



Ekonomik Gelişmişliği Etkileyen Bilgi Ekonomisi Değişkenlerinin Veri Madenciliği ile Belirlenmesi¹

Mehpare TİMOR *, Gonca YÜZBAŞI KÜNÇ**

ÖZ

Küreselleşmenin etkisiyle geçmişten günümüze biriken veri yığınları veri madenciliği tekniklerinin uygulama alanlarını genişletmiştir. Ülkelerin ekonomik olarak gelişme seviyelerini etkileyen faktörler bilgi çağında değişmiştir. Teknoloji çağında ekonomik gelişmeler ve ilerlemeler bilgi ekonomisini ortaya çıkarmıştır. Bilgi ekonomisinin temeli bilginin elde edilip işlenmesine ve dağıtılmasına kadar olan süreci kapsamaktadır. Genel ekonomik göstergeleri etkileyen bilgi ekonomisi değişkenleri daha önemli hale gelmiştir. Bu bağlamda, çalışmamızda öncelikle AB üye ülkeleri ve Türkiye, kümeleme analizi ile genel ekonomik göstergeler açısından gruplara ayrılmıştır. Ekonomik seviyelerine göre gruplara ayrılan ülkelere bilgi ekonomisi değişkenleriyle birlikte karar ağacı uygulanmıştır. Veri madenciliği uygulaması sonucu, ülkelerin ekonomik seviyelerini belirleyen bilgi ekonomisi değişkenleri tespit edilmiştir.

Anahtar Kelimeler: Veri Madenciliği, Bilgi Ekonomisi, Kümeleme Analizi, K-Ortalamalar, Ekonomi

JEL Sınıflandırması: C55, C81, C82

Determination of Knowledge Economy Variables that Affect Economic Development by Using Data Mining

ABSTRACT

The accumulated data from past to present expanded the application areas of data mining techniques with the effect of globalization. Factors affecting the economic development levels of countries have changed in information age. Economic developments and advances in the technology age have revealed the information economy. The basis of the knowledge economy covers the process from the acquisition of information to the distribution. In technology age, knowledge economy variables that affect general economic indicators have become more important. In this context, in our study, firstly, EU member countries and Turkey were divided into groups in terms of overall economic indicators with cluster analysis. The countries, which are divided into groups according to their economic levels, have been implemented a decision tree with information economy variables. As a result of data mining application, knowledge economy variables that determine the economic levels of countries were determined.

Keywords: Data Mining, Knowledge Economy, Cluster Analysis, K-means, Economy

JEL Classification: C55, C81, C82

*Geliş Tarihi / Received: 05.06.2020 Kabul Tarihi / Accepted: 15.08.2020
Doi: 10.17541/optimum.748237*

¹ Bu çalışma İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Sayısal Yöntemler Bilim Dalında sunulan “Veri Madenciliği Tekniklerinin Bilgi Ekonomisi Değişkenlerine Uygulanması” başlıklı doktora tezinden türetilmiştir.

* Prof. Dr., İstanbul Üniversitesi, İşletme Fakültesi, İşletme Bölümü, timorm@istanbul.edu.tr, ORCID: 0000-0002-9782-545X.

** Araş. Gör. Dr., Adıyaman Üniversitesi, İİBF, İşletme Bölümü, gyuzbasi@adiyaman.edu.tr, ORCID: 0000-0003-0213-7310.

1. GİRİŞ

Bilginin en basit formu farklı çevrelerden gelen insanlar tarafından bile yanlış anlaşılma riski en düşük olan ve bu nedenle açıkça kodlanabilen ve tamamen açık olabilen formudur. Ülkeler ekonomik ve sosyal açıdan ilerlemelerini kuvvetlendirmek için bilgiye daha çok önem vermek, daha etkili şekilde ürün ve hizmet üretmek, geliştirdikleri üretim yöntemlerinin bireylere düşük seviyede maliyetlerle erişmesini sağlamak gibi faaliyetleri hayata geçirmek zorunda kalmaktadırlar. Son yıllarda bilgi unsuru değer elde etmek için kullanılan önemli bir kaynak olarak görülmektedir. Ekonomi geçmişten günümüze bireyler, kurumlar, ülkeler vb. birçok taraf açısından farklı aşamalardan geçerek değişime uğramıştır. Bu ekonomik değişimin temelinde ise her zaman bilgi faktörü yer almıştır. Tarımla başlayan ekonomik ilerleme sanayi devrimiyle devam etmiştir ve günümüzde bilgiye dayalı yeni ekonomi anlayışıyla gelişimini sürdürmektedir. Ekonomik büyüme artık bilgiden bağımsız yönetilememektedir. Ekonomide klasik üretim faktörleri olan hammadde, sermaye, işgücü bilgiye dayalı tolumun ilerlemesiyle yetersiz kaldığı görülmüştür. Bu durum toplumun gelişmesi için üretim faktörleri arasında en çok katkı sağlayacak bilginin de girmesini sağlamıştır. Ülkelerin gelişmişlik seviyeleri açısından hangi bilgi ekonomisi değişkenlerinin belirleyici olduğu ise günümüzde hayati önem taşımaktadır.

1960'lı yıllarda bilgisayarların kullanılmasıyla elde edilen veriler kaydedilmeye başlanmıştır. 1980'li yıllarda biriktirilmiş verilere ulaşmak için veritabanları geliştirilmiştir. 1990'lı yıllara gelindiğinde ise bu veritabanları çoklu seviyelerde ilişkilendirilmeye başlanmıştır ve dinamik hale getirilmiştir. Geçmişte kaydedilen verilerin biriktirilmesiyle, bu verilerden anlamlı bilginin çıkarılması ihtiyacı doğmuştur. Günümüzde ise bu verilerden matematiksel ve istatistiksel yöntemlerle geleceğe yönelik anlamlı bilgiler elde edilmesini sağlayan veri madenciliği teknikleri geliştirilmiştir. Veri madenciliği sorunları çözmeye yardım eden tanımlayıcı ve tahmin edici modeller meydana getirerek ve veri analiz ederek işlemektedir. Veri madenciliği tekniklerinden olan karar ağaçları sınıflandırma amacıyla kullanılmaktadır. Ayrıca değişkenlere göre birimleri sınıflandıran bir yöntemdir. Kümeleme analizi ise nesnelerin benzerlik ya da farklılıklara göre alt gruplara ayrılmasını amaçlamaktadır.

Veri madenciliği teknikleri pazarlama, bankacılık, borsa, finans vb. gibi alanlarda sık kullanılmaktadır. Ekonomi alanında veri madenciliği uygulamalarına sık rastlanmamaktadır. Bu çalışmada benzer konularda yapılmış ulusal ve uluslararası çalışmalar incelenerek Avrupa Birliği üye ülkelerinin ve Türkiye'nin genel ekonomik gruplanmaları kümeleme analizi ile belirlenmiştir. Veri madenciliği tekniklerinden karar ağaçları algoritması kullanılmasıyla, kümeleme analizi ile elde edilen ülke gruplarını birbirinden ayıran bilgi ekonomisi değişkenleri belirlenmiştir.

1.1. Literatür Taraması

Bilgi ekonomisi ile ilgili literatürde ayrı ayrı birçok çalışma yer almaktadır. Bu çalışmalar aşağıdaki tabloda yer almaktadır (Tablo 1).

Tablo 1: Bilgi Ekonomisi Değişkenleri İle İlgili Çalışmalar

No	Yazar	Dönem (*)	Yöntem	Bilgi Ekonomisi Değişkenleri
1	Yapraklı S, Sağlam T	1980-2008	Granger Nedensellik Testi	Telekomünikasyon yatırımları, GSMH, İstihdam
2	Yeo BJ	1988-2007	Regresyon ve Korelasyon Analizi	İnovasyon kapasitesi, İktisadi büyüme
3	Driouchi A vd.	1995-2001	Zaman Serileri Analizi	Ekonomik teşvikler, Eğitim yatırımları, İnovasyon oranı
4	Yeloğlu HO	1995-1999	Kümeleme Analizi	Bilgi ve iletişim teknolojilerine ait üretim miktarları, Bilgisayar ve ilgili aktiviteler kısmına ait katma değer, İstihdam değişkeni, Bilgi ve iletişim sektöründe çalışan bireylere ödenen ücretlerde hesaplanan yüzde değişim, Bilgi ve iletişim teknolojileri sektöründe kullanılan örgüt sayılarına ait değişim verileri
5	Andrés AR vd.	1996-2010	Panel veri analizi	Bilimsel yayın sayısı, Bilgi ve iletişim teknolojileri kullanımı, inovasyon sayısı, Okullaşma oranı
6	Asongu SA	1996-2010	Zaman serileri analizi	Eğitim seviyesi, Bilgi altyapısı, İnovasyon sistemleri sayısı, Kurumsallaşma oranı
7	Işık N, Kılınç EC	2000-2010	Panel Veri Analizi	Patent başvuru, BİT sermayesi, BİT ürünleri ihracatı, İleri teknoloji ihracatı
8	Oktay E, Kaynak S	2005	Kanonik Korelasyon Analizi	Bilgi, iletişim ve teknoloji harcamaları, Kamu Ar-Ge harcamaları, İşletmelerin Ar-Ge harcamaları, Fen ve mühendislik bölümü mezunları sayısı, Kamu eğitim harcamalarının GSYİH içindeki yeri, İnsani gelişme endeksi, Yüksek teknoloji ürünlerinin ihracat içerisindeki payı, Amerikan Patent Enstitüsü'ne kayıtlı patent sayısı, Bilgi yoğun hizmetlerde çalışanların toplam istihdama oranı, Cep Telefonu aboneliği sayısı, İnternet erişimli işletmelerin yüzdesi, İnternet kullananların sayısı
9	Kaynak S, Arslan I	2005-2010	Panel Veri Analizi	Teşvik rakamları, Eğitim verileri, İnovasyon ve bit yatırımları
10	Raspe O, Van Oort F	2006	Çoklu Regresyon	Bilgi çalışanları sayısı, İnovasyon sayıları, Ar-Ge yatırımları
11	Yaylalı M vd.	2007	Veri Zarflama Analizi	BİT Harcamaları, Kamu Ar-Ge Harcamaları, Fen ve mühendislik mezun sayısı.
12	Adaçay FR	2007	Karşılaştırmalı Analiz	Ar-Ge yoğunluğu, Ar-Ge istihdamı, Bilimsel yayın sayısı, İleri teknoloji ihracatı, Patent başvuru sayısı, Ticari marka sayısı, Bilgi ve iletişim teknolojileri yatırımları
13	Berberoğlu B.	2010	Kümeleme Analizi, Diskriminant Analizi	Eğitilmiş insan sayısı, Bit harcamaları, Bit yetenekli kişilerin istihdam oranı
14	Yam RCM vd.	2011	Yapısal eşitlik modeli	BİT yoğunluğu, Eğitim seviyesi, İletişim becerileri, İnovasyon sayısı Araştırma geliştirme giderleri, İleri teknoloji üretimi

(*) Çalışmada dikkate alınan dönemi göstermektedir.

Ülkelerin ekonomik açıdan durumunu gösteren değişkenlerin başında GSYİH (Gayri Safi Yurtiçi Hasıla) göstergesi gelmektedir. Literatürde en fazla kullanılan makroekonomik gösterge GSYİH verileridir. Mah ve Kim (2006: 160) Kore'de vergiler ve makroekonomik değişkenler arasındaki uzun dönem ilişkisiyi göstermek için GSYİH değerlerini kullanmıştır. Punzi (2019: 1310) Var modeli kullanarak makroekonomik değişkenlerin enerji fiyatları üzerinde etkisini araştırmak için yaptığı çalışmada yine GSYİH'den faydalanmıştır. Todshki ve Ranjbaraki (2012: 397) saat ithalatı ve ihracatı üzerindeki makroekonomik değişkenlerin etkisini incelerken GSYİH, küresel petrol fiyatları ve popülasyon verilerini kullanmışlardır. Taghavi (2012: 1290), İran'a ait ekonomik gelişim üzerindeki ithalat ve ihracatın etkisini araştırırken ithalat, ihracat ve GSYİH verilerini kullanmıştır. Cansino vd. (2013: 280) biyoyakıt tesislerinin üretim kapasitesinin artırılmasının ekonomik etkilerini tahmin ettikleri çalışmada istihdam ve

GSYİH verilerini makroekonomik değişkenler olarak kullanılmışlardır. Bhargava (2014: 249) firmaların uzun dönem borçları ve toplam varlıkları ile makroekonomik değişkenler arasındaki ilişkiyi modellerken işsizlik verilerinden faydalanmıştır. Msaten ve Gnip (2019: 1190) kamu yatırımlarının makroekonomik etkilerini araştırırken işsizlik ve GSYİH verilerinden yararlanmıştır. Katırtıcıoğlu vd. (2015: 268) petrol fiyat hareketleri ve makroekonomik performans arasındaki ilişkiyi ortaya çıkarmak için panel veri analizi ile GSYİH, enflasyon ve işsizlik verilerini dikkate almıştır. Jalaee vd'nin (2019: 5) geliştirdikleri fonksiyonda enflasyon, işsizlik ve döviz kuru değişkenleriyle ARDL modeli ve yapay sinir ağları yöntemlerini uygulamışlardır. Şahin ve Özenç (2007: 198) GSYİH, enflasyon ve işsizlik verilerini makroekonomik değişkenler olarak aldıkları çalışmalarında kamu harcamaları ile nedensellik ilişkisi incelenmiştir. Özdemir ve Aksoy (2012: 120) makroekonomik değişkenlerinin istihdam üzerindeki etkilerinin vektör hata düzeltme modeli ile incelendikleri çalışmada GSYİH, faiz oranı ve ihracatın ithalatı karşılama oranı gibi değişkenleri kullanmışlardır. Bu çalışmada ise ülkelerin genel ekonomik seviyesini gösteren makroekonomik değişkenler olarak işsizlik, GSYİH ve ihracat ithalat oranı verileri kullanılmıştır.

2. VERİ MADENCİLİĞİ

Verilerin birikmesiyle günümüzde ortaya çıkarılan anlamlı bilgiler veri madenciliği alanını meydana getirmiştir. Veri madenciliği kavramı 1990'lı yıllarda ilk olarak veri tabanlarında bilgi keşfi süreciyle ortaya çıkmıştır ve bu sürecin aşamalarında yer almaktadır. Veri madenciliği büyük miktarlardaki veriyi farklı metotlarla hem anlaşılabilir hem de faydalı hale getiren ve beklenmedik ilişkileri bulmak için gözlemsel veri setlerinin analiz edilmesini sağlayan bir yöntemdir (Koh vd., 2004: 101). Bu ilişkileri ve gizli bilgileri ortaya çıkarılmasında dikkat edilmesi gereken önceden bilinmeyen bilgilere ulaşılmasıdır (Silahtaroglu, 2013: 12). Bu nedenle veri madenciliği faaliyetlerinin kurum ve kuruluşların karar mekanizmalarına destek sağlamada kritik öneme sahip olduğu söylenebilmektedir (Özkan, 2013: 38). Veri madenciliği veri tabanında önceden ortaya çıkarılmamış ve öngörülemeyen bağlantıların ikincil anlamda analizi olarak da tanımlanmaktadır. Aynı zamanda istatistik, matematik, veri tabanı teknolojileri, örüntü tanıma ve makine öğrenmesi ile ilişkili yeni ve farklı bir disiplin olarak açıklanmaktadır (Hand, 1998: 11). Veri madenciliği bilgi keşfi süreci olduğundan ilk olarak tanımlanan probleme uygun verilerle ilgili seçim yapılması ve önışleme gibi hazırlıklarla başlamaktadır. Amaca göre belirlenen veri madenciliği yöntemine uygun algoritmalarla analiz süreci devam etmektedir. Veri madenciliği tekniklerinin farklı algoritmaları bulunduğundan süreçteki adımlar kullanılan tekniğe göre değişebilmektedir. Son olarak elde edilen sonuçlar test edilmekte, yorumlanmakta ve değerlendirilmektedir (Tekerek, 2011: 3).

Veri madenciliğinde veri yığınlarındaki gizli bilgiler ortaya çıkarılırken amaçlarına göre farklı yöntemler kullanılmaktadır. Burada veri madenciliği tekniklerinin hangi amaçla kullanıldığının bilinmesi gerekmektedir. Sınıflama, kümeleme, tahmin etme, ilişki analizi, birliktelik analizi gibi amaçlara yönelik farklı veri madenciliği modelleri geliştirilmiştir. Bu çalışmada veri madenciliğine ait sınıflandırma uygulaması kullanılmıştır. Elde edilen bilgi filtrelenmekte, hazırlanmakta ve aynı zamanda faydalı kararlar ve stratejiler için sınıflandırılmaktadır (Hand, 1995: 1). Sınıflandırma tekniğinde algoritma eldeki eğitim kümesinden belirli bir dağılım öğrenmektedir. Sonraki adımda ise herhangi bir sınıfa ait olmayan test verilerini doğru bir şekilde sınıflara ayırmaktadırlar. Bu nedenle sınıflandırma işlemi genel olarak eğitim verileriyle öğrenme ve öğrenilen değerlerle test verileri üzerinde uygulanmaktadır (Şeker, 2013: 15). Gerek istatistik gerekse makine öğrenimi temelli çeşitli sınıflandırma yöntemleri geliştirilmiştir. İstatistik yöntemler içerisinde lineer ve lineer olmayan başlıca yöntemler lineer regresyon analizi, lojistik regresyon analizi, diskriminant analizi, bayes sınıflandırma teknikleridir. Makine öğrenimi yöntemleri içerisinde ise karar ağaçları, en yakın

komşu yöntemi, yapay sinir ağları ve destek vektör makineleri başlıca yöntemlerdir (Akpınar, 2014: 25). Sınıflandırma amacıyla kullanılan tekniklerden en yaygın olanları ise karar ağaçları ile yapay sinir ağlarıdır. Yapay sinir ağları insan sinirinin bir modelini meydana getirmektedir. Yapay Sinir ağları bir ya da daha fazla çıktı tahmin etmek için kullanılmaktadır. Yapay sinir ağları örneklerden öğrenmeye bağlı öğrenme tekniklerine göre uygulanmaktadır. Yani bir olay ile alakalı gerçekleşmiş örneklerin kullanılmasıyla, girdiler ve çıktılar arasında bulunan bağlantılar öğrenilmeye çalışılmakta ve bu bağlantılara göre yeni örnek çıktıları belirlenmektedir (Öztemel, 2006: 24). Karar ağacı uygulaması ise veri madenciliği alanında ucuz olma, yorumlanma kolaylığı, veri tabanı sistemleri ile basitçe bütünleşebilme, yüksek güvenilirlikte olma gibi özelliklerinden dolayı diğer sınıflandırma modellerine göre daha yaygın kullanılmaktadır (Özkes, 2003: 8). Bu nedenle çalışmada sınıflandırma yöntemlerinden karar ağaçları kullanılmıştır.

2.1. Karar Ağaçları

Sınıflama işlevlerinden en çok faydalanılan algoritma karar ağaçları algoritmalarıdır. Başka tekniklerle karşılaştırıldığında karar ağaçlarının yapılandırmanın ve anlamının daha basit olduğu söylenebilir (Agrawal ve Srikant, 1994: 488). Temel olarak iki adımdan oluştuğu söylenebilir: Birincisi ağacın kurulması, ikincisi de verilerin teker teker ağaca uygulanarak sınıflandırmanın gerçekleştirilmesi şeklindedir. Karar ağaçları meydana getirilirken hangi algoritmanın kullanılacağı önemlidir. Seçilen algoritmaya göre ağaç şekil değiştirebilmektedir. Farklı yapıya sahip ağaçlar farklı sınıflama sonucu verebilmektedir (Silahtaroglu, 2013: 68). Seçilen karar ağacı algoritmasına göre bölünme işlemi her aşamada ikili, üçlü veya çoklu olabilmektedir (Akpınar, 2014: 77). Karar ağaçlarına ait farklı algoritmalar yer almaktadır. Bunlar: entropiye dayalı ID3, C4.5, C5.0 algoritmaları, sınıflandırma ve regresyon ağaçları (CART) ve bellek tabanlı sınıflandırma algoritmalarıdır (Özkan, 2013: 118). Bu çalışma da elde edilen değişkenlere ait veriler sayısal nitelikte olduğu için entropiye dayalı algoritmalarından C5.0 kullanılmıştır.

2.1.1. C5.0 Algoritması

Sayısal değerleri olan özelliklerin karar ağaçlarında kullanılmasına imkân sağlamaktadır ve önceden bilinmeyen özellik değerlerine sahip olan örneğe ait kümeler için karar ağacının ne şekilde tasarlanacağı ile ilgili bir yol göstermektedir (Özkan, 2013: 122). ID3 algoritması ile aynı sürece sahiptir. Ek olarak C5.0 algoritmasında sayısal nitelikteki veriler belirli bir eşik değeriyle dönüştürülmektedir. İlk aşamada sayısal verilere sahip olan değişkenin değerlerinin orta noktası eşik değeri olarak belirlenir ve bu değerden küçük, eşit veya büyük olarak gruplanmaktadır. İkinci aşamada hedef niteliğe ait entropi değeri hesaplanır. Entropi karar ağaçlarının dallanmanın yapılmasında faydalanılan bir ölçüdür. Kısacası bir sistemdeki belirsizliğin ölçüsü olarak tanımlanmaktadır. Entropi değeri aşağıdaki formülle hesaplanmaktadır:

Herhangi bir sınıf niteliğine ait değişkenin alacağı değerlere göre k sınıfa ait değerler c_1, c_2, \dots, c_k olsun, t ise niteliğe ait tüm değerlerin sayısı olsun

$$P_t = c_1/t, c_2/t, \dots, c_k/t \text{ ise } k \text{ gruba ait olasılıklar ise entropi;}$$

$$H(t) = \sum P_i \log_2 P_i \text{ olarak hesaplanır.}$$

Üçüncü aşamada her bir sınıf niteliğine ait grupların hedef niteliğindeki karşılıkları incelenir. Bu grupların kendi içerisindeki entropileri hesaplanır. Hesaplanan entropi değerleri niteliğe ait grupların olasılıkları ile çarpılır ve daha sonra bu değerler toplanır.

$$H(x, t) = \sum t/t_i H(t_i)$$

Dördüncü aşamada kazanç ölçütü belirlenir. Bu ölçü ise hedef niteliğe ait entropi değerinden, incelenen niteliğe ait yukarıdaki hesaplanan değer çıkarılır.

$$\text{Kazanç}(x,t)=H(t)- H(x,t)$$

Yukarıda incelenen bütün işlemler her nitelik için uygulanmaktadır. En yüksek kazanç oranına sahip olan nitelik ile dallanma başlamaktadır. Dallardan itibaren aynı işlemler karar ağacı tamamlanana kadar iteratif bir şekilde devam etmektedir (Özkan, 2013: 130).

2.2. Uygulama

Uygulamanın ilk aşamasında 34 Avrupa Birliği üye ülkesi ve Türkiye'nin işsizlik, GSYİH ve ihracat ithalat oranı verileri kullanılarak kümeleme analizi yapılmıştır. Ülkelerin 2016-2019 yılları arasındaki verileri kullanılmıştır. Her yıl için yarı yarı kümeleme analizi yapılmıştır. İkinci aşamada kümeleme analizi ile gruplanan ülkeler hedef değişken, bilgi ekonomisi değişkenleri kullanılarak yine aynı yıllara ait verilerle karar ağacı uygulamasına katılmışlardır. Genel ekonomik seviye açısından gruplanan ülkeleri birbirinden ayıran bilgi ekonomisi değişkenleri tespit edilmiştir.

2.2.1. Kümeleme Analizi İle Ülkelerin Gruplanması

Bu aşamada 2016-2019 yılları arasında yer alan döneme ait GSYİH, ihracat ithalat oranı ve işsizlik değişkenleri göz önüne alınarak hiyerarşik olmayan kümeleme yöntemlerinden k-ortalama yöntemini kullanılmıştır. Diğer kümeleme yöntemleri ile aynı sonuçların elde edildiği görülmüştür.

Kümeleme analizinde istatistiksel olarak herhangi bir varsayıma ihtiyaç duyulmamaktadır. Değişkenler açısından birimleri ayırt edici özellik taşıması nedeniyle çoklu doğrusal bağlantı aranan bir durumdur. Bu nedenle her yıl için değişkenler arasındaki Pearson korelasyon katsayıları ve anlamlılık dereceleri hesaplanmıştır.

Tablo 2'de dikkate alınan değişkenler arasındaki korelasyon katsayıları ve anlamlılık seviyeleri yıllar bazında incelenmiştir. Tabloya bakıldığında GSYİH ve ihracat ithalat oranı değişkenleri arasındaki korelasyonların her yıl için %70'in üzerinde olduğu ve bu değerlerin %1 ve %5 'e göre anlamlı olduğu söylenebilir. GSYİH ve işsizlik rakamları değişkenlerine bakıldığında ise korelasyon katsayılarının yıllara göre %45 ile %50 arasında değiştiği ve bu değerlerin ise %5'e göre anlamlı olduğu görülmektedir. İhracat ithalat oranı ve işsizlik değişkenleri arasındaki korelasyonlarının ise %40 ile %58 arasında değişmekte olduğu ve 2018, 2019 yıllarındaki katsayıların %1 ve %5'e göre anlamlı olduğu; 2016, 2017 yıllarındaki değerlerin ise %5'e göre anlamlı olduğu görülmektedir.

Tablo 2: Kümeleme Analizi Değişkenlerinin Arasındaki Korelasyon Katsayılarının Anlamlılık Testleri

Yıllar	GDP/Exptoimp		GDP/Unemp		Exptoimp/Unemp	
	Pearson Korelasyon Katsayısı	Anlamlılık	Pearson Korelasyon Katsayısı	Anlamlılık	Pearson Korelasyon Katsayısı	Anlamlılık
2019	0,756	0,000	-0,478	0,036	-0,582	0,005
2018	0,780	0,000	-0,463	0,02	-0,432	0,002
2017	0,792	0,000	-0,45	0,04	-0,475	0,04
2016	0,710	0,000	-0,494	0,02	-0,409	0,02

Tablo 3'ten de görüldüğü gibi, %5 ve %1 hata payı ile tüm değişkenler istatistiksel olarak anlamlıdır. Bu doğrultuda, seçilen değişkenlerin ülkeleri ekonomik gelişmişlik açısından sınıflandırmak için uygun olduğu söylenebilir.

Tablo 3: Kümeleme Analizi Değişkenlerinin Anlamlılık Testleri

Yıllar	GDP		Unemp		Exptoimp	
	F İstatistiği	Anlamlılık	F İstatistiği	Anlamlılık	F İstatistiği	Anlamlılık
2019	149,256	0,000	4,162	0,022	15,452	0,000
2018	161,148	0,000	4,826	0,017	16,891	0,000
2017	170,958	0,000	6,341	0,004	9,160	0,000
2016	158,210	0,000	7,456	0,002	11,931	0,000

Tablo 4'de değişkenlerin ortalama değerleri (küme merkezleri) yer almaktadır. Böylece kümelerin içindeki gözlemlerin özelliklerini ortaya koymak mümkün olabilmektedir. Tüm yıllar için GSYİH değişkeni en iyi olan ülkeler 3.kümede yer almaktadır. Aynı şekilde ihracat ve ithalat oranı en iyi olan ve işsizliği en düşük olan ülkelerde 3. Kümede yer almaktadır. Bu nedenle 3.küme de ekonomik büyüklük seviyesi yüksek olan ülkeler yer almaktadır diyebiliriz. Tablodaki değerlere bakıldığında 1.kümede 2.kümedeki ülkelere göre daha iyi değerler olduğu için ekonomik büyüklük seviyesi ortalamada olan ülkeler olarak adlandırabiliriz. 2.kümedeki ülkeler ise diğer ülkelere göre ekonomik olarak seviyesi düşük olan ülkelerdir.

Tablo 4: Gruplara Ait Küme Merkezleri

Yıllar	GDP			Exptoimp			Unemp		
	1.Küme	2.Küme	3.Küme	1.Küme	2.Küme	3.Küme	1.Küme	2.Küme	3.Küme
2016	46181	13854	75280	1,08	0,93	1,31	7,99	12,78	5,02
2017	47849	13989	75763	1,10	0,96	1,28	8,18	13,11	5,01
2018	49002	14230	76874	1,06	0,89	1,32	7,93	11,66	4,32
2019	51293	14937	77855	1,07	0,9	1,3	7,16	10,73	4,08

Tablo 5’te elde edilen sonuçlara göre, her yıl için genel olarak aynı kümeleme analizi sonuçlarının elde edildiği görülmektedir. 1. grupta ekonomisi ortalamanın üzerinde olan ülkeler yer almaktadır. Bu ülkeler Belçika, Danimarka, Hollanda, İrlanda, Fransa, Avusturya, İsveç, İngiltere, İzlanda’dır. Ekonomik seviyesi yüksek olan ülkeler ise 3.grupta yer almaktadır. Bu ülkeler sadece 3 tanedir. Lüksemburg, Norveç ve İsviçre ülkeleri bu grubu oluşturmaktadır. Aralarında Türkiye’ nin de bulunduğu 2.grupta ekonomik olarak gelişmekte olan ülkeler yer almaktadır. Bulgaristan, Çek Cumhuriyeti, Estonya Yunanistan, İspanya, Hırvatistan, İtalya, Kıbrıs, Letonya, Litvanya, Makedonya, Malta, Polonya, Portekiz, Romanya, Slovenya, Slovakya, Karadağ, Yugoslavya ve Sırbistan bu grupta yer almaktadır.

Tablo 5: Kümeleme Analizi Gruplama Sonuçları

Ülkeler	2016	2017	2018	2019
Belçika	1	1	1	1
Bulgaristan	2	2	2	2
Çek Cumhuriyeti	2	2	2	2
Danimarka	1	1	1	1
Almanya	1	1	1	1
Estonya	2	2	2	2
İrlanda	1	1	1	1
Yunanistan	2	2	2	2
İspanya	2	2	2	2
Fransa	1	1	1	1
Hırvatistan	2	2	2	2
İtalya	2	2	2	2
Kıbrıs	2	2	2	2
Letonya	2	2	2	2
Litvanya	2	2	2	2
Lüksemburg	3	3	3	3
Macaristan	2	2	2	2

Malta	2	2	2	2
Hollanda	1	1	1	1
Avusturya	1	1	1	1
Polonya	2	2	2	2
Portekiz	2	2	2	2
Romanya	2	2	2	2
Slovenya	2	2	2	2
Slovakya	2	2	2	2
Finlandiya	1	1	1	1
İsveç	1	1	1	1
İngiltere	1	1	1	1
İzlanda	1	1	1	1
Norveç	3	3	3	3
İsviçre	3	3	3	3
Karadağ	2	2	2	2
Makedonya	2	2	2	2
Sırbistan	2	2	2	2
Türkiye	2	2	2	2

2.2.2. Karar Ağacı Uygulaması

Çalışmanın bu aşamasında, karar ağaçları için C5.0 algoritması uygulanmış olup, karar ağacı uygulamasında 34 bilgi ekonomisi değişkeni kullanılmıştır. C5.0 algoritması, sayısal değerlere sahip niteliklerinde karar ağaçlarında kullanılmasına imkân sağlayan denetimli bir sınıflandırma algoritmasıdır. İlk olarak hedef değişkenin bulunduğu bir veri kümesinin eğitilmesiyle model meydana gelmektedir. Elde edilen model yeni bir veri setiyle de test edilmekte ve modelin performansı gözlenmektedir. Bu nedenle eğitim veri seti olarak 140 adet veri içinden 105 tanesi rastgele seçilmiştir ve model oluşturulmuştur.

Önceki uygulama adımında yapılan kümeleme analizi sonucu ekonomik büyüklük seviyelerine göre gruplandırılan ülkeler bu aşamada hedef değişken olarak kullanılmıştır. Dolayısıyla hedef değişken ekonomik seviye olarak belirlenmiş olup 1, 2 ve 3 olarak gruplandırılmıştır. 1. grupta ekonomik açıdan orta seviyede olan ülkeler, 2.grupta ekonomik ortalamanın altında olan gelişmekte olan ülkeler ve 3.grupta ekonomik açıdan gelişmiş olan ve ortalama seviyenin üzerindeki ülkeler yer almaktadır. Tahminleyici alanlar olarak ise bilgi ekonomisi değişkenleri kullanılmıştır. Girdi değişkeni olarak kullanılan bilgi ekonomisi değişkenlerinin ayrıntılı açıklamaları aşağıda yer almaktadır.

Bu değişkenler uygulamada D1, D2, ..., D34 olarak kodlanmıştır. Kullanılan değişkenler:

D1: Düzenli olarak internet kullanan bireylerin yüzdesi

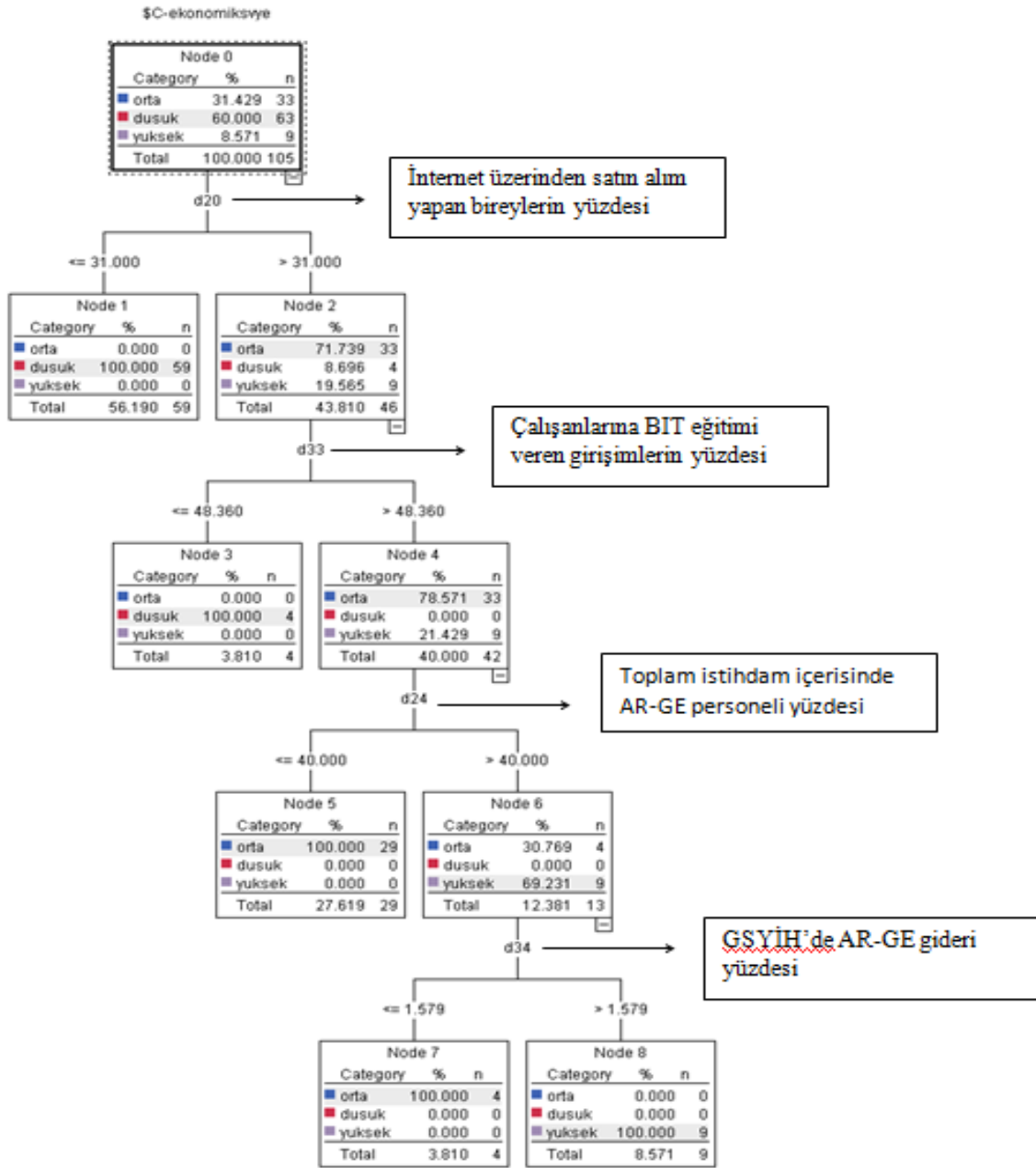
D2: E-bankacılık ve e-ticaret kullanım yüzdesi

- D3: Ağ bağlantısına sahip bilgisayarlar üzerinden sipariş alan girişimlerin yüzdesi
- D4: Evde internet erişimi olan hane halkı yüzdesi
- D5: Geniş Bant internet erişimine sahip hane halkı
- D6: Geniş Bant internet erişimine sahip girişimler
- D7: Geniş Bant internet erişimine sahip çalışanlar
- D8: Fonksiyonel departmanlar arasındaki bilgi paylaşımını sağlamak için kurumsal kaynak planlama yazılımlarına (ERP) sahip olan girişimlerin yüzdesi
- D9: Bilgisayar kullanan girişimlerin yüzdesi
- D10: İnternet ya da internet dışında başka ağlar aracılığıyla satış yapan girişimlerin yüzdesi
- D11: İstihdam edilen bireylerin internete erişim yüzdesi
- D12: Web site aracılığıyla kamu kuruluşlarından bilgi alma yüzdesi
- D13: İnternet üzerinden form doldurma ve gönderme yüzdesi
- D14: Kamu kuruluşlarıyla iletişim kurabilmek için interneti kullanan bireylerin yüzdesi
- D15: Girişimlerin e-ticaret iş hacmi
- D16: Online satın alım yapan girişimlerin yüzdesi
- D17: Web siteler aracılığıyla yapılan toplam satışlar
- D18: İnternet ya da internet dışında başka ağlar aracılığıyla yapılan satışlar ve satın alımların değeri
- D19: Web siteler aracılığıyla yapılan toplam satışlar
- D20: İnternet üzerinden satın alım yapan bireylerin yüzdesi
- D21: İleri teknoloji ihracat yüzdesi
- D22: Toplam istihdam içerisinde bilgi ve iletişim teknolojileri personeli yüzdesi
- D23: Mobil internet kullanım yüzdesi
- D24: Toplam istihdam içerisinde AR-GE personeli yüzdesi
- D25: BİT eğitilmiş çalışanları olan girişimlerin yüzdesi
- D26: Toplam AR-GE personel yüzdesi
- D27: Girişimlerin toplam AR-GE gideri
- D28: Patent uygulamalarının sayısı
- D29: Telekomünikasyon sektörü toplam gelirleri
- D30: Telekomünikasyon sektörü toplam yatırımları
- D31: Toplam istihdam içerisinde BİT personeli yüzdesi
- D32: Yüksek teknoloji ticaret hacmi
- D33: Çalışanlarına BİT eğitimi veren girişimlerin yüzdesi
- D34: GSYİH’de AR-GE gideri yüzdesi şeklinde ifade edilmektedir.

3. BULGULAR

Bu çalışmada genel ekonomik grupların ayrılmasında belirleyici olan bilgi ekonomisi değişkenlerinin tespiti yapılmıştır. Eğitim ve Ar-Ge değişkenlerinin genel ekonomik seviyeyi ayıran bilgi ekonomisi değişkenleri olduğu bulunmuştur. Bu amaçla çalışmada ilk olarak kümeleme analizi ve Karar Ağaçları algoritması kullanılmıştır. Karar ağaçları uygulamasında C5.0 algoritması kullanılmıştır. Kümeleme analizi sonucu, GSYİH, ihracat ithalat oranı ve işsizlik değişkenleri kullanılarak ülkeler 2016, 2017, 2018 ve 2019 yılları için 3 farklı gruba ayrılmışlardır. Kümeleme analizi sonucunda elde edilen 3 farklı grup kategorik değişken ve 34 bilgi ekonomisi sayısal değişkeni kullanılarak karar ağaçlarından C5.0 algoritması uygulanmıştır. Yapılan kümeleme analizi sonucu düşük, orta ve yüksek olarak 3 farklı ekonomik seviye grubu elde edilmiştir. Elde edilen seviye gruplarının C5.0 algoritması kullanılarak 4 farklı bilgi ekonomisi değişkeni ile ayrıldığı tespit edilmiştir.

Karar ağacı analizi sonucunda elde edilen dendograma göre (Şekil 1) bilgi ekonomisi değişkenlerinden internet üzerinden satın alım yapan bireylerin yüzdesi değişkenini ifade eden D20 değişkeni (internet üzerinden satın alım yapan bireylerin yüzdesi) değeri %32'den küçük olanlar ekonomik açıdan düşük seviyede olan ülkeler olarak sınıflandırılmaktadır. %32'den büyük olan ülkeler ise ekonomik açıdan ortalama seviyede olan ve ekonomik açıdan yüksek seviyede olan ülkeler olarak sınıflandırılmıştır. Aynı şekilde D33 değişkeni (çalışanlarına BİT eğitimi veren girişimlerin yüzdesi) değeri %48'in üstünde olan ülkeler de ekonomik açıdan ortalama değerlere sahip olan ve bu gruba göre ekonomik açıdan daha iyi seviyedeki ülkeler olarak elde edilmiştir. Ekonomik açıdan orta seviyedeki ve yüksek seviyedeki ülkeleri ayıran değişken ise D24 değişkeni olarak belirlenmiştir. D24 değişkeni, toplam istihdam içerisinde AR-GE personeli yüzdesidir. Bu değişkenin değeri %40'ın altında olan ülkeler ekonomik açıdan orta seviyedeki ülkeler olarak sınıflanırken, %40'ın üstünde olan ülkeler ise ekonomik açıdan yüksek seviyedeki ülkeler olarak sınıflanmıştır. Benzer şekilde D34 değişkeni (GSYİH'de AR-GE gideri yüzdesi) yani GSYİH'de AR-GE gideri yüzdesi değişkeni 1,579'nın altında olan ülkeler ekonomik açıdan orta seviyedeki ülkeler iken bu değer üstündeki ülkeler ekonomik açıdan yüksek seviyedeki ülkelerdir.



Şekil 1: Karar Ağacı Analizine Ait Dendrogram

Modelin doğruluğu analiz edildiğinde %95,2 oranında doğruluk payı olduğu ve yanlış payının %4,8 olduğu görülmektedir. Modeldeki doğru sınıflama başarısı, test veri setiyle ölçülerek %97,1 olarak elde edilmiştir. Elde edilen modelin sınıflandırma performansı ile ilgili tablo ise aşağıda yer almaktadır. Geriye kalan 35 veri ile model test edilmiştir. Her iki veri seti için sınıflandırma başarısı Tablo 6'de yer almaktadır.

Tablo 6: Karar Ağacı Analizi Sınıflandırma Performansı

Sınıflandırma	Eğitim Veri Kümesi		Test Veri Kümesi	
	Sayı	Yüzde	Sayı	Yüzde
Doğru	100	95,2	34	97,1
Yanlış	5	4,8	1	2,9

4. TARTIŞMA ve SONUÇ

Bu çalışmada genel ekonomik seviye ile bilgi ekonomisi göstergelerinin incelenmesi amacıyla kümeleme analizi ve veri madenciliği kullanılarak genel ekonomik düzeyi belirleyen bilgi ekonomisi göstergeleri tespit edilmeye çalışılmıştır. Bu amaçla, AB ülkeleri ile ülkemizin de yer aldığı ülkelerin GSYİH, işsizlik ve ihracat-ithalat oranı değişkenleri göz önüne alınarak kümeleme analizi yapılmıştır. Böylece ülkelerin her yıl için genel anlamda ekonomik gruplanmaları tespit edilmiştir. Analiz sonucu, ekonomik seviye değerleri ortalama değerlere sahip olan gelişmiş ülkeler 1.grubu oluşturmuştur. Bu ülkeler Belçika, Danimarka, Hollanda, İrlanda, Fransa, Avusturya, İsveç, İngiltere, İzlanda'dır. Ekonomik değerleri ortalamanın üzerinde olan ülkeler ise 3.grupta oluşturmuştur. Bu ülkeler Lüksemburg, Norveç ve İsviçre'dir. Aralarında Türkiye'nin de bulunduğu 2.grupta ekonomik seviye olarak ortalamanın altında olan, gelişmekte olan ülkeler yer almaktadır. Bulgaristan, Çek Cumhuriyeti, Estonya, Yunanistan, İspanya, Hırvatistan, İtalya, Kıbrıs, Letonya, Litvanya, Makedonya, Malta, Polonya, Portekiz, Romanya, Slovenya, Slovakya, Karadağ, Yugoslavya ve Sırbistan bu grupta yer almaktadır. Türkiye yapılan kümeleme analizi sonucunda gelişmekte olan ülkeler grubuna dahil olmuştur.

Bir sonraki adımda C5.0 algoritması kullanılarak karar ağacı uygulaması gerçekleştirilmiştir. Hedef değişken olarak önceki aşamada kümeleme analizi ile belirlenen genel ekonomik seviye gruplar, tahminleyici değişken olarak 35 tane bilgi ekonomisi değişkeni kullanılmıştır. Ekonomik seviye açısından en ayırıcı bilgi ekonomisi değişkenleri tespit edilmiştir. Bu analiz sonucunda ekonomik olarak gelişmekte olan ülkeleri (Bulgaristan, Çek Cumhuriyeti, Estonya Yunanistan, İspanya, Hırvatistan, İtalya, Kıbrıs, Letonya, Litvanya, Makedonya, Malta, Polonya, Portekiz, Romanya, Slovenya, Slovakya, Karadağ, Yugoslavya ve Sırbistan) diğer iki gruptan ayıran değişkenler internet üzerinden satın alım yapan bireylerin yüzdesi değişkeni ve çalışanlarına BİT eğitimi veren girişimlerin yüzdesi değişkeni olarak görülmüştür. Ekonomik olarak ortalama değerlere sahip ülkeler (Belçika, Danimarka, Hollanda, İrlanda, Fransa, Avusturya, İsveç, İngiltere, İzlanda) ve bu ülkelere nispeten daha iyi durumda olan ülkeleri (Lüksemburg, Norveç ve İsviçre) ayıran değişkenler ise toplam istihdam içerisinde Ar-Ge personeli yüzdesi değişkeni ile GSYİH'da yer alan Ar-Ge gideri yüzdesi değişkenleri olarak tespit edilmiştir.

Elde edilen sonuçlara göre gelişmekte olan ülkelerle gelişmiş olan ülkeleri ayıran bilgi ekonomisi değişkeni eğitim temelli değişkenlerdir. Gelişmiş olan ülkeler ile refah seviyesi olarak daha ileride olan ülkeleri ayıran bilgi ekonomisi değişkenleri ise Ar-Ge ve inovasyon değişkenleridir. Ülkelerin, ekonomik seviye ve refah olarak ilerleyebilmek için öncelikle eğitime, bunun yansısı Ar-Ge ve inovasyon faaliyetlerine önem vermeleri gerektiği tespit edilmiştir.

Araştırma ve Yayın Etiği Beyanı

Bu çalışma bilimsel araştırma ve yayın etiği kurallarına uygun olarak hazırlanmıştır.

Yazarların Makaleye Katkı Oranları

Birinci yazarın makaleye katkısı %60, İkinci yazarın makaleye katkısı %40'tır.

Çıkar Beyanı

Yazarlar açısından ya da üçüncü taraflar açısından çalışmadan kaynaklı çıkar çatışması bulunmamaktadır.

KAYNAKÇA

Adaçay, F. R. (2007). Bilgi ekonomisine ilişkin temel göstergeler açısından Avrupa Birliği ve Türkiye'nin karşılaştırılması, *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi* 19, 185-204.

Agrawal, R., & Srikant, R. 1994, Fast algorithms for mining association rules. In *Proc. 20th Int. Conf. Very Large Data Bases, VLDB* Vol. 1215, 487-499.

Akpınar, H. (2014). *Data: Veri madenciliği veri analizi*. Papatya Yayıncılık Eğitim.

Andrés, A. R., Asongu, S. A., & Amavilah, V. (2015). The impact of formal institutions on knowledge economy. *Journal of the Knowledge Economy*, 6(4), 1034-1062.

Asongu, S. A. (2014). Knowledge Economy and financial sector competition in African countries. *African Development Review*, 26(2), 333-346.

Berberoğlu, B. (2010). Yaşam boyu öğrenme ile bilgi ve iletişim teknolojilerin açısından türkiye' nin avrupa birliğindeki konumu. *Bilgi Ekonomisi ve Yönetimi Dergisi*, 5(2). 113-126.

Bhargava, A. (2014). Firms' fundamentals, macroeconomic variables and quarterly stock prices in the US. *Journal of Econometrics*, 183.2, 241-250.

Cansino, J. M., et al. (2013). Economic impacts of biofuels deployment in Andalusia. *Renewable and Sustainable Energy Reviews*, 27, 274-282.

Chye, K. H., Chin, T. W., & Peng, G. C. (2004). Credit scoring using data mining techniques. *Singapore Management Review*, 26(2), 25-48.

Hand, D. J. (1998). Data mining: statistics and more?. *The American Statistician*, 52(2), 112-118.

Driouchi, A., Azelmad, E. M., & Anders, G. C. (2006). An econometric analysis of the role of knowledge in economic performance. *The Journal of Technology Transfer*, 31(2), 241-255.

Galadima, M. D., & Aminu, A. W. (2019). Shocks effects of macroeconomic variables on natural gas consumption in Nigeria: Structural VAR with sign restrictions. *Energy policy*, 125, 135-144.

Işık, N., & Kılınç, E. C. (2013). Bilgi ekonomisi ve iktisadi büyüme: OECD ülkeleri üzerine bir uygulama. *Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 13(26), 21-54.

Jalae, S. A., Mehrdad L., and Amin, G. (2019). The Phillips curve in Iran: econometric versus artificial neural networks. *Heliyon*, 5(8),1-6.

Kaynak, S., & Arslan, I. (2012). In the context of the knowledge economy of parameters: econometrics analysis of the Organisation for Economic Co-Operation and Development (OECD) Economies. *African Journal of Business Management*, 6(12), 4454.

Katircioğlu, S. T., Sertoğlu, K., Candemir, M., & Mercan, M. (2015). Oil price movements and macroeconomic performance: evidence from twenty-six OECD countries. *Renewable and Sustainable Energy Reviews*, 44, 257-270.

Mah, Jai S.; Kim, Yong Dae. (2006). Antidumping duties and macroeconomic variables: the case of Korea. *Journal of Policy Modeling*, 28.2: 157-162.

Masten, I., & Gnip, A. G. (2019). Macroeconomic effects of public investment in South-East Europe. *Journal of Policy Modeling*, 41(6), 1179-1194.

Oktay, E., & Kaynak, S. (2007). Türkiye ve Avrupa Birliği ülkelerinin bilgi ekonomisi girdi ve çıktı değişkenleri arasındaki kanonik ilişkinin araştırılması. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 10(2), 419-440.

- Özdemir, Z. A., & Aksoy, E. (2012). Türkiye’de makroekonomik değişkenler ve istihdam teşviklerinin istihdam üzerindeki etkileri. *Akdeniz Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 12(24), 102-124.
- Özekes, S. (2003). Veri madenciliği modelleri ve uygulama alanları. *Istanbul Ticaret Üniversitesi Dergisi*, 3, 65-82.
- Özkan, Y. (2008). *Veri madenciliği yöntemleri*. Papatya Yayıncılık, İstanbul.
- Öztemel, E. (2003). *Yapay sinir ağları*. Papatya Yayıncılık, İstanbul.
- Pradhan, R. P., Arvin, M. B., Bahmani, S., & Bennett, S. E. (2017). The innovation-growth link in OECD countries: could other macroeconomic variables matter?. *Technology in Society*, 51, 113-123.
- Punzi, M. T. (2019). The impact of energy price uncertainty on macroeconomic variables. *Energy policy*, 129: 1306-1319.
- Raspe, O., & Van Oort, F. (2006). The knowledge economy and urban economic growth. *European Planning Studies*, 14(9), 1209-1234.
- Silahtaroglu, G. (2008). *Veri madenciliği*. Papatya Yayınları, İstanbul.
- Şahin, M., & Özenç, Ç. (2007). Kamu harcamaları ile makroekonomik değişkenler arasındaki nedensellik ilişkileri. *Yönetim Bilimleri Dergisi*, 5(2), 177-200.
- Şeker, Ş. E., İş zekâsı ve veri madenciliği. (2013). Cinius Yayınları, 1. Baskı, İstanbul.
- Taghavi, M. et al. (2012). Study on the impact of export and import on economic growth in Iran, *Journal of Basic and Applied Scientific Research*, 2(12), 12787-12794
- Tekerek, A. (2011). Veri madenciliği süreçleri ve açık kaynak kodlu veri madenciliği araçları. *Akademik Bilişim’11 - XIII. Akademik Bilişim Konferansı Bildirileri, 2-4 Şubat 2011, İnönü Üniversitesi, Malatya*, (11), 161-169.
- Todshki, N. E., & Ranjbaraki, A. (2016). The impact of major macroeconomic variables on Iran's steel import and export. *Procedia Economics and Finance*, 36, 390-398.
- Yam, R. C., Guan, J. C., Pun, K. F., & Tang, E. P. (2004). An audit of technological innovation capabilities in Chinese firms: some empirical findings in Beijing, China. *Research Policy*, 33(8), 1123-1140.
- Yapraklı, S., & Sağlam, T. (2010). Türkiye’de bilgi iletişim teknolojileri ve ekonomik büyüme: ekonometrik bir analiz (1980-2008). *Ege Akademik Bakış*, 10(2), 575-582.
- Yaylalı, M., Oktay, E., Akan, Y., & Kaynak, S. (2007). Türkiye ve Avrupa birliğine üye ülkelerin bilgi ekonomisi performanslarının veri zarflama analizi metoduyla karşılaştırılması. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 22(1), 1-25.
- Yeloğlu, H. O. (2009). Bilgi ekonomisi değişkenlerine yönelik ilk izlenimler: Türkiye OECD ülkeleri karşılaştırmaları (1995-1999). *Bilgi Dünyası*, 10(2), 245-260.
- Yeo, B. J. (2010). Driving the knowledge economy: explaining the impact of regional innovation capacity on economic performance. *Contemporary Management Research*, 6(1), 71-86.

Extended Summary

Determination of Knowledge Economy Variables that Affect Economic Development By Using Data Mining

Economic growth can no longer be achieved without knowledge. The raw materials, capital, labor force, which are the classical production factors in the economy, were insufficient with the development of the society based on knowledge. This situation has enabled information to be included among the production factors. There is a strong relationship between the general level of economic development and the knowledge economy.

The motivation of the study is to reveal the decisive indicators of this relationship. The purpose of the study is to determine which knowledge economy indicators are distinctive for the general economic level.

Different studies exist on knowledge economy. In the study of Işık and Kılınç, (2013:23), the relationship between knowledge economy and economic growth was examined by applying panel data analysis on OECD countries. In their study, it is explained that there is a directly proportional relationship between economic development and knowledge economy variables. The number of patent applications, the growth rate of ICT capital services, the export of ICT goods and the increase of high technology exports were used as knowledge economy variables. Yapraklı and Sağlam (2010:576) examined the direction and magnitude of the relationship between knowledge and economic growth in Turkey's economy. In their study, telecommunications investments are used as variables. GNP data and employment data were used as growth variables. As a result of their study it has been found a positive relationship between economic growth and physical capital, labor, human capital and knowledge in the long run. Adaçay (2007:189) compared Turkey and EU countries in terms of some basic indicators of the knowledge economy. The variables used in the study are R&D intensity, R&D employment, number of scientific publications, high technology export figures, patent application numbers, trademark applications and information and communication technology expenditures. A research has been done by Berberoğlu (2010: 115), which is on European Union member countries and Turkey with clustering analysis. In the study of Berberoğlu, variables that are 25-64 years old adults, population, ICT expenses in GDP, percentage of individuals who use information technologies actively in general employment, percent of employees in information technology expertise level in general employment were used. In Mah and Kim's study (2006: 160) it is used GDP values to show the long-term relationship between taxes and macroeconomic variables in Korea. In Taghavi's study (2012: 1290) it is used import, export and GDP data while investigating the impact of imports and exports on Iran's economic development. In Cansino et al.'s study (2013: 280) it is used employment and GDP data as macroeconomic variables in the study where they estimated the economic effects of increasing the production capacity of biofuels.

The difference of our study from the studies given above is the application of the C5.0 decision tree algorithm, one of the data mining techniques as well as Cluster Analysis. Another unique feature of our study is the determination of knowledge economy variables that determine the economic levels of countries.

In our study, the data used were obtained from the website of EUROSTAT. Cluster analysis has been applied in the first stage of the study. Variables, used in cluster analysis, are unemployment, GDP, import and export rate data of 34 European Union member states and Turkey for 2016, 2017, 2018 and 2019.

Since numerical variables were used, in the second stage of our study, one of the data mining techniques, C5.0 decision tree algorithm has been applied. In this step of our study, K-means technique, one of the cluster analysis methods, was used. The same results were obtained with other cluster analysis techniques. As a result of the cluster analysis, countries with an average economy are included in the first group. These countries: Belgium, Denmark, Netherlands, Ireland, France, Austria, Sweden, England, Iceland. Countries which economic levels are high were belongs in the third group. Only three countries are included in this group: Luxembourg, Norway and Switzerland. Group 2, which includes Turkey, is economically developing countries. Bulgaria, Czech Republic, Estonia, Greece, Spain, Croatia, Italy, Cyprus, Latvia, Lithuania, Macedonia, Malta, Poland, Portugal, Romania, Slovenia, Slovakia, Montenegro, Yugoslavia and Serbia are included in this group.

In our study, according to the decision tree model applied, countries less than 32% value of D20 variable, are classified as economically low countries. Countries, value of D20 variable (Percentage of individuals who make purchases online) less than 32%, are classified as economically low countries. Countries, according to D20 that are larger than 32%, are average level and high level in terms of economy. The countries, which above 48% according to variable D33 (% of enterprises that give ICT training to their employees), are average level and high level in terms of economy. The variable, that economically separates average level and high level countries, is determined as the D24 variable (% of R&D personnel in total employment). Countries, that value of this variable are below 40%, are in economically average level. Countries, that value of this variable are above 40%, are economically high level. Countries that value of D34 variable are below 1,579, are in economically average level. Countries that value of D34 variable (Percentage of R&D expense in GDP) are above 1,579, are in economically high level.

According to the results obtained, knowledge economy variables that distinguish developing countries and developed countries are internet usage and education variables. The knowledge economy variables that distinguish developed countries and developing countries that are more advanced in terms of welfare are R&D and innovation variables.

In conclusion, by this study, it has been determined that countries need to give more importance to education and technology, as well as R&D and innovation activities in order to progress in terms of economic level and prosperity.

