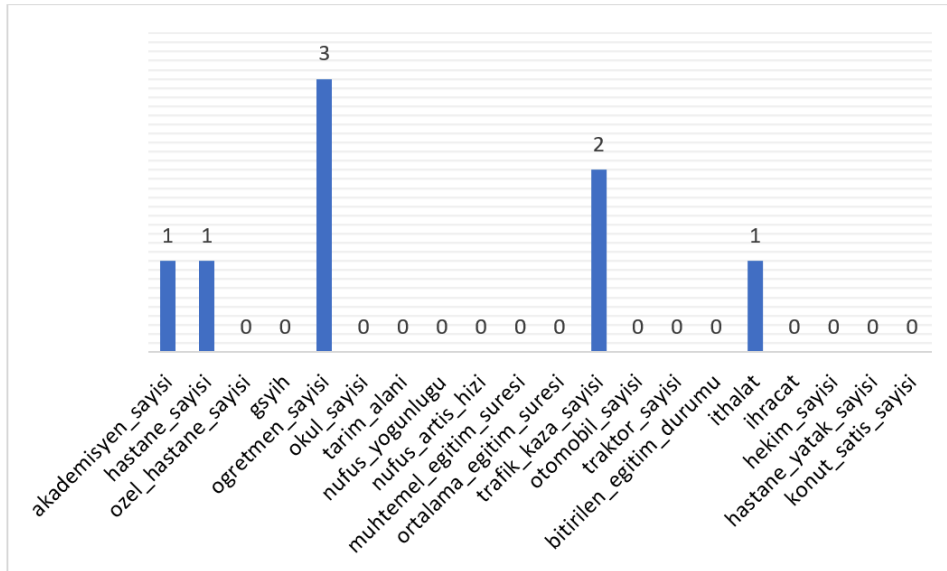
Verilen Göç Bağımlı Değişkeni için Tüm Algoritmalar Dikkate Alınarak İlk Beşte Yer Alan Değişkenlerin Sıklıkları



Figür 10’da verilen göç için tüm algoritmalar dikkate alınarak ilk beşte yer alan değişkenlerin sıklıkları verilmiştir. Tüm algoritmalar tarafından en önemli olarak seçilen değişken (en çok seçilen değişken) (8: Sekiz algoritma tarafından önemli olarak seçilen değişken): *trafik kaza sayısı*. Sırasıyla ilk beşte yer alan diğer en önemli değişkenler ise şu şekildedir: *öğretmen sayısı*, *hastane sayısı*, *okul sayısı*, *otomobil sayısı*, *ithalat*, *hekim sayısı*, *özel hastane sayısı*, *konut satış sayısı*, *akademisyen sayısı*, *nüfus yoğunluğu*, *ihracat* ve *hastane yatak sayısı*. İlk beşte yer almayan değişkenler ise GSYİH, tarım alanı, nüfus artış hızı, muhtemel eğitim süresi, ortalama eğitim süresi, traktör sayısı ve bitirilen eğitim durumudur.

Figür 11

Verilen Göç Bağımlı Değişkeni için Tüm Algoritmalar Dikkate Alınarak İlk Sırada Yer Alan Değişkenlerin Sıklıkları



Figür 11’de verilen göç için tüm algoritmalar dikkate alınarak ilk sırada yer alan değişkenlerin sıklıkları verilmiştir. Algoritmaların üç tanesinde ilk sırada *öğretmen sayısı*, iki tanesinde *trafik kaza sayısı*, birer tanesinde *akademisyen sayısı*, *hastane sayısı* ve *ithalat* değişkeni ilk sırada yer almaktadır yani en önemli değişkenlerdir.

Göç Tahminine İlişkin Sonuçlar

Tüm değişkenler ve en önemli değişkenler dikkate alınarak; basit doğrusal regresyon (LM), Destek Vektör Regresyonu (SVM), Karar Ağaçları (RPART), Rassal Ormanlar (RF), Güçlendirme (Boosting), Gradyan Artırma Makinesi (GBM), Çantalama (Bagging) ve Yığınlama (Stacking) algoritmaları kullanılarak tahmin yapılmıştır. Performans değerlendirme metrikleri kullanılarak modeller karşılaştırılmıştır. Bu metrikler ile tam model ile önemli değişkenleri kullanılan model arasındaki farklar görülür ve eğer önemli değişkenlerle oluşturulan modelin performansı kabul edilebilir seviyedeysen, daha önce değinilen avantajlarından dolayı bu modeli tercih edilmelidir.

Tablo 7 ve 8’de model sonuçları verilmiştir. Tüm değişkenler dikkate alınarak elde edilen sonuçlar (Tüm Değişkenler), tüm öznitelik seçim algoritmalarında önem sırası olarak ilk onda yer alan değişkenlere bakılıp bunlardan en az beş öznitelik önem algoritması tarafından seçilen değişkenler (İlk On) ve en az on ortak önem skoru değerine sahip değişkenler (Ortak Önem Skoru) dikkate alınarak model sonuçları verilmiştir.

Tablo 7 incelendiğinde, öznitelik seçimi sonrası algoritmaların performansının genellikle iyileştiği görülmüştür yani daha düşük hata değerine sahiptir. Ayrıca, çoğu algoritma ilk on öznitelik seçimine nazaran ortak önem skoru yaklaşımı ile daha iyi performans değerine sahip olduğu görülmüştür. En iyi sonucu veren algoritma (rmse ve mse için) ise ortak önem skoru yöntemi ile değişken önemi sonrası stacking (XGBOOST, GBM, BAGGING) algoritmasıdır. MAE değeri için en iyi sonucu veren tüm değişkenlerle stacking (SVM, XGBOOST, GBM) algoritmasıdır. Ancak daha önce de değindiğimiz gibi; avantajlarından dolayı hata kabul edilebilir seviyede ise daha az değişkene sahip model tercih edilmelidir. Sonuçlar incelendiğinde arada çok fark olmayan ortak önem skoru yöntemi ile değişken önemi sonrası stacking (RF, XGBOOST, GBM) algoritması tercih edilebilir.

Tablo 7

Alınan Göç Bağımlı Değişkenine İlişkin Sonuçlar

Model	Öznitelik	MSE	RMSE	MAE
LM	Tüm Değişkenler	0,031	0,173	0,109
	İlk On	0,031	0,177	0,108
	Ortak Önem Skoru	0,032	0,179	0,112
RPART	Tüm Değişkenler	0,128	0,357	0,198
	İlk On	0,128	0,357	0,198
	Ortak Önem Skoru	0,128	0,357	0,198
RF	Tüm Değişkenler	0,039	0,198	0,096
	İlk On	0,041	0,203	0,105
	Ortak Önem Skoru	0,038	0,194	0,094
SVM	Tüm Değişkenler	0,028	0,166	0,102
	İlk On	0,029	0,171	0,105
	Ortak Önem Skoru	0,028	0,167	0,106
XGBOOST	Tüm Değişkenler	0,036	0,191	0,098
	İlk On	0,040	0,200	0,109
	Ortak Önem Skoru	0,035	0,187	0,094
GBM	Tüm Değişkenler	0,228	0,477	0,276
	İlk On	0,395	0,628	0,279
	Ortak Önem Skoru	0,242	0,492	0,269
BAGGING	Tüm Değişkenler	0,035	0,186	0,131

Model	Öznitelik	MSE	RMSE	MAE
	İlk On	0,039	0,198	0,133
	Ortak Önem Skoru	0,034	0,184	0,132
	Tüm Değişkenler	0,029	0,170	0,089
LM, RF, SVM, XGBTREE	İlk On	0,036	0,190	0,105
	Ortak Önem Skoru	0,021	0,145	0,092
	Tüm Değişkenler	0,025	0,158	0,099
LM, RPART, SVM	İlk On	0,029	0,171	0,101
	Ortak Önem Skoru	0,028	0,168	0,101
	Tüm Değişkenler	0,040	0,200	0,094
LM, RPART, RF	İlk On	0,041	0,202	0,103
	Ortak Önem Skoru	0,038	0,196	0,091
	Tüm Değişkenler	0,036	0,190	0,091
LM, RPART, XGBTREE	İlk On	0,029	0,169	0,108
	Ortak Önem Skoru	0,029	0,172	0,103
	Tüm Değişkenler	0,023	0,152	0,098
LM, RPART, GBM	İlk On	0,023	0,152	0,105
	Ortak Önem Skoru	0,025	0,158	0,107
	Tüm Değişkenler	0,021	0,146	0,104
LM, RPART, BAGGING	İlk On	0,023	0,153	0,105
	Ortak Önem Skoru	0,025	0,160	0,109
	Tüm Değişkenler	0,030	0,174	0,093
LM, SVM, RF	İlk On	0,027	0,165	0,097
	Ortak Önem Skoru	0,027	0,164	0,094
	Tüm Değişkenler	0,027	0,163	0,091
LM, SVM, XGBOOST	İlk On	0,027	0,165	0,098
	Ortak Önem Skoru	0,022	0,150	0,090
	Tüm Değişkenler	0,021	0,145	0,089
LM, SVM, GBM	İlk On	0,031	0,176	0,104
	Ortak Önem Skoru	0,025	0,160	0,098
	Tüm Değişkenler	0,026	0,162	0,101
LM, SVM, BAGGING	İlk On	0,033	0,182	0,104
	Ortak Önem Skoru	0,029	0,171	0,103
	Tüm Değişkenler	0,039	0,199	0,086
LM, RF, XGBOOST	İlk On	0,046	0,216	0,105
	Ortak Önem Skoru	0,036	0,191	0,091
	Tüm Değişkenler	0,042	0,206	0,093
LM, RF, GBM	İlk On	0,048	0,219	0,106
	Ortak Önem Skoru	0,041	0,201	0,094
	Tüm Değişkenler	0,045	0,212	0,095
LM, RF, BAGGING	İlk On	0,044	0,210	0,103
	Ortak Önem Skoru	0,040	0,199	0,094
	Tüm Değişkenler	0,034	0,184	0,090
RPART, SVM, RF	Tüm Değişkenler	0,034	0,184	0,090

Model	Öznitelik	MSE	RMSE	MAE
	İlk On	0,027	0,164	0,099
	Ortak Önem Skoru	0,031	0,176	0,096
	Tüm Değişkenler	0,026	0,161	0,091
RPART, SVM, XGBOOST	İlk On	0,027	0,164	0,101
	Ortak Önem Skoru	0,024	0,154	0,092
	Tüm Değişkenler	0,021	0,146	0,088
RPART, SVM, GBM	İlk On	0,027	0,165	0,101
	Ortak Önem Skoru	0,027	0,163	0,100
	Tüm Değişkenler	0,024	0,158	0,101
RPART, SVM, BAGGING	İlk On	0,027	0,169	0,104
	Ortak Önem Skoru	0,029	0,170	0,105
	Tüm Değişkenler	0,045	0,213	0,093
RPART, RF, XGBOOST	İlk On	0,040	0,199	0,101
	Ortak Önem Skoru	0,029	0,172	0,079
	Tüm Değişkenler	0,042	0,206	0,093
RPART, RF, GBM	İlk On	0,042	0,205	0,103
	Ortak Önem Skoru	0,047	0,217	0,096
	Tüm Değişkenler	0,042	0,206	0,095
RPART, RF, BAGGING	İlk On	0,046	0,215	0,111
	Ortak Önem Skoru	0,046	0,214	0,095
	Tüm Değişkenler	0,022	0,150	0,084
RPART, XGBOOST, GBM	İlk On	0,031	0,177	0,105
	Ortak Önem Skoru	0,018	0,135	0,085
	Tüm Değişkenler	0,020	0,141	0,083
RPART, XGBOOST, BAGGING	İlk On	0,032	0,178	0,106
	Ortak Önem Skoru	0,018	0,136	0,085
	Tüm Değişkenler	0,033	0,181	0,085
SVM, RF, XGBOOST	İlk On	0,027	0,163	0,098
	Ortak Önem Skoru	0,031	0,175	0,096
	Tüm Değişkenler	0,032	0,180	0,087
SVM, RF, GBM	İlk On	0,029	0,171	0,101
	Ortak Önem Skoru	0,030	0,173	0,095
	Tüm Değişkenler	0,036	0,190	0,092
SVM, RF, BAGGING	İlk On	0,029	0,170	0,100
	Ortak Önem Skoru	0,031	0,176	0,097
	Tüm Değişkenler	0,018	0,134	0,071
SVM, XGBOOST, GBM	İlk On	0,029	0,171	0,102
	Ortak Önem Skoru	0,024	0,155	0,094
	Tüm Değişkenler	0,020	0,142	0,078
SVM, XGBOOST, BAGGING	İlk On	0,029	0,171	0,102
	Ortak Önem Skoru	0,025	0,158	0,096
	Tüm Değişkenler	0,021	0,146	0,092

Model	Öznitelik	MSE	RMSE	MAE
RF, XGBOOST, GBM	İlk On	0,028	0,168	0,101
	Ortak Önem Skoru	0,026	0,162	0,100
	Tüm Değişkenler	0,034	0,183	0,087
	İlk On	0,049	0,222	0,107
	Ortak Önem Skoru	0,025	0,159	0,078
	Tüm Değişkenler	0,035	0,187	0,090
RF, XGBOOST, BAGGING	İlk On	0,047	0,216	0,104
	Ortak Önem Skoru	0,025	0,158	0,079
	Tüm Değişkenler	0,052	0,228	0,097
RF, GBM, BAGGING	İlk On	0,047	0,218	0,105
	Ortak Önem Skoru	0,045	0,213	0,099
	Tüm Değişkenler	0,040	0,199	0,098
XGBOOST, GBM, BAGGING	İlk On	0,051	0,225	0,106
	Ortak Önem Skoru	0,017	0,130	0,0802
	Tüm Değişkenler	0,040	0,199	0,098

Tablo 8 incelendiğinde, öznitelik seçimi sonrası algoritmaların performansının genellikle iyileştiği görülmüştür yani daha düşük hata değerine sahiptir. Ayrıca, çoğu algoritma ortak önem skoru yaklaşımı nazaran ilk on öznitelik seçimi yaklaşımı ile daha iyi performans değerine sahip olduğu görülmüştür. En iyi sonucu veren algoritma (rmse ve mse için) tüm değişkenlerle XGBOOST algoritmasıdır. MAE değeri için en iyi sonucu veren tüm değişkenlerle RF algoritmasıdır. Ancak daha önce de değindiğimiz gibi; avantajlarından dolayı hata kabul edilebilir seviyede ise daha az değişkene sahip model tercih edilmelidir. Sonuçlar incelendiğinde arada çok fark olmayan ilk on öznitelik seçim yaklaşımı ile (rmse ve mse için) en iyi sonucu veren XGBOOST algoritması; MAE değeri için ise ortak önem skoru yaklaşımı ile değişken önemi sonrası RF algoritması tercih edilebilir.

Tablo 8

Verilen göç bağımlı değişkenine ilişkin sonuçlar

Model	Öznitelik	MSE	RMSE	MAE
LM	Tüm Değişkenler	0,034	0,186	0,108
	İlk On	0,040	0,200	0,119
	Ortak Önem Skoru	0,034	0,183	0,109
RPART	Tüm Değişkenler	0,208	0,456	0,163
	İlk On	0,208	0,456	0,163
	Ortak Önem Skoru	0,208	0,456	0,163
RF	Tüm Değişkenler	0,017	0,131	0,069
	İlk On	0,021	0,144	0,076
	Ortak Önem Skoru	0,019	0,137	0,074
SVM	Tüm Değişkenler	0,015	0,121	0,083
	İlk On	0,015	0,123	0,086
	Ortak Önem Skoru	0,016	0,125	0,090
XGBOOST	Tüm Değişkenler	0,012	0,108	0,071
	İlk On	0,013	0,116	0,079
	Ortak Önem Skoru	0,014	0,120	0,081
GBM	Tüm Değişkenler	0,265	0,515	0,281

Model	Öznitelik	MSE	RMSE	MAE
	İlk On	0,384	0,620	0,292
	Ortak Önem Skoru	0,351	0,593	0,316
	Tüm Değişkenler	0,054	0,232	0,116
BAGGING	İlk On	0,055	0,234	0,116
	Ortak Önem Skoru	0,055	0,235	0,117
	Tüm Değişkenler	0,027	0,165	0,097
LM, RF, SVM, XGBTREE	İlk On	0,021	0,146	0,103
	Ortak Önem Skoru	0,027	0,165	0,106
	Tüm Değişkenler	0,018	0,135	0,086
LM, RPART, SVM	İlk On	0,018	0,135	0,092
	Ortak Önem Skoru	0,017	0,130	0,088
	Tüm Değişkenler	0,032	0,180	0,108
LM, RPART, RF	İlk On	0,024	0,156	0,095
	Ortak Önem Skoru	0,032	0,178	0,109
	Tüm Değişkenler	0,017	0,132	0,079
LM, RPART, XGBTREE	İlk On	0,017	0,130	0,089
	Ortak Önem Skoru	0,025	0,160	0,086
	Tüm Değişkenler	0,026	0,163	0,098
LM, RPART, GBM	İlk On	0,035	0,187	0,115
	Ortak Önem Skoru	0,028	0,167	0,103
	Tüm Değişkenler	0,025	0,159	0,097
LM, RPART, BAGGING	İlk On	0,036	0,190	0,114
	Ortak Önem Skoru	0,027	0,165	0,102
	Tüm Değişkenler	0,026	0,160	0,104
LM, SVM, RF	İlk On	0,019	0,139	0,104
	Ortak Önem Skoru	0,041	0,203	0,130
	Tüm Değişkenler	0,019	0,139	0,091
LM, SVM, XGBOOST	İlk On	0,019	0,136	0,091
	Ortak Önem Skoru	0,021	0,145	0,087
	Tüm Değişkenler	0,019	0,137	0,094
LM, SVM, GBM	İlk On	0,018	0,135	0,093
	Ortak Önem Skoru	0,015	0,124	0,0903
	Tüm Değişkenler	0,018	0,135	0,087
LM, SVM, BAGGING	İlk On	0,018	0,135	0,093
	Ortak Önem Skoru	0,017	0,130	0,090
	Tüm Değişkenler	0,027	0,163	0,094
LM, RF, XGBOOST	İlk On	0,015	0,122	0,087
	Ortak Önem Skoru	0,065	0,255	0,142
	Tüm Değişkenler	0,027	0,166	0,098
LM, RF, GBM	İlk On	0,023	0,153	0,094
	Ortak Önem Skoru	0,034	0,185	0,112
	Tüm Değişkenler	0,028	0,169	0,101

Model	Öznitelik	MSE	RMSE	MAE
	İlk On	0,024	0,154	0,095
	Ortak Önem Skoru	0,032	0,180	0,109
	Tüm Değişkenler	0,028	0,169	0,101
RPART, SVM, RF	İlk On	0,034	0,185	0,124
	Ortak Önem Skoru	0,035	0,188	0,129
	Tüm Değişkenler	0,015	0,124	0,076
RPART, SVM, XGBOOST	İlk On	0,014	0,117	0,083
	Ortak Önem Skoru	0,020	0,142	0,082
	Tüm Değişkenler	0,017	0,130	0,092
RPART, SVM, GBM	İlk On	0,015	0,1243	0,086
	Ortak Önem Skoru	0,014	0,130	0,089
	Tüm Değişkenler	0,017	0,131	0,0830
RPART, SVM, BAGGING	İlk On	0,016	0,126	0,085
	Ortak Önem Skoru	0,018	0,136	0,087
	Tüm Değişkenler	0,018	0,136	0,096
RPART, RF, XGBOOST	İlk On	0,022	0,150	0,089
	Ortak Önem Skoru	0,024	0,154	0,100
	Tüm Değişkenler	0,043	0,208	0,0882
RPART, RF, GBM	İlk On	0,035	0,187	0,0819
	Ortak Önem Skoru	0,038	0,196	0,087
	Tüm Değişkenler	0,047	0,217	0,088
RPART, RF, BAGGING	İlk On	0,038	0,195	0,083
	Ortak Önem Skoru	0,041	0,201	0,088
	Tüm Değişkenler	0,029	0,171	0,084
RPART, XGBOOST, GBM	İlk On	0,035	0,188	0,106
	Ortak Önem Skoru	0,026	0,162	0,085
	Tüm Değişkenler	0,031	0,176	0,087
RPART, XGBOOST, BAGGING	İlk On	0,038	0,195	0,109
	Ortak Önem Skoru	0,027	0,166	0,087
	Tüm Değişkenler	0,022	0,148	0,101
SVM, RF, XGBOOST	İlk On	0,020	0,142	0,099
	Ortak Önem Skoru	0,053	0,230	0,135
	Tüm Değişkenler	0,024	0,154	0,106
SVM, RF, GBM	İlk On	0,026	0,162	0,107
	Ortak Önem Skoru	0,044	0,209	0,131
	Tüm Değişkenler	0,022	0,149	0,106
SVM, RF, BAGGING	İlk On	0,026	0,162	0,110
	Ortak Önem Skoru	0,048	0,219	0,138
	Tüm Değişkenler	0,020	0,143	0,099
SVM, XGBOOST, GBM	İlk On	0,017	0,132	0,090
	Ortak Önem Skoru	0,015	0,124	0,084
	Tüm Değişkenler	0,020	0,140	0,092

Model	Öznitelik	MSE	RMSE	MAE
SVM, GBM, BAGGING	İlk On	0,017	0,131	0,090
	Ortak Önem Skoru	0,018	0,135	0,087
	Tüm Değişkenler	0,021	0,146	0,105
	İlk On	0,015	0,124	0,085
	Ortak Önem Skoru	0,017	0,129	0,090
	Tüm Değişkenler	0,015	0,124	0,081
RF, XGBOOST, GBM	İlk On	0,014	0,120	0,076
	Ortak Önem Skoru	0,033	0,181	0,105
	Tüm Değişkenler	0,014	0,118	0,079
RF, XGBOOST, BAGGING	İlk On	0,015	0,124	0,076
	Ortak Önem Skoru	0,034	0,185	0,108
	Tüm Değişkenler	0,044	0,210	0,088
RF, GBM, BAGGING	İlk On	0,039	0,198	0,084
	Ortak Önem Skoru	0,038	0,197	0,084
	Tüm Değişkenler	0,016	0,128	0,0823
XGBOOST, GBM, BAGGING	İlk On	0,034	0,185	0,098
	Ortak Önem Skoru	0,024	0,156	0,091

SONUÇ, TARTIŞMA VE ÖNERİLER

İç göç Türkiye için önemli bir sorundur ve iç göçle ilgili çalışmalar çok önemlidir. Bu durum nedeniyle, bu çalışmanın amacı Türkiye’de 2018-2022 yılları arasında iç göç için alınan göç ve verilen göç ayrı ayrı ele alınarak değişken önemleri bulunup; verilen göç ve alınan göçün tahmini tüm değişkenler ve değişken önemi sonrası en önemli olan değişkenler dikkate alınarak analizler yapılarak performansları karşılaştırmaktır.

Analiz sonuçlarına göre, trafik kaza sayısı, öğretmen sayısı, akademisyen sayısı, hastane sayısı ve konut satış sayısı gibi değişkenler hem alınan göç hem de verilen göç için en önemli belirleyiciler arasında yer almıştır. Değişken önemi ile belirlenen bu faktörlerin çoğu, literatürde iç göç üzerinde etkili olduğu belirtilen ekonomik, sosyal ve demografik unsurlar ile uyum göstermektedir. Özellikle topluluk öğrenme (ensemble learning) algoritmalarının, özellikle de stacking yönteminin yüksek doğrulukta tahminler sunduğu gözlemlenmiştir. Ayrıca, öznitelik seçimi sonrasında yapılan modellemelerde hata oranlarının genellikle düştüğü yani algoritmaların performansının genellikle iyileştiği görülmüştür. Bu bulgu, veri boyutunun azaltılmasının ve önemsiz değişkenlerin dışlanması model performansı üzerinde olumlu etkiler yarattığını ortaya koymaktadır.

Bu çalışmayla alınan ve verilen göçte önemli değişkenler belirlenmiştir ve makine öğrenmesi yaklaşımıyla düşük hata değeri ile göç tahmini yapılmıştır. Önemli değişkenlerin üzerinde durularak politikalar geliştirilebilir ve iyi performans veren modeller kullanılarak göç tahmini yapılabilir. Elde edilen bulgulara göre yapılabilecek politika önerileri şunlardır:

Eğitim Altyapısının Güçlendirilmesi: Çalışmada öne çıkan önemli değişkenlerden biri öğretmen sayısıdır. Bu bulgu, eğitim altyapısının bulunduğu ildeki nüfus hareketliliğiyle doğrudan ilişkili olduğunu göstermektedir. Bu doğrultuda göç veren illerde özellikle kırsal alanlarda öğretmen istihdamı artırılmalı, eğitimde bölgesel dengesizlikler azaltılmalıdır. Yerel okul sayıları ve niteliği artırılarak yerinde eğitim teşvik edilmelidir. Teşvik programları aracılığıyla öğretmenlerin göç veren bölgelere atanması özendirilmelidir (örneğin maaş artışı, lojman desteği, kariyer imkanları).

Sağlık Hizmetlerine Erişimin Artırılması: Hastane sayısı, özel hastane sayısı, hekim sayısı ve hastane yatak kapasitesi, göç kararını etkileyen önemli sağlık göstergeleridir. Göç veren bölgelerde sağlık altyapısı güçlendirilmeli, özellikle temel sağlık hizmetlerine erişim yaygınlaştırılmalıdır. Nitelikli uzman hekimlerin dezavantajlı bölgelere atanmasını teşvik edecek sağlık politikaları geliştirilmelidir.

Ulaşım ve Trafik Güvenliği Yatırımları: Trafik kaza sayısının hem alınan hem de verilen göç açısından kritik önemde olduğu görülmüştür. Bu durum, ulaşım altyapısı ve güvenliğinin yaşam kalitesine etkisini göstermektedir. Kaza oranı yüksek illerde karayolu altyapısı modernize edilmeli; kavşak düzenlemeleri ve yol genişletmeleri yapılmalıdır. Toplu taşıma sistemleri geliştirilerek ulaşım maliyetleri düşürülmeli ve konfor artırılmalıdır. Trafik güvenliğine yönelik yerel eğitim ve denetim politikaları yaygınlaştırılmalıdır.

Konut Politikalarının Gözden Geçirilmesi: Konut satış sayısının göç kararlarında belirleyici olması, barınma olanaklarının göç hareketlerini etkilediğini göstermektedir. Göç alan şehirlerde konut arz-talep dengesi sağlanmalı, dar gelirli gruplar için sosyal konut projeleri teşvik edilmelidir. Göç veren şehirlerde ise kamu destekli konut iyileştirme projeleriyle yaşam kalitesi artırılabilir.

Bölgesel Kalkınma Odaklı Entegre Politikalar: Göç eğilimleri bölgesel kalkınma planlarıyla entegre edilerek, sosyo-ekonomik yatırımlar göç veren bölgelerde yoğunlaştırılmalıdır. GAP ve DAP benzeri büyük ölçekli projelerin yaygınlaştırılması ile göç baskısı altında olan bölgelerde yaşam kalitesi artırılabilir. Merkezi idare ile yerel yönetimler arasında koordinasyon güçlendirilerek, göçün yönünü dengelemeye yönelik sürdürülebilir politikalar üretilebilir.

Çalışmanın katkısından bahsedecek olursak, yapılan literatür incelemesinde Türkiye’de iç göçe ilişkin yapılan çalışmalarda çoğunlukla ekonometrik yöntemler kullanıldığı görülmüştür. Makine öğrenmesi yaklaşımı ile değişken önemi ve göç tahminine ilişkin çalışmaya rastlanmamıştır. Ayrıca yapılan çalışmada kapsamlı model denemesi yapılmıştır. Bu bağlamda çalışma hem yöntemsel çeşitliliği hem de uygulamalı niteliğiyle iç göç üzerine yapılan literatüre özgün katkılar sunmakta, gelecekteki çalışmalara metodolojik bir zemin oluşturmaktadır.

Gelecek çalışmalarda, SHAP (SHapley Additive exPlanations) gibi açıklanabilir yapay zekâ yöntemleri kullanılarak modellerin daha şeffaf hale getirilmesi ve karar mekanizmalarının daha anlaşılır olması sağlanabilir. Kullanılan modellerin hiper parametreleri çapraz doğrulama veya grid search gibi yöntemlerle optimize edilip denenebilir. Ayrıca, düzenleme (regularization) tabanlı öznelik seçimi yöntemleri, örneğin Lasso gibi farklı öznelik seçim yöntemleri kullanılarak performansı incelenebilir. Veri eksikliği nedeniyle modele dahil edilemeyen istihdam ve güvenlik gibi faktörlerin de değerlendirilmesi, iç göçü etkileyen yapısal dinamiklerin daha bütüncül analizine katkı sunacaktır. Türkiye örneği, benzer yapıya sahip gelişmekte olan ülkelerle karşılaştırılarak, iç göç dinamiklerinin küresel düzeydeki benzerlik ve farklılıkları ortaya konulabilir.



Hakem Değerlendirmesi
Çıkar Çatışması
Finansal Destek

Dış bağımsız.
Yazar çıkar çatışması bildirmemiştir.
Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

Peer Review
Conflict of Interest
Grant Support

Externally peer reviewed.
The author has no conflict of interest to declare.
The author declared that this study has received no financial support.

Yazar Bilgileri
Author Details

Meryem Pulat Öztürk
1 Fırat Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, Elâzığ, Türkiye.



Kaynakça | References

- Akay, E. Ç. (2018). Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi. *Sosyal Bilimler Araştırma Dergisi*, 7(2), 41-53.
- Akar, Ö., Güngör, O. (2012). Classification of multispectral images using Random Forest algorithm. *Journal of Geodesy and Geoinformation*, 1(2), 105-112.
- Anavatan, A. (2017). Türkiye’de iç göç ve belirleyicileri: mekansal veri analizi. *Social Sciences Studies Journal*, 3(6), 1109-1116.
- Atalay, M., Öztürk, Ş. (2016). Türkiye’deki illerin göç ve işsizlik istatistiklerine göre kümelenmesi. In *2nd International Congress on Applied Sciences: "Migration, Poverty and Employment"*, 23-25 Eylül, Konya/Turkey, *Bildiriler Kitabı*.
- Bahçıvan, S., Kılıç, R. (2024). Türkiye’de iç göçün belirleyicilerine ilişkin panel veri analizi. *Toplum Ekonomi ve Yönetim Dergisi*, 5(2), 289-302.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
- Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- Breiman L., (2002), Manual on setting up, using, and understanding random forests V3.1, http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf.
- Breiman, L. (1996). Stacked Regression. *Machine Learning*, 24(1), 49-64.
- Bilgin, M. (2018). *Makine Öğrenmesi*. İstanbul: Papatya Yayınları.
- Bülbül, S., Kose, A. (2010). Türkiye’de bölgelerarası iç göç hareketlerinin çok boyutlu ölçekleme yöntemi ile incelenmesi. *İstanbul Üniversitesi İşletme Fakültesi Dergisi*, 39(1), 75-94.
- Chen, T., Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (785-794).
- Cheng, J., Greiner, R., Kelly, J., Bell, D., Liu, W. (2011). Learning Bayesian networks from data: An information-theory based approach. *Artificial Intelligence*, 137(1-2), 43-90.
- Çatalbaş, G. K., Yazar, Ö. (2015). Türkiye’deki bölgeler arası iç göçü etkileyen faktörlerin panel veri analizi ile belirlenmesi. *Alphanumeric Journal*, 3(1), 99-117.
- Çelik, R., Arslan, I. (2018). Göç ve işsizlik arasındaki ilişki: ampirik bir uygulama. In *Journal of Social Policy Conferences*, 74, 65-75.
- Deka, P. C. (2014). Support Vector Machine Applications in the Field of Hydrology: A Review. *Applied Soft Computing*, 19, 372-386.
- Doğan, Ö. S. (2017). Population Movements In Turkey: Internal Migration. *Current Researches in Geography*, 49.
- Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. J., & Vapnik, V. N. (1997). Support Vector Regression Machines. *Advances in Neural Information Processing Systems*, 9, 155-161.
- Dücan, E. (2016). Türkiye’de iç göçün sosyo-ekonomik nedenlerinin bölgesel analizi. *Ekonomik ve Sosyal Araştırmalar Dergisi*, 12(2), 167-183.
- Emeç, H., Birecikli, Ş. Ü., Güler, B. K. (2019). İstanbul’a yönelik iç göç hareketlerinin ekonometrik analizi. *İnsan ve insan*, 6(22), 785-808.
- Ercilasun, M., Gencer, E. A. H., Ersin, Ö. Ö. (2011, October). Türkiye’deki İç Göçleri Belirleyen Faktörlerin Modellenmesi. In *International Conference on Eurasian Economies* (ss. 12-14).
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.
- Gezici, F., Keskin, B. (2005, August). Interaction between regional inequalities and internal migration in Turkey. In *ERSA Conference Papers* (ss. 1-18).
- Gözen, S., Çelikay, F. (2024). Türkiye’de Kamu Harcamaları ve Yoksulluğun İç Göç Üzerindeki Mekânsal Etkileri: Ampirik Bir İnceleme. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (81), 72-91.
- Guyon, I., Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157-1182.
- Han, J., Pei, J., Kamber, M. (2011). *Data Mining: Concepts and Techniques (Third Edition)*. San Francisco: Morgan Kaufmann.
- Hasan, M. A. M., Nasser, M., Ahmad, S., Molla, K. I. (2016). Feature selection for intrusion detection using random forest. *Journal of information security*, 7(3), 129-140.
- Hussain, M., Wajid, S. K., Elzaat, A., Berbar, M. (2011). A Comparison of SVM Kernel Functions for Breast Cancer Detection. In *2011 Eight International Conference Computer Graphics, Imaging and Visualization* (ss. 145-150), Washington, DC, United States. 17-19 Ağustos 2011.
- James, G., Witten, D., Hastie, T., Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York: Springer.

- Kavitha, S., Varuna, S., Ramya, R. (2016, November). A comparative analysis on linear regression and support vector regression. In *2016 online international conference on green engineering and technologies (IC-GET)* (ss. 1-5).
- Kavzoğlu, T., Çölkesen, G., Case, K. (2010). Karar Ağaçları ile Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği. *Harita Teknolojileri Elektronik Dergisi*, 2(1), 36-45.
- Kursa, M. B., Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of statistical software*, 36, 1-13.
- Liaw, A., Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18-22.
- Manavgat, G., Saygılı, R. F. (2016, December). Türkiye’de içgöçü etkileyen faktörler üzerine bir uygulama: Mekânsal panel veri analizi. In *2nd International Conference on Applied Economics and Finance (ICOAEF)* (ss. 5-6).
- Maulud, D., Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of applied science and technology trends*, 1(2), 140-147.
- Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., Brown, S. D. (2004). An Introduction to Decision Tree Modeling. *Journal of Chemometrics*, 18(6), 275-285.
- Ozsari, M., Food, K. (2016). An Analysis of the Reasons of Internal Migration in Turkey with Logit Method. *Scholar Journal of Applied Sciences and Research*, 2(1), 3-9.
- Özdemir, D. (2018). Türkiye’de bölgelerarası iç göç hareketlerinin belirleyicileri. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 22(3), 1337-1349.
- Özdemir, F., Doğan, Ö. S. (2022). Türkiye’de bölgelerarası iç göçlerin kentleşmeye etkileri: istanbul örneği. *Sosyal, Beşerî ve İdari Bilimler Dergisi*, 5(4), 432-450.
- Öztemel, E. (2003). *Yapay Sinir Ağları*. Papatya Yayıncılık, İstanbul.
- Ribeiro, M. T., Singh, S., Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ss. 1135-1144).
- Salunkhe, U. R., Mali, S. N. (2016). Classifier Ensemble Design for Imbalanced Data Classification: A Hybrid Approach. *Procedia Computer Science*, 85, 725-732.
- Schapiro, R. E. (1990). The strength of weak learnability. *Machine Learning*, 5(2), 197-227.
- Strobl, C., Boulesteix, A.-L., Zeileis, A., Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources, and a solution. *BMC Bioinformatics*, 8(1), 25.
- Şahin, M., Kılıç, M. E. (2024). Türkiye’de İller Arası Alınan Göç: Çok Boyutlu Panel Veri Analizi. *İşletme Bilimi Dergisi*, 12(1), 1-13.
- Ustebay, S., Turgut, Z., Aydın, M. A. (2018, December). Intrusion detection system with recursive feature elimination by using random forest and deep learning classifier. In *2018 international congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT)* (ss. 71-76). IEEE.
- Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.
- Wang, Z., Wang, Y., Srinivasan, R. S. (2018). A Novel Ensemble Learning Approach to Support Building Energy Use Prediction. *Energy and Buildings*, 159, 109-122.
- Witten, I. H., Frank, E., Hall, M. A. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. USA: Morgan Kaufmann Publishers.
- Wolpert, D. H. (1992). Stacked Generalization. *Neural Networks*, 5(2), 241-259.

