İSTANBUL UNIVERSITY PRESS

Istanbul Business Research, 50(2), 465-493
DOI: 10.26650/ibr.2021.50.015019
http://ibr.istanbul.edu.tr/
http://dergipark.gov.tr/ibr

Istanbul Business Research

Submitted: 10.01.2019
Revision Requested: 20.05.2021
Last Revision Received: 03.06.2021
Accepted: 10.06.2021
Published Online: 07.10.2021

RESEARCH ARTICLE

Veri Madenciliği Karar Ağaçları Kullanarak Ülkelerin İnovasyon Değerlerinin Tahmini ve Doğrusal Regresyon Modeli ile Karşılaştırmalı Bir Uygulama

Merve Doğruel [1], Seniye Ümit Fırat[2]

Öz

Ülkelerin sahip olduğu inovasyon seviyeleri ve kapasiteleri, günümüzde hem rekabet edebilirlik hem de yaşamakta olduğumuz Endüstri 4.0 Devrimi açısından son derece önemlidir. Bu kapsamda bakıldığında, ülkeler açısından kapasite ve seviye göreceli bir kavram olarak kalmaktadır ve küresel karşılaştırmalar açısından ortak bir ölçme sistemine gereksinim vardır. Bu ihtiyacı önemli ölçüde karşılayan Ağ Yapılara Hazır Olma Endeksi (AYHOE) ve Küresel İnovasyon Endeksi (KİE), ülkelerin inovasyon seviyelerinin belirlenmesinde etkili ve kapsamlı endekslerdir. Ayrıca her iki endeks de akademik altyapıya sahiptir ve bu nedenle araştırmacılar için önemli bir veri kaynağıdır. Bu çalışma, KİE değeri ve AYHOE endeksine ait alt endekslerin boyutlarında yer alan göstergeler kullanılarak, denetimli makine öğrenmesi temellerine dayanan bir veri madenciliği tekniği olan regresyon ağacı analizi ve doğrusal regresyon analizi uygulamalarını ve karşılaştırmasını içermektedir. Araştırmanın amacı regresyon ağacı uygulayarak, AYHOE göstergelerinden hareketle KİE tahminlemesi yapmak ve en iyi ayrılmayı sağlayan KİE göstergelerini belirlemektir. Analiz için Sınıflandırma ve Regresyon Ağacı ((SRA) - Clasification and Regression Tree (CART)) algoritması kullanılmıştır. Analiz sonucunda AYHOE kapsamındaki hangi göstergelerin, KİE değerleri tahmininde ve ülke sıralamasında kullanılabileceği belirlenmiştir. Aynı veri seti kullanılarak doğrusal regresyon analizi uygulanmıştır. SRA algoritması ile elde edilen regresyon ağacı sonuçları, doğrusal regresyon modelinden elde edilen çıkarımlar ile karşılaştırılmıştır.

Anahtar Kelimeler

Ağ Yapılara Hazır Olma Endeksi, İnovasyon, Karar Ağacı Öğrenmesi, Küresel İnovasyon Endeksi, Sınıflandırma ve Regresyon Ağacı

Prediction of Innovation Values of Countries Using Data Mining Decision Trees and a Comparative Application with Linear Regression Model

Abstract

Innovation levels and capacities of countries are two very important factors for competitiveness as well as the current Industrial 4.0 Revolution. In this context, capacity and level are relative concepts, with a great need for a common measurement system on global-based comparisons. The Network Readiness Index (NRI) and the Global Innovation Index (GII), which meet this need to a significant extent, are globally important indices with an effective and academic infrastructure to determine the innovation levels of countries. This study includes regression tree analysis and linear regression analysis and comparison using the indicators within the dimensions below the subscales of the GII score and

1 **Sorumlu Yazar:** Merve Doğruel (Dr. Öğr. Üyesi), Nişantaşı Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Endüstri Mühendisliği Bölümü, İstanbul, Türkiye. E-posta: merve.dogruel@nisantasi.edu.tr ORCID: 0000-0003-2299-7182

2 Seniye Ümit Fırat (Prof. Dr.), İstanbul Gedik Üniversitesi, Mühendislik Fakültesi, Endüstri Mühendisliği, İstanbul, Türkiye. E-posta: ümit.firat@gedik.edu.tr ORCID: 0000-0002-0271-5865

Atıf: Dogruel, M., & Firat, S. U. (2021). Veri Madenciliği Karar Ağaçları Kullanarak Ülkelerin İnovasyon Değerlerinin Tahmini ve Doğrusal Regresyon Modeli ile Karşılaştırmalı Bir Uygulama. *Istanbul Business Research, 50*(2), 465-493. http://doi.org/10.26650/ibr.2021.50.015019

NRI index based on supervised machine learning. The regression tree application aimed to make the GII estimation based on the NRI indicators and determine the best discriminating GII indicators. Therefore, the Classification and Regression Tree (CART) algorithm is used for analysis. The analysis result determined the indicators within the scope of NRI that are used in the GII scores and country ranking estimation. Linear regression analysis was performed with the same data set, and the regression tree obtained by the CART algorithm was compared with the linear regression model.

Keywords
Networked Readiness Index, Innovation, Decision Tree Learning, Global Innovation Index, Classification and Regression Tree

Extended Summary

In Industry 4.0, which is built on the digital revolution infrastructure, countries are expected to adapt to the characteristics of this age in terms of competitiveness, as with companies of all sizes. Understanding and using factors that triggered the Industrial 4.0 revolution is important to adapt to this era. The most important of these elements is information technology. Therefore, the readiness level in terms of speed, capacity, number, and quality of the equipment that operates in the global network i.e., the adequacy of information infrastructure and network systems, should be addressed. Globally, the information and communications technology (ICT)-based developments of countries were measured by the World Economic Forum and its collaboration with Network Readiness Index (NRI).

Innovation is an important factor of the Industrial 4.0 revolution, since it plays an active role in the growth of countries as a factor that strengthens people and policies and becomes the main tool in creating economic prosperity. Globally, the innovation levels of countries are measured by Global Innovation Index (GII) that was prepared in collaboration with Cornell University, INSEAD, and the World Intellectual Property Organization.

Decision trees in data mining are learning algorithms based on supervised machine learning. The decision tree algorithm that is chosen in this study, like other learning techniques, aimed to create the most appropriate model from the training data. Afterward, the validity of the model created with the training data is evaluated by the test data and used to predict the approved model.

The data set of the research consists of NRI indicator values and GII scores of 123 countries in 2016. This study aimed to construct a classification model and linear regression equation to estimate GII scores by using the NRI indicators. Thus, NRI indicator values are used to determine the level of innovation of countries. Therefore, firstly, decision trees analysis, one of the classification techniques, was used as a predictive data mining task. In the decision tree analysis stage, regression tree analysis was applied, which is rarely encountered in the literature. The regression tree was created by the Classification and Regression Tree (CART) algorithm in R software programming. To develop the predictive model, 0.60 of the data set

was used as a training set for the learning stage of the algorithm, and samples were chosen by cross-validation. The optimal tree was reached by post pruning and by selecting the model that gives least-squares deviation.

Secondly, NRI indicators were determined as independent variables and the overall GII score as the dependent variable to construct an equational cause-effect model. Then, linear regression analysis was applied to the data set using the Statistical Package for the Social Sciences software package. Principal Component Analysis (PCA) was applied to independent variables before linear regression application for dimension reduction. Additionally, all linear regression analysis assumptions were checked using related tests and graphical tools to detect violations on preconditions.

The root node was determined as "ICT Patent Cooperation Treaty (PCT) patent applications per million populations" and the terminal node was determined as "PCT patent applications per million populations." to estimate the GII according to the optimal tree. These two variables with high factor loadings are included in factor 1 extracted from PCA. Contrarily, factor 1 has the most effect on predicting the dependent variable GII in linear regression analysis. According to these findings, the CART algorithm provides more in-depth information and linear regression provides more superficial information. Further, a more easily interpretable result was obtained by regression tree analysis. In these respects, estimation with the CART algorithm has superior aspects compared to linear regression analysis.

Veri Madenciliği Karar Ağaçları Kullanarak Ülkelerin İnovasyon Değerlerinin Tahmini ve Doğrusal Regresyon Modeli ile Karşılaştırmalı Bir Uygulama

Ekonomik refah yaratmanın temel aracı olan inovasyon; daha geniş çapta iklim değişikliği ile mücadelede katkı sağlamakta, sürdürülebilir kalkınmayı tetiklemekte ve sosyal uyumu da teşvik etmektedir (Gault, 2018, p. 617). Endüstri 4.0 Devrimi'ni üçüncü devrimin bir uzantısı olarak kabul eden görüşler de olmakla birlikte, yaşanan bu yeni evrimi diğerlerinden çok farklı kılan üç özelliğe dikkat çekilmektedir. i-Hız: Bu devrim öncekilerin tersine doğrusal olmayıp, üstel bir hızla gelmektedir. Çok yönlü küresel dünyanın, yeni teknolojilerin sürekli daha yeni ve daha üstün yetenekli teknolojileri üretmesi ile ilgilidir. ii- Genişlik ve Derinlik: Bu devrim dijital teknolojiler alt yapısı üzerine inşa edilmektedir ve iş dünyasında, toplumun diğer kesimlerinde, bireylerde benzeri görülmemiş paradigmalar ile ilerlemekte, hızlı teknolojik değişimler ve donanımlar geliştirmektedir. iii- Sistem Etkisi: Endüstri 4.0 Devrimi, tüm yönetim bilişim sistemleri yanında her şeyin her şeye bağlantılı olabildiği bir ağ sistemi üzerinden gelişirken, dünyada koskocaman sistemlerin, bütünleşik bir dönüşümünü kapsamaktadır. Bu üç özellik aslında kökeninde inovasyona dayanmaktadır. Günümüzde inovasyon ve onunla etkileşim içinde olan tüm alanlar büyük önem kazanmaktadır (Fırat ve Fırat, 2017a).

İnovasyon sürekli olarak dış çevreyi ve yaşam tarzlarını değiştirdiğinden, tedarik zincirleri, topluluklar, kuruluşlar, kurumlar, bölgeler ve ülkeler için sürdürülebilirlikte kilit unsurdur. Aslında literatür, inovasyon odaklı yaklaşımlara dayanarak sürdürülebilirliğin ele alınması gerektiğini kabul etmektedir. Hatta literatürde, sürdürülebilir kalkınmanın çevresel ve sosyal boyutlarına etkilerinin düşük ve yüksek olması temel alınarak geleneksel inovasyon, sosyal inovasyon, sürdürülebilir inovasyon ve yeşil inovasyon olmak üzere tipleri ele alınmaktadır (Silvestre ve Trc 2019). Küresel sorunlara bu kadar geniş yelpazede çözüm arayan inovasyon, özellikle son 50 yılda sürdürülebilir küresel rekabet kavramı bakımından da son derece önemli hale gelmiştir.

Günümüzün rekabetçi dünyasında, hem gelişmiş hem de gelişmekte olan ülkelerin küresel zorluklara karşı ortak yenilikçi çözümler bulması ve eşzamanlı olarak kendi nüfuslarının acil ihtiyaçlarını da karşılaması gerekmektedir. İnovasyon, ulusal sınırları aşarak insanı ve politikaları güçlendiren bir unsur olup, ülkelerin büyümesinde kritik bir faktördür.

Artan küresel bağlantı eğilimi, bireysel, toplumsal, bölgesel ve küresel her düzeyde sorunları çözebilme yeteneği ve standartlaştırılmış bir yol gerektirmektedir. Anahtar göstergeler yoluyla inovasyon verilerinin ölçülmesi ve analiz edilmesi mümkün olabilmektedir. Küresel İnovasyon Endeksi (KİE) 2007'den bu yana inovasyon kabiliyetlerine ve sonuçlarına göre, diğer önemli parametreler yanında 2016 yılında; patent uygulamaları, eğitim harcamaları, yaratıcı ürünlerin ihracatı ve diğer uluslararası boyutların ölçütlerini içeren 82 gösterge kullanarak dünya ekonomilerini sıralamaktadır.

Diğer yandan yeni ve güçlü seçenekler ile dijital, biyolojik ve fiziksel teknolojileri bir araya getiren, yeni sistemler setine geçişi temsil eden Dördüncü Endüstri Devrimi'ni yaşamaktayız. Bu yeni sistemler dijital devrimin altyapısı üzerine inşa edilmektedir. Endüstri 4.0'ın hem alt yapısı hem de araçları bilişim teknolojilerine dayanmaktadır. Bu nedenle küresel ağ da faaliyet gösterebilecek hız, kapasite, sayı ve donanım kalitesinde "hazır olma" düzeyi, yani bilişim altyapısı ve ağ sistemlerinin yeterliliği önem kazanmaktadır (Fırat ve Fırat, 2017b). Küresel Bilgi Teknolojileri Raporu 2016 (Global Information Technologies Report 2016), ülkelerin gelişmekte olan teknolojilerin nimetlerinden faydalanma konusunda hazır olma durumu ile dijital devrimin ve ötesinin sunduğu fırsatları değerlendiren bir içeriğe sahiptir. Bu içerikte KİE, küresel inovasyon bakımından incelemeler için çok elverişlidir.

Bilgi ve İletişim Teknolojilerinin ((BİT) - Information and Communication Technologies (ICT)) gelişimi; bilgi sistemlerindeki inovasyon, yönetme yetkinliklerinin ve profesyonel yeteneklerin sürekli olarak eğitilmesi ve iyileştirilmesi konusunda etkilidir (Kowal ve Paliwoda-Pękosz, 2017, p. 304). BİT devriminin etmenleri AYHOE ile küresel çapta ölçülebilmektedir.

İnsan sermayesi ((İS) - Human capital (HC)) ve BİT perspektifinden inovasyon; işletmeler, hükümetler ya da sosyal topluluklar için yeni BİT bilgisi, beceriler, sosyal ve yönetimsel yetkinlikler gibi yeni yetenekler geliştirme kapasitesi anlamına gelmektedir. Bu değişimler, dünyadaki ülkelerin ve ekonomilerin inovasyon performansı hakkında ayrıntılı kriterler içeren KİE ile ölçülmektedir (Kowal ve Paliwoda-Pękosz, 2017, p. 304).

Ülkelerin inovasyon seviyelerinin belirlenmesinde en etkili göstergelerden biri kabul edilen KİE ve inovasyonu BİT temelli olarak ele alan AYHOE, akademik alt yapıya sahip, inovasyon alanındaki önemli iki indekstir. Bu iki endeks gerek kapsam, gerek işlev olarak birbirleriyle benzerlik göstermektedir ve ilişkili görünmektedir.

Her gün, exabyte'ler seviyesinde yeni veri, İnternet Protokolü ((İP) - Internet Protocol (IP)) ağları üzerinden oluşturulmakta ve taşınmaktadır. 2016 yılında dünya "zettabyte dönemi" ne geçmiştir ve küresel IP trafiği 1,1 zettabyte'ye veya 1 trilyon gigabayttan daha fazla bir kapasiteye ulaşmıştır. 2020'ye kadar küresel IP trafiğinin 2.3 zettabyte'ye ulaşacağı tahmin edilmektedir. Bu veri büyümesi ekonomileri körüklemekte ve yaratıcılık dalgaları oluşturarak inovasyonu tetiklemektedir. 2016 yılı Küresel Bilgi Teknolojileri Raporu, küresel yeniliği teşvik etmek için teknolojinin ve özellikle de geniş bantın rolünü vurgulamaktadır (World Economic Forum, INSEAD ve Cornell University, 2016). İnternet ağ yapıları olmadan hiçbir inovasyonun gerçekleşmesi mümkün görünmemektedir. IP ağları; her bir kişiyi, her ülkeyi ve her IP özellikli cihazı bağlama kapasitesine sahiptir. Küresel ağlar, verilerin üretimden süreçlere kadar pek çok alanda engelsiz, hızlı bir şekilde büyümesini ve işbirlikçi inovasyonu mümkün kılmalarını sağlamaktadır. Dijital aktiviteyi teşvik etmekte donanımlı olan ülkeler, yeni sektörlerin ortaya çıkmasına ve geleneksel sektörlerin hızla gelişmesine katkıda bulunmaya devam etmektedir.

Donanımların, yazılımların ve hizmetlerin rolü; hükümetler, işletmeler ve bireyler için daha da kritik öneme sahiptir ve bu nedenle, yüksek hızlı geniş bantlı IP ağları, günlük yaşamın bir parçası haline gelmiştir. Aslında, 2020 yılına kadar 26 milyardan fazla internet bağlantılı cihaz ve 4 milyardan fazla küresel internet kullanıcısı olacağı tahmin edilmektedir. Genişbant internet, sosyal yapıları ve tüm ekonomileri önemli ölçüde etkileme yeteneği ile dünyanın en önemli genel amaçlı teknolojilerinden biri olarak kategorize edilmektedir (World Economic Forum ve ark., 2016). Bunlar dikkate alındığında, inovasyonun ağ sistemleri ile ilişkisi daha açık olarak ortaya çıkmaktadır. Ağ sistemleri, gerçek zamanlı, hızlı ve çok büyük miktarlarda veri akışları sağlamakta ve toplanan/biriken veriler, depolamadan analize, özetlemeden modellemeye kadar bilgi işlemenin her aşamasında geleneksel teknikler dışında, yeni metodolojilere gereksinimi arttırmaktadır. Veri madenciliği bu ihtiyaçların itici gücü ile son yılların en popüler alanlarından biri olmuştur.

Veri madenciliği, veriler içinde gizli kalmış, bilinmeyen, ilginç ilişkileri keşfetmeye yarayan bir prosedürdür ve tahmin etme konusunda da başarılı bir yaklaşım olarak hem bilimsel çalışmalarda, hem de endüstriyel uygulamalarda yaygın olarak kullanılmaktadır (Purohit ve Sharma, 2017). Özellikle büyük verinin gündeme gelmesiyle birlikte çığ gibi büyüyen yığınlar arasından faydalı ve değerli bilgiyi bulup çıkarmak ve kullanıma sunmak için veri madenciliği teknikleri ve algoritmaları vazgeçilmez araçlardır. Veri madenciliği dijital dönüşüm çağının hızı, depolama hacimlerindeki büyüme ve çeşitlenen ham veri karşısında güçlü analiz teknikleri sunmaktadır.

Veri madenciliği, tanımlayıcı (descriptive) ve tahminleyici (predictive) olmak üzere iki ana kategori de toplanabilen işlevler için kullanılmaktadır. Bu işlevleri yerine getirmek için; veri madenciliğinin kümeleme (clustering), birliktelik kuralı madenciliği (association rule mining), ve sınıflandırma (classification) gibi çok çeşitli ve farklı görevleri bulunmaktadır (Agarwal, Mittal ve Pareek, 2016). Bu üç görev alanında da sayısız algoritma yer almaktadır.

Literatürde, en etkili 10 veri madenciliği algoritmaları arasında SRA da yer almaktadır (Wu ve diğerleri, 2008). Araştırmacılar arasında bu görüş yaygın olmakla birlikte SRA'nın regresyon ağacı uygulamalarına son derece nadirdir. Bu çalışmada; veri madenciliği sınıflandırma algoritmalarından karar ağaçları uygulanmış ve doğrusal regresyon analizi uygulaması sonuçları ile karşılaştırılma yapılmıştır. Karar ağaçları için, literatürde sınıflandırma ağacına göre daha az çalışılmış olan regresyon ağacı analizi uygulaması gerçekleştirilmiştir. Böylece regresyon ağacı analizine ilişkin uygulama eksikliğine katkıda bulunmak ve sonuçlarının doğrusal regresyon ile değerlendirilmesi hedeflenmiştir. Araştırmada son elli yılda küresel çapta ele alınan inovasyon kapsamında, ülkelerin inovasyon seviyelerinin belirlenmesinde en önemli indekslerden olan KİE için tahminleme çalışması yapılmıştır. KİE, teması "Dijital Ekonomide İnovasyon (Innovating in the Digital Economy)" olan BİT temellerine dayanan AYHOE göstergeleri kullanılarak tahminlenmiştir. Günümüzde küresel çerçevede en çok ta-

kip edilen konulardan biri inovasyon performanslarıdır. Bu araştırmada farklı algoritma ve teknikler ile çeşitli göstergeler kullanarak inovasyon performansları için tahmin modelleri sunulmaktadır.

SRA algoritması kullanılarak oluşturulan regresyon ağacı modelinde KİE tahminlemesini yaparken, en iyi ayrılmayı sağlayan AYHOE göstergelerinin belirlenmesi de amaçlanmıştır. Tahmin modelinin en uygun model olması için, en küçük hata kareleri ortalamasını veren model denemeleri yapılmış ve en düşük hata kareleri ortalaması ile KİE'yi tahmin eden optimal ağaç modeli elde edilmiştir. Elde edilen optimal SRA karar ağacı modeli ile; görsel olarak yorumlanabilen, tahmin hataları düşük ve tahminleme yorumu kolay olan bir model oluşturulmuştur. Bu model ile gelecek yıllardaki veriler kullanılarak tahminleme yapılması da öngörülmektedir. Aynı veri seti kullanılarak, doğrusal regresyon modeli de kurulmuş ve SRA algoritması kullanılarak oluşturulan model ile doğrusal regresyon analizi ile elde edilen modelin karşılaştırması yapılmıştır.

İnovasyon ve İlgili Küresel Endeksler

İnovasyonun önemi ilk kez 20. yüzyılın başlarında Schumpeter tarafından vurgulanmıştır. Farklı tanımları yapılan inovasyon kavramı için, 2005 yılında Eurostat ve OECD tarafından ortaklaşa geliştirilen The Oslo Manuel'de tüm yaklaşımlar için kullanılabilecek, ortak kabul görmüş inovasyon tanımı şu şekilde yapılmıştır: "iş uygulamalarında, iş yeri organizasyonunda ya da dış ilişkilerde; yeni veya önemli derecede geliştirilmiş bir ürünün (mal ya da hizmet) ya da sürecin, yeni bir pazarlama yöntemi ya da yeni bir organizasyonel yöntemin uygulanması" (OECD, 2005, p. 46).

Yeni teknolojilerin geliştirilmesi ve yaygınlaştırılması, resmi ve resmi olmayan ağlar ile bu etkileşimleri düzenleyen kurumsal kaynaklardan oluşan aktörler inovasyon sistemlerinin temel yapısal öğesidir. Firmalar, araştırma kuruluşları, devlet daireleri, STK'lar ve diğer aracı kuruluşlar inovasyonun gelişmesine ve yayılmasına katkıda bulunan başrol oyuncularıdır (Binz ve Truffera, 2017, p. 1286).

İnovasyon temel olarak buluş, patent, lisans, diğer fikri mülkiyet hakları, endüstriyel dizayn katogorilerini içeren fikri mülkiyeti kapsamaktadır (Mataradzija, Rovcanin ve Mataradzija, 2013).

İnovasyon küresel gündemin en önemli konularından biridir. Rekabet açısından, daha da önemlisi sürmekte olan dijital devrim (Endüstri 4.0) kapsamında ülkelerin konumlarının belirlenmesi son derece önemlidir. Bu bağlamda World Economic Forum ve diğer uluslararası kuruluşlar bu konuda yoğun çalışmalar gerçekleştirmektedir. Bu çalışmada, ülke karşılaştırmalarında en çok dikkat çeken ve içerdiği göstergeler bakımından akademik alt yapıya sahip iki küresel index (KİE, AYHOE) gözönüne alınmıştır.

Ağ Yapılara Hazır Olma Endeksi ((AYHOE) - The Networked Readiness Index (NRI))

2016 yılında Dünya Ekonomik Forum'un INSEAD ve Cornell Üniversitesi işbirliği ile yayınladığı Küresel Bilgi Teknolojileri Raporu, Ağ Yapılara Hazır Olma Endeksi (AYHOE) kullanılarak BİT devriminin etmenlerini küresel çapta ölçmektedir. 2016 yılı indeksinde 139 ülke kapsanmıştır. Küresel Bilgi Teknolojileri Raporu 2016 yılı teması "Dijital Ekonomide İnovasyon (Innovating in the Digital Economy)" olarak belirlenmiştir. "Dijital Ekonomide İnovasyon" teması ile hazırlanan Küresel Bilgi Teknolojileri Raporu 2016, dijital devrimin hem inovasyonun doğasını değiştirdiğine, hem de firmaların sürekli olarak inovasyon yapma konusunda artan bir baskıya maruz kaldığına vurgu yapmaktadır. Rapordan 4 anahtar mesaj çıkarılmıştır:

i. Dijital devrim inovasyonun doğasını değiştiriyor.

ii. Firmalar sürekli inovasyon yapmak için artan bir baskıyla karşı karşıya kalıyor.

iii. İşletmeler ve hükümetler hızla büyüyen dijital nüfusun ihtiyaçlarını karşılamakta eksik kalabiliyor.

iv. Yönetişim ve düzenlemelerde acil inovasyon isteyen yeni bir ekonomi şekilleniyor.

Bu temel sonuçlar, inovasyonu şekillendiren, tetikleyen ve karşılıklı etkileşim içinde olduğu alanın BİT olduğuna bir kez daha dikkati çekmektedir.

Küresel Bilgi Teknolojileri Raporu ilk olarak 2001 yılından yayınlanmış ve zaman içinde gelişmiştir. 2016 basımında Şekil 1'de görüldüğü gibi, AYHOE yapısını oluşturan 4 alt endeks bulunmaktadır. Alt endekslerin oluşturulmasında 10 boyut, bu boyutların içinde ise 53 gösterge bulunmaktadır (World Economic Forum ve ark., 2016).

Şekil 1. Ağ Yapılara Hazır Olma Endeksi 2016 Çerçevesi

Küresel İnovasyon Endeksi ((KİE) - The Global Innovation Index (GII))

İlk kez 2007 yılında yayınlanan Küresel İnovasyon Endeksinin (KİE), 2016 basımı Cornell Üniversitesi, INSEAD ve Dünya Fikri Mülkiyet Örgütü (World Intellectual Property Organization (WIPO)) iş birliği ile yayınlanmıştır. KİE, inovasyon faktörlerinin sürekli değerlendirildiği bir ortam yaratmaya yardımcı olmaktadır. 2016 yılı KİE modelinde 128 ülke kapsanmıştır.

Yıllar içinde KİE, herhangi bir ulusun inovasyon kapasitesinin; sadece yerel olarak ne düzeye geldiği ile değil, aynı zamanda tüm dünyayı nasıl etkilediği ile de ölçüldüğünü göstermiştir. Yoksulluk, sağlık, kentleşme, suya erişim ve iklim değişikliği gibi konular küresel niteliktedir. Ancak aynı zamanda hem zorluklar hem de çözümlerin yerel sonuçları vardır.

Bu nedenle, gelişmekte olan ülkelerde yerel çözümler sunan yenilikçi atılımlar küresel bir etkiye sahip olabilir ve diğer gelişmekte olan ülkeler arasında karşılıklı yarar sağlamak için paylaşım olanağı sağlayabilir. Bu yaklaşım içinde KİE raporunun 2016 teması "Küresel İnovasyon ile Kazanmak (Winning with Global Innovation)" olarak belirlenmiştir.

KİE dört ölçü hesaplamaktadır:

i. İnovasyon Girdi Alt Endeksi

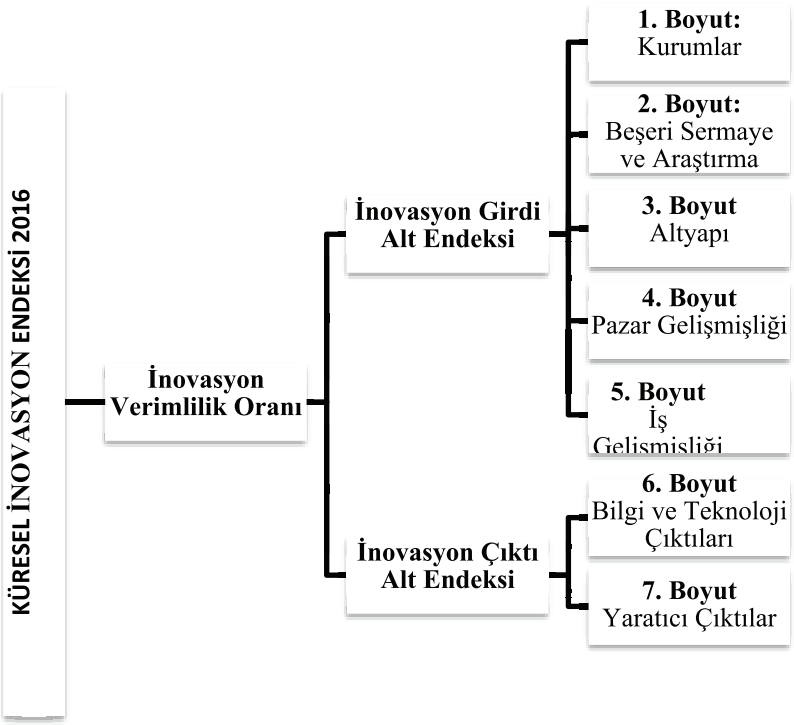
ii. İnovasyon Çıktı Alt Endeksi

iii. Genel KİE puanı

iv. İnovasyon Verimlilik Oranı

KİE puanları, girdi ve çıktı alt endeks değerlerinin basit ortalaması olarak hesaplanmaktadır. (Cornell University, INSEAD ve WIPO, 2016).

Şekil 2'de görüldüğü gibi, 2016 yılında girdi alt endeksi 5 boyuttan, çıktı alt endeksi 2 boyuttan, her bir boyut ise 3 alt boyuttan oluşmakta olup, bu alt boyutların içinde ise toplam 82 gösterge bulunmaktadır.



Şekil 2. Küresel İnovasyon Endeksi 2016 Çerçevesi

Bu çerçeve incelendiğinde, girdi alt endeksindeki 3. boyut olan "altyapı" ve çıktı alt endeksindeki 6. boyut olan "bilgi ve teknoloji çıktıları" boyutlarının tamamen BİT alanını temsil ettiği görülmektedir. Yani KİE, içinde AYHOE boyutlarına benzer ve ortak değişkenler de barındırmaktadır.

AYHOE ve KİE'nin Birlikte Değerlendirilmesi

Teknoloji, inovasyon ve bilgi son elli yılda dünya ekonomisi evrimi ve uluslararası iş geliştirmenin temelinde bulunan üç önemli kavramdır (Andersson, Das, Mudambi ve Pedersen, 2016).

BİT, gelişmiş ve gelişmekte olan ülkeler için inovasyon ve büyümeye imkan sağlayan önemli kaynaklardan biridir. BİT'nin gelişmiş pazarlarda ekonomik büyüme için en önemli kaynaklardan biri olduğu ve inovasyona olanak sağladığı gösterilmiştir (Amiri ve Woodside, 2017).

Kononova, 2013 yılında 96 ülke için KİE ile AYHOE arasında 0,94'lük bir korelasyon katsayısı hesaplamıştır (Kononova, 2015). Tabii ki bu korelasyonların anlamlı ve yüksek olmasında iki endeksin kapsadığı ortak sayılabilecek değişkenlerin de etkisi bulunmaktadır. Ancak inovasyon, teorik olarak ağ – internet yapıları gerektirdiği için bu bulgu pratik olarak açıklanabilmektedir.

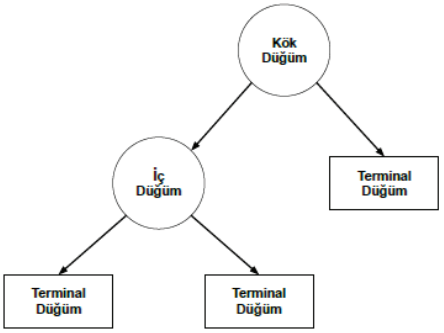
Preda ve diğerleri 2015 yılında Avrupa Birliğine bağlı 28 ülke için KİE ve AYHOE arasında tek değişkenli regresyon denklemi kurmuş ve korelasyon katsayısı olan R değerini 0,918 olarak bulunmuştur (Preda, Crişan, Stănică ve Samuel, 2016). Zoroja'nın araştırmasında ise, BİT'nin yenilikçilik üzerinde pozitif etkisi olduğu belirtilmiştir (Zoroja, 2016). "BİT olmadan inovasyon olmaz" yaygın görüşünü destekleyici sonuçlardır.

Doğruel Anuşlu ve Fırat ise, 2016 KİE ve AYHOE verileri ile yaptıkları betimsel çalışmada, ülkelerin bu iki endeks açısından sıralamalarının benzer olduğunu belirtmiştir. İnovasyonda ön sıralarda olan ülkelerin ağ yapılarının da iyi olduğuna dikkat çekilmiştir (Doğruel Anuşlu ve Fırat, 2019).

Veri Madenciliği Sınıflandırma Modelleri İçin Karar Ağacı Algoritmaları

Veri madenciliğinde bir karar ağacı, hem sınıflandırıcıları hem de regresyon modellerini temsil etmek için kullanılabilen nonparametrik bir tahmin modelidir ve araştırmalarda kararların hiyerarşik modellerini ve sonuçlarını belirtmek için kullanılmaktadır (Rokach ve Maimon, 2015).

Şekil 3'te örneklendirildiği gibi, karar ağaçları düğümler ve uçlardan oluşan hiyerarşik, yönlü bir ağaç yapısındadır. Yaprak olmayan bir düğüm iç ya da bölünmüş düğüm olarak adlandırılırken, yaprak düğüm ise terminal bir düğüm olarak adlandırılmaktadır (Guller, 2015). Ağaç yapısında iki düğüm birbirine ok ile bağlandığında, okun çıkış yaptığı düğüme ebeveyn düğüm, okun giriş yaptığı düğüme de çocuk düğüm denilmektedir.



Şekil 3. Ağaç Yapısı

Karar ağaçları oluşturulurken kullanılan tüm metotlarda tahmin edilecek olan değişken için final değerler kümesine ulaştıran "eğer-öyleyse" kurallar kümesi oluşturulmaktadır. Elde edilen final değerleri kategorik bir değişkenin olasılıkları ise yaratılan karar ağacı sınıflandırma ağacı; sürekli bir değişkenin nicelikleri ise yaratılan karar ağacı regresyon ağacı olarak nitelendirilmektedir (Putler ve Krider, 2015). Bir sınıflandırma ağacı, karar kurallarının özetini temsil eden bir algoritmadır. Bağımlı değişken kategorik olan bir hedef değişken iken, bağımsız değişkenler tahminleyicilerdir. Her iç düğüm, bir tahmine dayanan kararı temsil etmektedir. Her uç, potansiyel gelecek karara kılavuzluk etmektedir. Her yaprak, bir sınıf ile etiketlenir. Amaç; tahminleyicilerin değerlerine uygun olarak, kökten yapraklara kadar uzanan bir yol izleyerek sınıflandırma yapmaktır. Regresyon ağacı özet ağaç ile temsil edilen bir algoritmadır ancak hedef değişken, sınıf yerine gerçek bir niceliktir. Karar düğümleri sınıflandırma ağacına benzemektedir ancak her yaprak amaç değişkeni için bir nicelik ile etiketlendirilme yapılmaktadır (Khoshgoftaar, Allen ve Deng, 2005).

Veri madenciliğinde karar ağaçları, denetimli makine öğrenmesi temellerine dayanan öğrenen bir algoritmalar topluluğudur. Diğer öğrenen algoritmalarda da olduğu gibi karar ağaçları öğrenmesinde de, seçilen karar ağacı algoritması, eğitim verisinden en uygun modeli oluşturmayı hedeflemektedir. Sonrasında test verisi ile oluşturulan modelin geçerliliği sınanmakta ve eğer model onaylanırsa tahminleme yapmak için kullanılmaktadır.

Regresyon Ağacı Yapısı ve Aşamaları

Regresyon ağacının amacı, sürekli ve kategorik bağımsız değişkenlerden yararlanarak sürekli bağımlı değişkeni (hedef değişkeni) tahmin etmektir.

Regresyon ağaçlarının temelleri 1963 yılında Morgan ve Sonquist tarafından AID (Automatic Interaction Detection) algoritmasının geliştirilmesiyle atılmıştır. 1984 yılında Breinman ve diğerleri tarafından en popüler versiyon olan SRA (CART - Classification and Regression Trees) algoritması geliştirilmiştir (Tutz, 2012). Literatürde GUIDE, M5, SUPPORT, SECRET, MART, SMOTI, MAUVE, BART, SERT gibi farklı algoritmalar da bulunmaktadır (Yang, Liu, Tsoka, Papageorgiou, 2017).

Regresyon ağacı modelleri oluşturma ve kullanma süreçleri, üç temel algoritmik alt görev içermektedir (Parziale ve ark., 2016, p. 193):

i. Regresyon ağacı büyütme

ii. Regresyon ağacı budama

iii. Regresyon ağacı tahmin etme

SRA Algoritması İle Regresyon Ağacı

SRA (Sınıflandırma ve Regresyon Ağacı) algoritması hem sınıflandırma hem de regresyon ağacı oluşturmak için kullanılan, tekrarlanan ikili ayırmaya göre tahmin yapan, parametrik olmayan (non-parametrik) ve doğrusal olmayan (non-linear) bir karar ağacı algoritmasıdır. Hedef değişken kategorik ise oluşturulan ağaç bir sınıflandırma ağacı ((SA) (Classification Tree, CT)), hedef değişken sürekli ise oluşturulan ağaç bir regresyon ağacı ((RA) (Regression Tree, RT)) adını almaktadır.

SRA, regresyon analizindeki varsayımların sağlanmadığı durumlar için regresyon analizinin bir alternatifidir. Ayrıca SRA, veri setinin karmaşık bir yapıya sahip olduğu durumlarda bile bağımlı değişkeni etkileyen bağımsız değişkenlerin belirlenmesi ve bu değişkenlerin modeldeki önemlerini göstererek birbirleri arasındaki ilişkilerin anlaşılabilir bir görsellikte sunmasından dolayı da regresyon analizinin alternatifi olarak kullanılmaktadır (Ceyhan, 2014).

SRA'nın en önemli avantajlarından biri de eksik veriler olduğunda dahi tamamen otomatik ve etkili bir mekanizma ile çalışabilmesidir (Kuzey, 2012).

Bir regresyon ağacı oluşturmak için, SRA algoritmasının süreci aşağıdaki gibi özetlenebilir (Yohannes ve Webb, 1999; Sumathi ve Paneerselvam, 2010):

1- SRA, kök düğümden başlayarak bağımsız değişkenlerin her birinde olası tüm ayrılmaları gerçekleştirir, her ayrılma önceden tanımlanmış düğüm safsızlık ölçütü (node impurity measure) uygular.

2- Elde edilen safsızlıktaki azalmayı belirler.

3- SRA daha sonra, ayrılma uyum kriterlerini (goodness-of-split criteria) uygulayarak en iyi ayrılmayı gerçekleştirir ve veri setini sağ-sol çocuk düğümlere ayırır.

4- SRA özyinelemeli olduğu için, terminal olmayan her bir düğüm için 1'den 3'e kadar olan adımları tekrarlar, mümkün olan en büyük ağacı üretir.

5- Son olarak SRA, elde edilen ağaca budama algoritmasını uygular.

SRA ile Regresyon Ağacı Büyütme

Regresyon ağacını büyütmek için, her adımda girdi değişkenlerinden birisi örnekleri ayırmak için seçilmektedir. Seçilen değişken süresince ayrılma noktasına nitelik değer testi (attribute value test) uygulanmakta ve iç düğümün sonraki düğümlere bölünmesi için en iyi ayrılma noktası belirlenmektedir (Kim ve Hong, 2017). SRA, regresyon ağaçlarının oluşturulmasında ayrılmaişlemleri için en küçük kareler sapması (least squares deviation) veya en küçük mutlak sapma (least absolute deviation) ölçülerini kullanmaktadır (Kuzey, 2012).

Ağacı büyütürken amaç; tahmini çıktılar ile gerçek çıktılar arasında daha küçük hatalar elde etmek için girdi alanını bölmektir. Genel olarak tahmini çıktılar, bir terminal düğümünden alınan eğitim örneklerinin gerçek çıktılarının ortalaması kullanılarak aşağıdaki şekilde saptanmaktadır (Kim ve Hong, 2017):

$$\widehat{y_i} = \frac{\sum_{j \epsilon t_i} y_j}{|t_i|}$$

t_i: i yaprak düğümü

$|t_i|$: i yaprak düğümündeki örnek sayısı

Ayrılma kriteri en küçük kareler sapması safsızlık ölçüsüne dayanır.

$$I(t_i) = \sum_{j \epsilon t_i} (y_j - \hat{y}_i)^2$$

$I(t_i)$ i düğümündeki safsızlık ölçüsü

En küçük kareler sapması kullanarak ayrılma kriteri aşağıdaki gibi hesaplanır (Kim ve Hong, 2017):

$$\Delta I = I(t_p) - P_l I(t_l) - P_r I(t_r)$$

t_p ebeveyn düğüm ve t_l ile t_r, t_p'nin iki çocuk düğümünü, P_l ve P_r ise sırasıyla sağ ve sol çocuk düğümlere atanan örneklerin oranlarıdır. Ayrılma noktası ΔI maksimuma çıkarmak için belirlenir.

Nümerik ya da ordinal değişken kullanarak ayrılma kuralı üretilirse ve çocuk düğüm sayısı iki ise, ebeveyn düğümdeki örnekler $\{x{:}x_k > s\}$ ve $\{x{:}x_k \leq s\}$ olarak iki alt kümeye ayrılır. Burada x_k seçilen değişkeni, s ise ayrılma noktasını ifade etmektedir. Nominal tahminleyiciler için de aynı yaklaşım kullanılır fakat q kategorili sırasız kategorik tahminleyiciler için 2^{q-1} mümkün ayrım bulunmaktadır (Kim ve Hong, 2017, p. 40).

Ardışık bir sıra izleyen süreç, durdurma kuralı uygulanmamışsa, homojenlik kriterleri gerçekleşip maksimum ağaç olana kadar ya da bazı durdurma kuralları uygulanana kadar devam etmektedir (Ceyhan, 2014).

SRA ile Regresyon Ağacı Budama

Eğitim verisi ile öğrenme yapılırken, ağaç yapısı aşırı büyük bir şekilde oluşturulursa, ağacın her bir yaprağı tek bir eğitim durumunda olan, sıfır hataya sahip bir ağaç modeli oluşturulur. Özellikle küçük örnekler ile çalıştığında model, daha önce karşılaşılamayan durumlara karşı neredeyse hiç genelleme yapamamakta ve dolayısıyla tahminler doğru olmamaktadır. Bu durum, eğitim verisine aşırı uyum (overfitting) olarak bilinir. Bu problemi minimize

etmek için ağacın büyütmesini durdurmada kullanılan ya ön-budama (pre-pruning) olarak bilinen budama kuralları ya da ağacı büyütüldükten sonra yapılan son-budama (post-pruning) yaklaşımı uygulanmaktadır (Soman, Diwakar ve Ajay, 2009).

SRA algoritmasında karmaşıklığı hesaba katmak için popüler bir çözüm, karmaşıklık için açık bir ceza içeren, maliyet karmaşıklığı (cost complexity) olarak adlandırılan bir işlev tanımlamak olmuştur (Berk, 2016). Minimize edilmeye çalışılan hata karmaşıklığı ölçüsü (the error-complexity measure) T ağaç için sınıflandırma hatasının toplam maliyeti ve karmaşıklık için ceza olmak üzere iki kısımdan oluşmaktadır (Sutton, 2005):

$$R_\alpha (T) = R(T) + \alpha|T|$$

R(T): T ağaç için sınıflandırma hatasının toplam maliyeti

|T|: terminal düğümlerin sayısı

α: her bir terminal düğüme uygulanan ceza değeri

α değeri sıfıra eşit veya sıfırdan büyük bir değerdir. Eğer α = 0 ise, ceza değeri yoktur, maliyet karmaşıklığı maksimum seviyededir ve doymuş bir ağaçtır. Eğer α değeri büyütülürse, R(T) değerini azaltan ağacın altında kalan ayrımlar kesileceği için maliyet karmaşıklığı azalacaktır (Kuzey, 2012).

Budama hakkındaki yeni çalışmalarda α, cp ile yer değiştirmektedir (Berk, 2016).

$$R_{cp} (T) \equiv R(T) + cp|T|R(T_1)$$

T_1: ayrılmamış bir ağaç

|T|: bir ağaç için ayrılmaların sayısı

R: ağacın riski

Bir T ağacı için, K terminal düğümündeki genel risk ise aşağıdaki gibidir (Berk, 2016):

$$R(T) = \sum_{j=1}^{K} P(A_j)\, R(A_j)$$

Bu değer, her düğümle ilişkili riskin terminal düğümler üzerinden toplamıdır.

Cp değeri 0 ile 1 arasındadır. Cp=0 olduğunda, doymuş bir ağaç varken, cp=1 olduğunda ayrılma yoktur.

Bağımsız test verisi veya çapraz doğrulama (cross-validation) kullanılarak farklı aday ağaçları arasından optimum büyüklükte ağaç seçilmektedir (Cho ve Kurup, 2011). Eğer veri

seti yeteri kadar büyük değilse, hesaplama karmaşıklığına rağmen çapraz doğrulama yönteminin kullanılması önerilmektedir (Maimon ve Rokach, 2005).

SRA ile Regresyon Ağacı Tahmin Etme

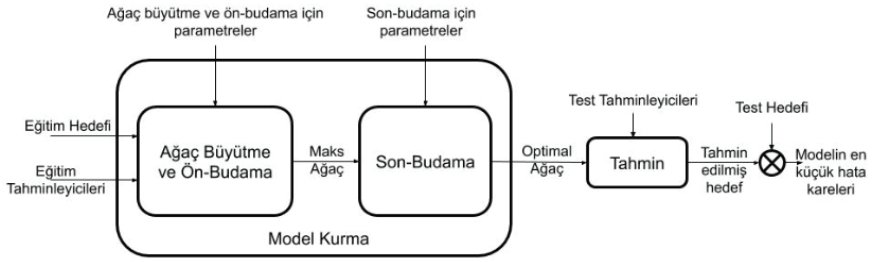
Optimal ağaç seçildikten sonra SRA, her bir terminal düğüm için özet istatistikler hesaplamaktadır. Eğer ayırma kuralı olarak en küçük kareler sapması seçilmiş ise SRA, bağımlı değişkenin ortalamasını ve standart sapmasını hesaplamaktadır. Terminal düğümün ortalaması, bu terminal düğümdeki durumlar için bağımlı değişkenin tahmin edilen değeri olmaktadır. Eğer en küçük mutlak sapma seçilmişse, SRA bağımlı değişkenin medyanını ve mutlak ortalama sapmalarının ortalamasını üretmektedir. Terminal düğüm için, medyan bağımlı değişkenin tahmin edilen değeri olmaktadır (Yohannes ve Webb, 1999).

Uygulama: Küresel İnovasyon Endeksleri İle Tahmin Çalışması

Uygulama için, ilk aşamada SRA algoritması ile optimum regresyon ağacı elde edilmesi, sonra doğrusal regresyon analizi çalışması yapılması ve son aşamada opimum regresyon ağacı ile doğrusal regresyon analizinin karşılaştırılması hedeflenmiştir.

SRA Algoritması İle Regresyon Ağacı

Uygulama aşamalarında SRA algoritması kullanılarak regresyon ağacını oluşturma ve kullanma süreçleri için Şekil 4'teki akış geliştirilerek uygulama bu akış planı doğrultusunda gerçekleştirilmiştir. Uygulamada ilk amaç AYHOE göstergelerini tahminleyici olarak kullanarak hedef değişken KİE'ni tahminlemektir. Diğer yandan, bu tahmin yapılırken en iyi ayrılmayı sağlayan AYHOE göstergelerinin belirlenmesi amaçlanmaktadır. Bu iki amacı bir arada sağlayacak en uygun modelin, SRA analizi ile oluşturulabileceği belirlenmiştir.



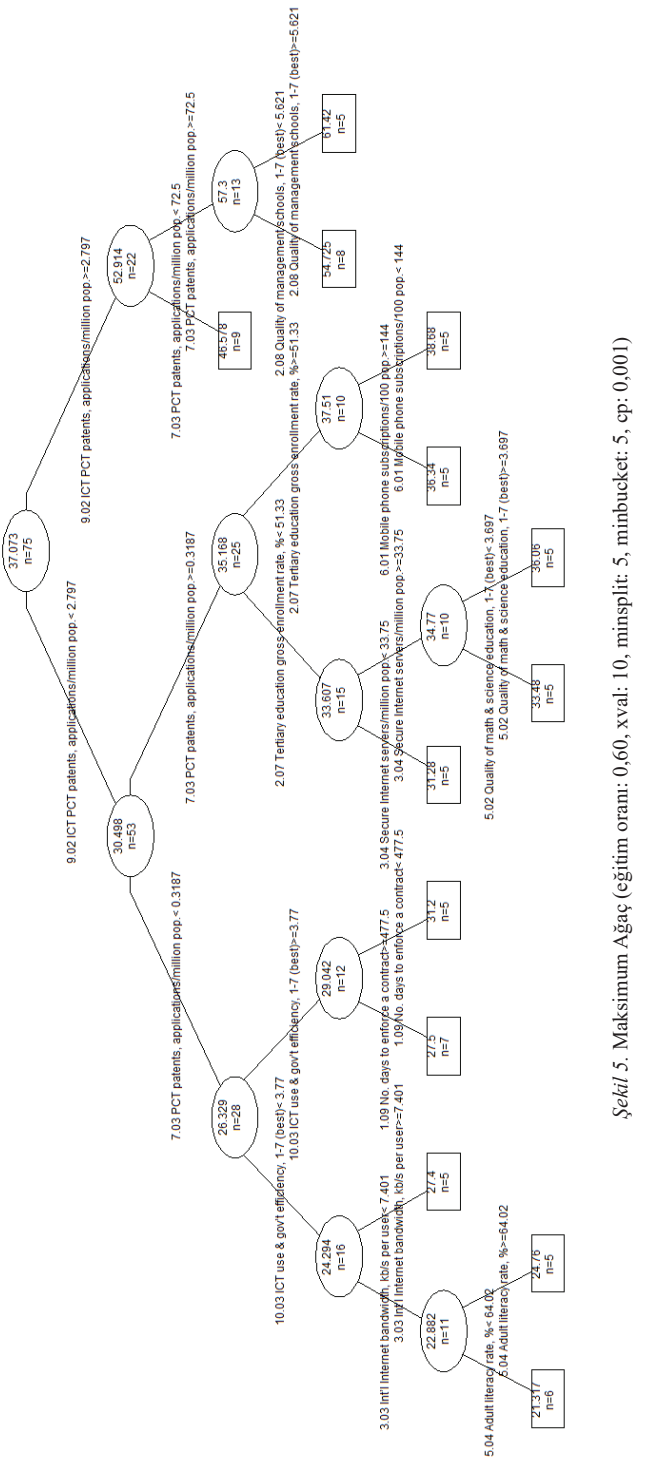
Şekil 4. SRA ile Regresyon Ağacının Blok Diyagramı

Analizin uygulanması için R programlama dilinden yararlanılmıştır. Veri seti 2016 yılına ait; AYHOE göstergelerinden oluşan tahminleyiciler ile hedef değişken olan KİE'den oluşturulacaktır. Veri ön işleme sürecinde R dilinde; sadece 2016 yılına ait olan ilgili verilerin seçilmesi, AYHOE göstergelerinin KİE ile eşleşmesini sağlamak amacıyla transpoze işleminin uygulanması, AYHOE gösterge verilerinde ve KİE verisinde farklı şekilde yazılmış ancak aynı ülkeye ait olan verilerin kullanılabilmesi için ülke isimlerinin eşleştirilmesinin sağlanması, AYHOE göstergelerine ait veride olan ancak KİE verisinde olmayan 26 ülkenin ve KİE verisinde olan ancak AYHOE göstergelerine ait veride bulunmayan 3 ülkenin belirlenerek kapsam dışında tutulması, AYHOE gösterge verilerinin ve KİE verisinin ülke isimlerine göre birleştirilmesi ve göstergeleri tamamen boş verilerden oluşan ülkelerin kapsam dışında tutulması gibi çeşitli işlemler yapılmıştır. Sonuçta 123 ülkenin; AYHOE gösterge verilerine ait 53 tahminleyici değişken ve 1 hedef değişken (KİE) olmak üzere 54 değişkenden oluşan bir veri seti elde edilmiştir.

Modelin öğrenme yapabilmesi için; veri setinin 0,60'ı eğitim kümesi olarak kullanılmış ve geriye kalan 0,40'ı ise test verisi olarak ayrılmıştır.

SRA algoritması ile maksimum regresyon ağacını büyütmek için R programında "rpart" kütüphanesi kullanılmıştır. İlgili rpart kütüphanesi indirildikten sonra, ağacı büyütmek için gerekli argüman değerleri rpart fonksiyonuna tanımlanmıştır. Bu amaçla çapraz doğrulama sayısı olan "xval" değeri 10, düğümde bulunması gereken minimum gözlem sayısı değeri olan "minsplit" değeri 5, herhangi bir terminal düğümde bulunması gereken minimum gözlem sayısı değeri olan "minbucket" değeri 5 ve ön-budama yapılması için karmaşıklık parameter değeri olan "cp" değeri ise 0,001 olarak belirlenmiştir. Cp değeri belirlenirken ağacın çok kompleks olmamasına ancak ayrılmaların da optimum düzeyde belirlenmesine olanak sağlayacak bir değer olmasına özen gösterilmiştir.

Belirtilen argümanlarla büyütülen ağaçta, hedef değişken olan KİE tahmininde kullanılan en büyük ağaç için ayrılmayı sağlayan 12 iç düğüm ve 13 terminal düğümden oluşan Şekil 5'teki yapı elde edilmiştir.



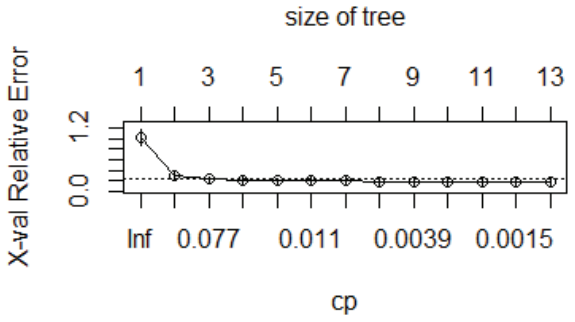
Şekil 5. Maksimum Ağaç (eğitim oranı: 0,60, xval: 10, minsplit: 5, minbucket: 5, cp: 0,001)

Belirlenen argüman değerlerine göre oluşturulan en büyük ağaçtan, optimal boyuttaki ağacı bulmak için de son-budama uygulaması yapılacaktır. Son-budama işleminin yapılması için elde edilen maksimum ağacın en küçük çapraz doğrulama hata değerine sahip cp değerinin seçilmesi gerekmektedir.

Tablo 1 ve Şekil 6'da görüldüğü gibi 0,00312 cp değeri ile küçük çapraz doğrulama hata değeri elde edilmiştir.

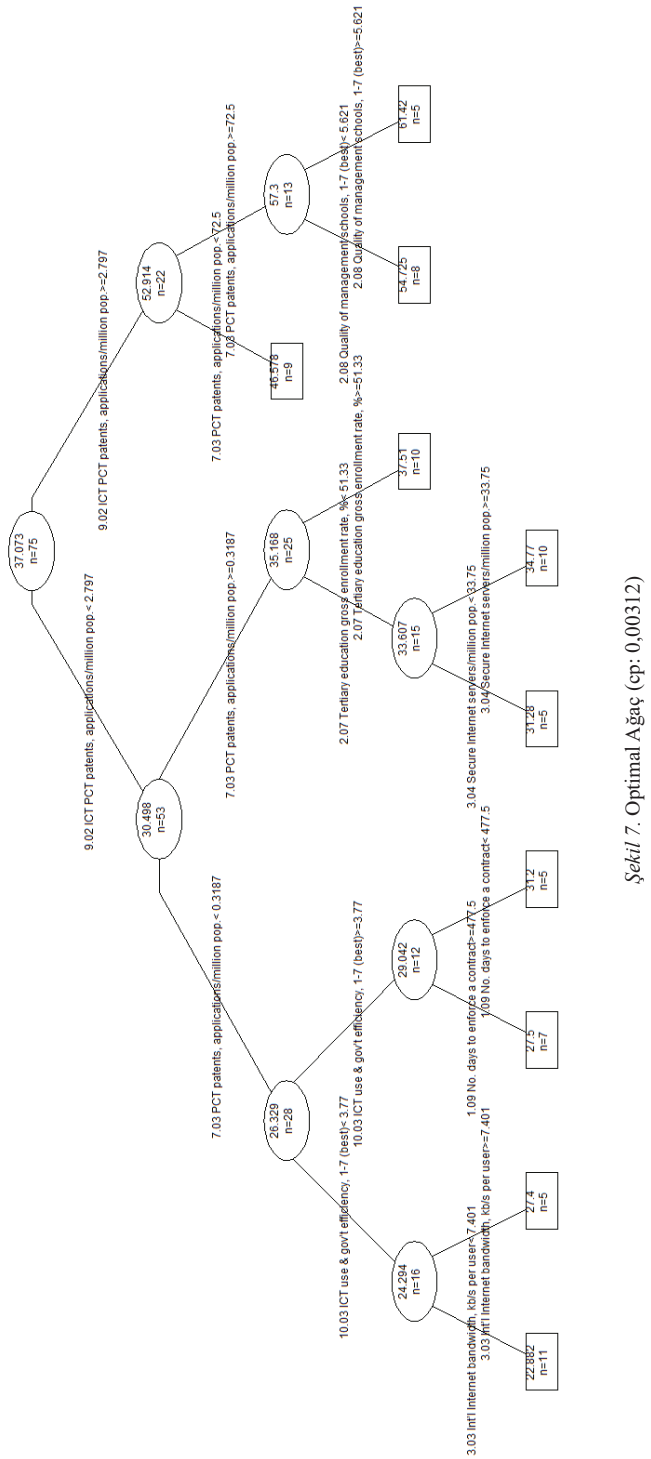
Tablo 1
Çapraz Doğrulama Hataları

	CP	Nsplit	gör.hata	xhata	xssapma
1	0,75311	0	1	1,01565	0,14136
2	0,09949	1	0,24689	0,30702	0,04912
3	0,05895	2	0,14740	0,23740	0,04296
4	0,01490	3	0,08845	0,21487	0,04949
5	0,01330	4	0,07355	0,20895	0,04800
6	0,00881	5	0,06025	0,21170	0,04896
7	0,00677	6	0,05144	0,20485	0,05855
8	0,00391	7	0,04467	0,19666	0,05793
9	0,00385	8	0,04076	0,19687	0,05791
10	0,00312	9	0,03691	0,19478	0,05781
11	0,00160	10	0,03379	0,19623	0,05812
12	0,00132	11	0,03219	0,19530	0,05815
-13	0,00100	12	0,03087	0,19723	0,05826



Şekil 6. Çapraz Doğrulama Hataları

0,00312 cp değeri ile Şekil 7'deki optimal ağaç elde edilmiş olup, optimal ağaçta ayrılmayı sağlayan 9 iç düğüm ve 10 terminal düğüm bulunmaktadır.



Şekil 7. Optimal Ağaç (cp: 0,00312)

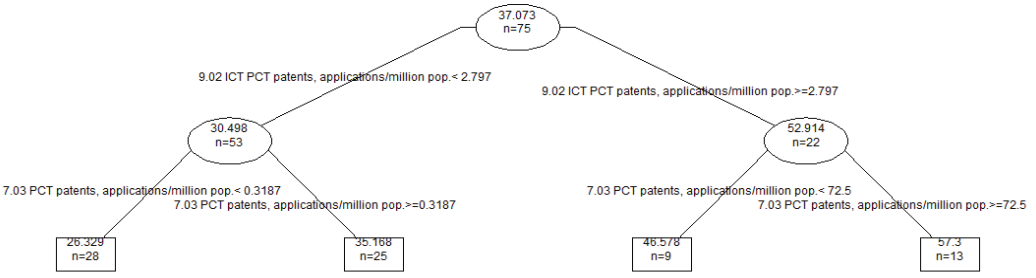
Elde edilen optimal ağaç ile test veri kümesi kullanarak KİE değerleri tahmin edilmiştir. Tahmin edilen KİE değerleri ile gerçek KİE değerlerini kıyaslamak için hata kareleri ortalaması hesaplanmış ve bu değer 15,927 olarak bulunmuştur.

En uygun modelin oluşturulması amacıyla, Tablo 2'de yer alan daha düşük hata kareleri ortalamasının olup olmadığını araştırmak için farklı eğitim küme yüzdeleri ve xval değerleri için (diğer argümanlar sabit tutularak) model denemeleri yapılmıştır.

Tablo 2
Farklı Eğitim Yüzdeleri ve XVAL Değerleri İçin Kurulan Modellerin Hata Kareleri Ortalaması

Eğitim/ XVAL	**5**	**10**	**15**	**20**	**25**
0,6	26,345	15,927	*15,550	19,111	15,883
0,7	23,496	23,496	25,604	22,787	22,780
0,8	24,985	24,985	25,215	26,884	26,426

Yapılan denemelerde en düşük hata kareleri ortalamasını veren 0,60 oranındaki eğitim kümesi yüzdesi ve 15 çapraz doğrulama sayısıyla maksimum ağacın büyütülmesine karar verilmiştir. Bu maksimum ağacın en küçük çağraz doğrulama hata değerine sahip cp değeri (0,0149) ile elde edilen optimal ağaç ise Şekil 8'de gösterilmiş olup, tahmin yapmak için bu model kullanılacaktır.



Şekil 8. Optimal Ağaç (cp: 0,0149)

Şekil 8'deki optimal ağaca göre KİE tahmin etmek için kök düğüm "milyon nüfus başına Bilgi İletişim Teknolojileri (BİT) Patent İşbirliği Anlaşması (PİA) patent başvuruları (ICT PCT patent applications per million population)" ve terminal düğüm "milyon kişi başına düşen Patent İşbirliği Anlaşması (PİA) patent başvuruları (PCT patent applications per million population)" olarak belirlenmiştir.

Doğrusal Regresyon Analizi

123 ülkenin AYHOE gösterge değerleri olan 53 bağımsız ve KİE olan 1 bağımlı değişkenden oluşan veri setinde, 13 bağımsız değişken için eksik gözlemler olduğu belirlenmiş

ve bu değerler değişken ortalamaları ile tamamlanmıştır. Araştırmanın veri setinde satır/sütun sayısının matris işlemleri için regresyon analizi yapılmasına uygun olmaması nedeniyle, AYHOE'ne ait 53 bağımsız değişken, ait oldukları alt endeksler bazında toplanarak yeni değişkenler tanımlanmıştır. Böylece bağımsız değişken sayısı 10'a indirgenmiştir.

10 bağımsız ve 1 bağımlı değişkenden oluşan veri setine ait korelasyon matrisi incelendiğinde, 10 bağımsız değişken arasında oldukça yüksek ve anlamlı korelasyonlar olduğu görülmüş, yani çoklu doğrusal bağlantı problemi ile karşılaşılmıştır. Çoklu doğrusal bağlantı problemine çözüm üretebilmek ve daha yorumlanabilir bir regresyon modeli elde etmek için, 10 bağımsız değişken ile Temel Bileşenler Analizi (TBA - Principal Component Analysis (PCA)) uygulaması yapılmıştır. Kaiser-Meyer-Olkin (KMO) testi (0,783>0,5) sonucuna göre örneklemin yeterli olduğu saptanmıştır. Bartlett test sonucu da anlamlı bulunarak (sig. =,000) korelasyon matrisin birim matristen anlamlı bir şekilde farklı olduğu sonucuna ulaşılmıştır. Bu iki testin anlamlı bulunmasıyla TBA'nın bu veri seti için uygulanabilir olduğu kabul edilmiştir. Ayrıca anti-image matrisi kontrol edilmiş ve tüm anti-image korelasyonları da anlamlı bulunmuştur. Dirsek yöntemi (elbow method - scree plot) göre de yatay şekil alan noktanın 2 faktör olduğu kabul edilmiş ve varimax rotasyonu ile elde edilen toplam açıklayıcılık yüzdesinin 2 faktör için %63,747 olduğu saptanmıştır. Rotasyon ile elde edilen bileşen matrisi (rotated companent matrix) Şekil 9'da verilmiş olup, incelendiğinde; 10 değişkenden 7'sinin (sosyal etkiler, kamu yönetiminin kullanımı, ekonomik etkiler, bireysel kullanım, işletmelerin kullanımı, altyapı) 1. faktörde, 3'ünün ise (iş ve inovasyon çevresi, erişilebilirlik, beceriler) 2. faktörde olduğu görülmektedir. Elde edilen faktörler incelenerek, 1. faktöre "bireysel ve çevresel etkiler" , 2. faktöre ise "inovasyonel ve gelişimsel etkiler" adı verilmiştir.

Rotated Component Matrix[a]

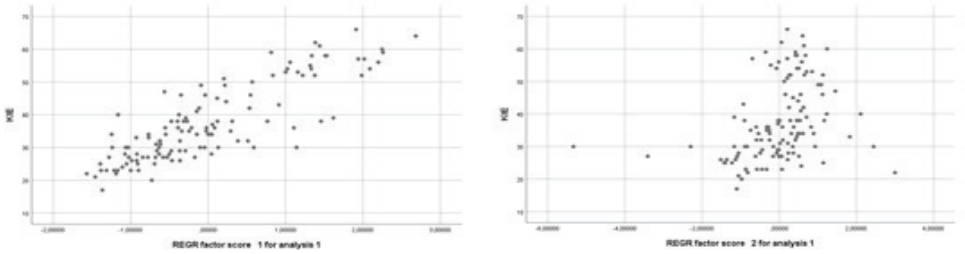
	Component	
	1	2
Sosyal_Etkiler	,920	,090
Kamu_Yonetiminin_Kullanimi	,844	-,260
Ekonomik_Etkiler	,795	,279
Bireysel_Kullanim	,788	,471
Isletmelerin_Kullanimi	,760	,228
Altyapi	,682	,223
Siyasi_ve_Duzenleyici_Cevre	-,545	,187
Is_ve_Inovasyon_Cevre	,042	,759
Erisilebilirlik	,022	-,675
Beceriler	,544	,617

Şekil 9. Birleştirilmiş Bileşen Matrisi

Doğrusal regresyon analizi uygulaması için bireysel ve çevresel etkiler olarak tanımlanan faktör 1 puanları ile inovasyonel ve gelişimsel etkiler olarak tanımlanan faktör 2 puanları bağımsız değişkenler, KİE ise bağımlı değişken olarak belirlenmiştir. Regresyon modelinin kurulmasından önce ise doğrusal regresyon analizine ilişkin çoklu doğrusal bağlantı probleminin olmaması, doğrusallık, hataların eşit varyanslılığı, hataların bağımsızlığı ve hataların normal dağılması varsayımlarının geçerliliği araştırılmıştır.

Faktör analizi yapılarak çoklu doğrusal bağlantı probleminin (multicollinearity) ortadan kalkabileceği bilinmekle birlikte (Fvero ve Belfiore, 2019), Stepwise yöntemi ile regresyon modeli kurulmuş ve daha düşük hata kareleri ortalaması veren modelde iki bağımsız değişken de kalmıştır.

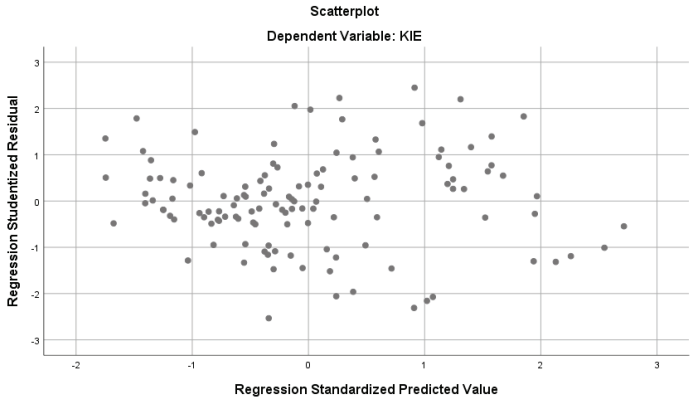
Doğrusallık varsayımının araştırılması için, bağımlı değişken ile bağımsız değişkenler arasında Şekil 10'daki serpilme grafikleri çizilmiş ve doğrusal ilişkiler olduğu kabul edilmiştir. Şekil 10 (a)'da KİE ile bireysel ve çevresel etkiler arasında, Şekil 10 (b)'de KİE ile inovasyonel ve gelişimsel etkiler arasındaki serpilme grafiklerikleri görülmektedir. Ayrıca bağımlı değişken ile bağımsız değişkenler arasında korelasyon katsayıları hesaplanmış ve anlamlı (sig.=0,000) bulunmuştur.



10 (a) 10 (b)

Şekil 10. Doğrusallık İçin Serpilme Grafikleri

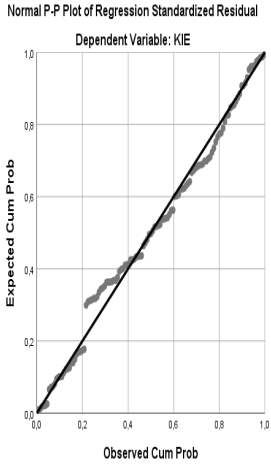
Hataların eşit varyanslılığının (homoscedasticity) kontrolü için standardize tahmini değerleri (ZPED) ile Student (t-dağılım) dağılmış hatalar (SRESID) arasında Şekil 11'deki serpilme grafiği çizilmiş ve bir örüntü olmadığı saptandığı için bu varsayımın da geçerliliği sağlanmıştır.



Şekil 11. Eşit Varyanslılık İçin Serpilme Grafiği

Hatalar arasında ilişki olmaması, bir diğer deyişle hataların bağımsızlığı yani otokorelasyon olmaması varsayımının araştırılması için Durbin-Watson testi yapılmış ve sonuç 1,801 bulunmuştur. %5 anlamlılık seviyesinde Durbin-Watson tablosunda bağımsız örnek sayısı (k) 2 için, örnek birim sayıları 100 ve 150 kabul edilerek alt ve üst sınır değerleri kontrol edilmiş ve elde edilen 1,801 değerinin otokorelasyon olmayan bölgede olduğu saptanmıştır. Dolayısıyla otokorelasyon olmadığı varsayımı kabul edilmiştir.

Hataların normalliği varsayımının araştırılması için ise Şekil 12'de görülen standardize edilmiş hataların normal olasılık grafiği (normal probability plot - NPP) çizilmiş ve elde edilen grafik ile hataların normal dağıldığı kabul edilmiştir.



Şekil 12. Normallik İçin P-P Grafiği

Regresyon modelinin geçerliliği için gerekli tüm varsayımlar sınanmış ve kabul edilmiştir. Varsayımların kabulünden sonra regresyon modeli kurulmuş ve modelin denklemi KİE=37,350 + 9,991 Bireysel ve çevresel etkiler + 3,992 İnovasyonel ve gelişimsel etkilerolarak belirlenmiştir. Model anlamlı (sig.=0,000) olup, R^2=0,833 olarak bulunmuştur. Bireysel ve çevresel etkiler bağımsız değişkenine ait kısmi korelasyon katsayısı 0,901 iken, inovasyonel ve gelişimsel etkiler değişkenine ait kısmi korelasyon katsayısı ise 0,638'dir.

Tartışma ve Sonuç

KİE değerlerini, AYHOE aracılığı ile SRA algoritması kullanarak tahminlemeyi amaçlayan uygulamalı bölümde, nihai olarak elde edilen optimal ağaca dayanarak, tahminleyici olarak önerilebilecek değişkenler belirlenmiştir. KİE tahmin etmek için en etkili olan iki tahminleyicinin "milyon nüfus başına Bilgi İletişim Teknolojileri (BİT) Patent İşbirliği Anlaşması (PİA) patent başvuruları" ve "milyon kişi başına düşen Patent İşbirliği Anlaşması (PİA) patent başvuruları" göstergeleri olduğu görülmektedir. Kök düğümü oluşturan "milyon nüfus başına BİT PİA patent başvuruları" tahminleyicisi, AYHOE'yi oluşturan "Etki Alt Endeksi" içinde bulunan 9. boyut olan "ekonomik etki" göstergelerinden biridir. Diğer anlamlı tahminleyici olan "milyon kişi başına düşen PİA patent başvuruları" ise AYHOE'yi oluşturan "Kullanım Alt Endeksi" içinde yer alan 7. boyut olan "işletmelerin kullanımı" göstergelerinden biridir.

SRA regresyon ağacı algoritması ile geliştirilen modelin bulgularına dayanarak ulaşılan sonuçlar aşağıdaki gibi özetlenebilir. Modelde KİE'yi tahmin ederken ayrılmayı en iyi sağlayan değişken "milyon nüfus başına BİT PİA patent" başvuruları değişkenidir. Optimal ağaç modeline göre 4 tane temel kural oluşturulabilmektedir:

KİE'yi tahmin ederken ayrılmayı en iyi sağlayan değişken "milyon nüfus başına BİT PİA patent" başvuruları değişkenidir. Optimal ağaç modeline göre 4 tane temel kural oluşturulabilmektedir:

i- Eğer; milyon nüfus başına BİT PİA patent başvuruları < 2,797 ise ve milyon kişi başına düşen PİA patent başvuruları < 0,319 ise KİE = 26,329.

ii- Eğer; milyon nüfus başına BİT PİA patent başvuruları <2,797 ise ve milyon kişi başına düşen PİA patent başvuruları ≥ 0,319 ise KİE = 35,168.

iii- Eğer; milyon nüfus başına BİT PİA patent başvuruları ≥ 2,797 ise ve milyon kişi başına düşen PİA patent başvuruları < 72,469 ise KİE = 45,578.

iv- Eğer; milyon nüfus başına BİT PİA patent başvuruları ≥ 2,797 ise ve milyon kişi başına düşen PİA patent başvuruları ≥ 72,469 ise KİE = 57,300.

Oluşturulan optimum ağaç modeli kullanılarak 123 ülke için tahmin edilen KİE değerleri ve gerçek KİE değerleri (2016 yılı için) EK-1'de bulunmaktadır. Göz ile incelemede farklar küçük görünse de bir istatistiki metrik ile doğrulamak için korelasyon analizi yapılması uygun görülmüştür. KİE'nin tahmin edilen ve gerçek değerler arasında hesaplanan Pearson korelasyon katsayısı 0,95 (alfa= 0.01) bulunmuş olup son derece yüksek ve anlamlıdır. Bu iki AYHOE göstergesinin etkin olduğu model kullanarak tahmin edilecek KİE değerleri gerçek değerlere son derece yakın olmuştur. Bu çalışmada ulaşılan, araştırma bulgularına dayanarak "Bu iki göstergenin KİE değerlerini tahminlemede kullanılmasının çok uygun olacağı" şeklindeki önerimizi güçlü kılmaktadır. Diğer yandan kullanılan algoritmanın ve elde edilen optimal ağacın metriklerinin de son derece iyi ve geçerli bulunmuş olması, uygulamada ulaşılan kavramsal anlamlılığın güvenilirliğini arttırıcı niteliktedir.

Doğrusal regresyon modeli uygulamasında ise, veri setine 53 değişken ayrı ayrı dahil edilememiş, bağımsız değişkenler faktör analizi uygulamasından elde edilen iki faktör ile regresyon analizi gerçekleştirilmiştir. 7 değişkenden oluşan faktör 1'in (bireysel ve çevresel etkiler) kısmi korelasyon katsayısı (0,901) faktör 2'nin (inovasyonel ve gelişimsel etkiler) kısmi korelasyon katsayısından (0,638) çok daha yüksektir. Faktör 1'in KİE üzerinde, faktör 2'ye göre daha fazla etkisi olduğu açıkça görülmektedir.

Doğrusal regresyon analizi elde edilen sonuç ile SRA algoritması ile edilen optimum ağaç karşılaştırıldığında, optimum ağaçta yer alan kök düğüm (milyon nüfus başına BİT PİA patent başvuruları) ve terminal düğümün (milyon kişi başına düşen PİA patent başvuruları) her ikisinin de doğrusal regresyonda KİE üzerinde daha etkili olduğu belirlenen faktör 1'de yer aldığı görülmektedir. Ancak doğrusal regresyonda KİE üzerinde daha etkili olduğu belirlenen "bireysel ve çevresel etkiler" adlı faktörün, SRA algoritması ile elde edilen optimum ağaçta daha derinlemesine bir bilgi edinecek şekilde hangi göstergeler (milyon nüfus başına BİT PİA patent başvuruları ve milyon kişi başına düşen PİA patent başvuruları) ile iyi bir tahmin vereceği de belirlenebilmiştir. SRA algoritması ile ayrıca görsel olarak da yorumlanması kolay bir sonuç elde edilmiştir. Bu açılardan SRA algoritması ile tahminlemenin, doğrusal regresyon analizine göre açıklama ve yorumlama açısından daha üstün yönleri olduğunu söylemek mümkündür. Aynı veri seti için, faktör analizi sonrası kullanılabilen doğrusal regresyon analizi, bazı değişkenlerin bireysel bazda katkılarını görebilmeyi kısıtlamaktadır. Ayrıca doğrusal regresyon analizinin varsayımların test edilmesi gibi ek analizler gerektirmesi, SRA algoritması lehine bir durum ortaya koymaktadır.

En etkili on veri madenciliği algoritması arasında gösterilen SRA'nın, regresyon ağacı sınıflandırması literatürde çok nadir olarak yer almaktadır. Bu çalışmanın hem algoritmanın kullanımı hem de inovasyon göstergeleri tahminleme konusundaki ampirik araştırmalar arasında ilklerden olması nedeniyle literatüre önemli bir katkı sağlayacağı düşünülmektedir.

Sonraki çalışmalarda ve yıllık raporlarda, modeldeki AYHOE göstergeleri aynı kalmak koşulu ile AYHOE veri seti kullanılarak ve önerilen modeli uygulayarak SRA algoritması ile KİE değerlerinin başarılı bir şekilde tahmin edilebileceği öngörülmektedir.

Hakem Değerlendirmesi: Dış bağımsız.
Çıkar Çatışması: Yazarlar çıkar çatışması bildirmemiştir.
Finansal Destek: Yazarlar bu çalışma için finansal destek almadığını beyan etmiştir.
Yazar Katkısı: Çalışma Konsepti/Tasarımı: S.Ü.F., M.D.; Veri Toplama: M.D.; Veri Analizi /Yorumlama: S.Ü.F., M.D.; Yazı Taslağı: M.D.; İçeriğin Eleştirel İncelemesi: S.Ü.F.; Son Onay ve Sorumluluk: S.Ü.F., M.D.

Peer-review: Externally peer-reviewed.
Conflict of Interest: The authors have no conflict of interest to declare.
Grant Support: The authors declared that this study has received no financial support.
Author Contributions: Conception/Design of study: S.Ü.F., M.D.; Data Acquisition: M.D.; Data Analysis/Interpretation: S.Ü.F., M.D.; Drafting Manuscript: M.D.; Critical Revision of Manuscript: S.Ü.F.; Final Approval and Accountability: S.Ü.F., M.D. Accountability: E.K.P., S.A.Ş., B.H

Kaynakça/References

Agarwal, R., Mittal, M. & Pareek, S. (2016). Loss profit estimation using temporal association rule mining. *International Journal of Business Analytics, 3(1),* 45-57.

Amiri, S. & Woodside, J. M. (2017). Emerging markets: The impact of ICT on the economy and society. *Digital Policy, Regulation and Governance*, *19*(5), 383-396.

Andersson, U., Das, ., Mudambi, R. & Pedersen, T. (2016). Technology, innovation and knowledge: The importance of ideas and international connectivity. *Journal of World Business*, *51*, 153-162.

Berk, R. A. (2016). *Statistical learning from a regression perspective*, (2nd ed.). Cham, Switzerland: Springer International Publishing.

Binz, C. & Truffera, B. (2017). Global innovation systems - A conceptual framework for innovation dynamics in transnational contexts. *Research Policy, 46*, 1284-1298.

Ceyhan, G. (2014). *Üniversite öğrencilerinin yansıtıcı düşünme düzeyleri ve araştırmaya yönelik kaygılarının çeşitli değişkenler açısından CART analizi ile incelenmesi*. (Yüksek Lisans Tezi). Yüzüncü Yıl Üniversitesi Eğitim Bilimleri Enstitüsü, Van.

Cho, J. H. & Kurup, P. U. (2011). Decision tree approach for classification and dimensionality reduction of electronic nose data. *Sensors and Actuators B: Chemical*, *160*, 542-548.

Cornell University, INSEAD & WIPO. (2016). *The Global Innovation Index 2016: Winning with Global Innovation*. Ithaca, Fontainebleau and Geneva.

Doğruel Anuşlu, M. ve Fırat, S. Ü. (2019). Endüstri 4.0 ve sürdürülebilirlik etkileşimi: Küresel endesklerle değerlendirmeler. İçinde E. S. Bayrak Meydanoğlu, M. Klein, ve D. Kurt (Edler). Dijital dönüşüm trendleri (ss 56-100). Vefa, İstanbul: Filiz Kitapevi.

Fvero, L. P. & Belfiore, P. (2019). *Data science for business and decision making*. United Kingdom, UK: Academic Press.

Fırat, O. Z. ve Fırat, S. Ü. (2017a). Endüstri 4.0 yolculuğunda trendler ve robotlar. *Istanbul University Journal of the School of Business, 46-2, 211-223.*

Fırat, S. Ü. ve Fırat, O. Z. (2017b). Sanayi 4.0 Devrimi üzerine karşılaştırmalı bir inceleme: Kavramlar, küresel gelişmeler ve Türkiye. *Toprak İşveren Dergisi, 114*, 10-23.

Gault, F. (2018). Defining and measuring innovation in all sectors of the economy. *Research Policy, 47,* 617-622.

Guller, M. (2015). *Big data analytics with spark: a practitioner's guide to using spark for large scale data analysis.* New York, NY: Apress

Khoshgoftaar, T. M., Allen E. B. & Deng, J. (2005). Using regression trees to classify fault-prone software modules. In D. Zhang & J. J. P. Tsai (Eds.), *Machine learning application in software engineering* (pp. 87-94). 5 Toh Tuck Link, Singapore: World Scientific Publishing Co. Pte. Ltd.

Kim, K. & Hong, J. (2017). A hybrid decision tree algorithm for mixed numeric and categorical data in regression analysis. *Pattern Recognition Letters*, *98*, 39-45.

Kononova, K. (2015). Some aspects of ICT measurement: Comparative analysis of e-indexes. In *Proceedings of the 7th International Conference on Information and Communication Technologies in Agriculture, Food and Environment (HAICTA 2015)*. Kavala, Greece.

Kowal, J & Paliwoda-Pękosz G. (2017). ICT for global competitiveness and economic growth in emerging economies: Economic, cultural, and social innovations for human capital in transition economies. *Information Systems Management, 34(10),* 304-307.

Kuzey, C. (2012). *Veri madenciliğinde destek vektör makinaları ve karar ağaçları yöntemlerini kullanarak bilgi çalışanlarının kurum performansı üzerine etkisinin ölçülmesi ve bir uygulama*. (Doktora Tezi). İstanbul Üniversitesi İşletme Anabilim Dalı Sayısal Yöntemler Bilim Dalı, İstanbul.

Maimon, O. & Rokach, L. (2005). Decision tree. In O. Maimon & L. Rokach (Eds.), *The data mining and knowledge discovery handbook* (pp. 165-192). New York, NY: Springer Science+Business Media, Inc.

Mataradzija, A., Rovcanin, A. & Mataradzija, A. (2013). Innovation and innovative performance in the European Union. In *Proceedings of the Management, Knowledge and Learning International Conference.* Bangkok, Thailand; Celje, Slovenia; Lublin, Poland: ToKnowPress.

Silvestre, B. S. & Trc, D. M. (2019). Innovations for sustainable development: Moving toward a sustainable future. *Journal of Cleaner Production, (208)*. 325-332.

Organisation For Economic Co-operation and Development (OECD). (2005). *Oslo Manual: Guidelines for collecting and interpreting innovation data (*3rd ed.*)*. Paris, France: OECD Publishing.

Parziale, L., Benke, O., Favero, W., Kumar, R., Lafalce, S., Madera, C. & Muszytowski, S. (2016). *Enable real-time analytics on IBM z systems platform*. Retrieved from http://www.redbooks.ibm.com/redbooks/pdfs/sg248272.pdf

Preda, A., Crișan, D. A., Stănică, J. L. & Samuel, A. N. A. (2016). Transectional analysis between innovation and ICT readiness for the european union countries. *Journal of Information Systems & Operations Management, 10*(2)*,* 393-403.

Purohit, S. K. & Sharma, A. K. (2017), Development of data mining driven software tool to forecast the customer requirement for quality function deployment. *International Journal of Business Analytics, (4) (1),* 56-86.

Putler, D. S. & Krider, R. E. (2015). *Customer and business analytics: Applied data mining for business decision making using R*. Boca Raton, FL: CRS Press.

Rokach, L. & Maimon O. (2015). *Data mining with decision trees: Theory and applications*, (2nd ed.). 5 Toh Tuck Link, Singapore: World Scientific Publishing Co. Pte. Ltd.

Soman, K. P., Diwakar, S. & Ajay, V. (2009). *Data mining: Theory and practice*. Patparganj Industrial Area, Delhi: PHI Learning Private Limited.

Sumathi, S. & Paneerselvam, S. (2010). *Computational intelligence paradigms: Theory & applications using MATLAB*. Boca Raton, FL: CRS Press.

Sutton, C. D. (2005). Classification and regression trees, bagging, and boosting. In C. R. Rao, E. J. Wegman & J. L. Solka (Eds). *Data mining and data visualization* (pp. 303-328). Amsterdam, The Netherlands: Elsevier B.V.

Tutz, G. (2012). *Regression for categorical data.* New York, NY: Cambridge University Press.

World Economic Forum, INSEAD & Cornell University. (2016). *The Global Information Technology Report 2016: Innovating in the Digital Economy*. Geneva, Fontainebleau and Ithaca.

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q. & Motoda, H. (2008). Top 10 algorithms in data mining, Knowl Inf Syst, 14, 1-37.

Yang, L., Liu, S., Tsoka, S. & Papageorgiou, L. G. (2017). Regression tree approach using mathematical programming. *Expert Systems With Applications*, *78*, 347-357.

Yohannes, Y. & Webb, P. (1999). *Classification and regression trees, CART: A user manuel for identifying indicators of vulnerabilty to famine and chronic food insecurity*. New York, NY: International Food Policy Research Institute.

Zoroja, J. (2016). Impact of ICTs on innovation activities: Indication for selected european countries. *Naše gospodarstvo/Our Economy*, 62(3), 39-51.