

ÜLKELERİN GELİŞİMİŞLİK SEVİYELERİNİN TAHMİNİNDE KULLANILAN SINIFLANDIRMA ALGORİTMALARININ KARŞILAŞTIRILMASI

COMPARISON OF CLASSIFICATION ALGORITHMS USED FOR ESTIMATION OF DEVELOPMENT LEVELS OF COUNTRIES

Esranur DEMİRCİ*, Meltem KARAATLI**

* Yüksek Lisans Öğrencisi, Süleyman Demirel Üniversitesi, Sosyal Bilimler Enstitüsü, İşletme ABD, esra09393@gmail.com, <https://orcid.org/0000-0002-7840-2398>.

** Doç. Dr., Süleyman Demirel Üniversitesi, İİBF, İşletme Bölümü, meltemkaraatli@sdu.edu.tr, <https://orcid.org/0000-0002-7403-9587>.

ÖZ

Ülkelerin gelişmişlik düzeylerini ekonomi, sağlık, teknoloji, eğitim gibi birçok faktör etkilemektedir. Bu çalışma da birçok değişken dikkate alınarak ülkelerin gelişmişlik seviyelerinin tahmininde kullanılan sınıflandırma algoritmalarının karşılaştırması yapılmıştır. Brüt sermaye oluşumu, doğumda beklenen yaşam süresi, yetersiz beslenme yaygınlığı, ileri teknoloji ihracatı, sanayi (inşaat dahil) katma değeri, nüfus artışı, kentsel nüfus, kişi başına gayri safi milli hasıla (GSMH), gayri safi yurtiçi hasıla, toplam işsizlik (toplam işgücünün yüzdesi), enflasyon (GSYH deflatörü) ve brüt sermaye çalışmada kullanılan değişkenler seti olarak belirlenmiştir. Uygulama için veri madenciliğinde kullanılan Waikato Üniversitesinde Java programlama diliyle geliştirilmiş olan WEKA (Waikato Environment for Knowledge Analysis) programı tercih edilmiştir.

Anahtar Kelimeler: Veri Madenciliği, Sınıflandırma, Gelişmişlik.

Jel Kodu: C40

ABSTRACT

Several factors such as economy, health, technology and education affect the development levels of countries. In this study, classification algorithms, used in estimating development levels of countries are compared, considering a number of variables. Gross capital formation, life expectancy at birth, prevalence of malnutrition, advanced technology exports, industrial added value (including construction), population growth, urban population, GDP per capita, gross domestic product, total unemployment (as the percentage of labor force), inflation (GDP deflator) and gross capital were determined as the set of variables used in the study. For the study, WEKA (Waikato Environment for Knowledge Analysis) program developed with Java programming language at Waikato University which is used in data mining was used.

Keywords: Data Mining, Classification, Development.

Jel Code: C40

1. GİRİŞ

Toplumlar eski çağlardan bu yana daha iyi yaşam standartlarına sahip olmayı hedeflemişlerdir. Bu amacı gerçekleştirebilmek için uzmanlar az miktarda

bulunan kaynakları daha verimli bir hale getirebilmek ve üretimde artış sağlayabilmek amacıyla birçok çalışma ve ülkelerin kaynak üretimini artırıp ekonomik

düzeylerinin yükselmesini hedeflemişleridir. Fakat ülkelerin gelişmişlik seviyelerini belirlemedeki tek etken ekonomi değildir. Sosyal ve kültürel gelişmişlik de ülkelerin gelişmişlik seviyelerini arttırmak için son derece önemlidir (Kubar, 2016:66). Ülkelerin sosyo-ekonomik gelişiminde sağlık, teknoloji, ekonomi, nüfus ve eğitim gibi birçok unsur etkili olmaktadır.

Ülkelerde ekonomik gelişmişliği belirlemede en çok göz önünde bulundurulan faktör Gayrisafi Yurtiçi Hâsıla (GSYİH)'dir. GSYİH seviyesi diğer ülkelerden fazla olan ülkelerin gelişmişlik seviyeleri daha yüksek görülmektedir. Fakat GSYİH'nin tek başına bir gelişmişlik ölçütü olarak alınması bir tartışma konusudur. Yüksek GSYİH'e sahip olan ülkelerde yaşam kalitelerinin ve standartlarının düşük olması, gelir dağılımlarında adaletsizlikler olması, çevresel koşulların kötü olması ve hava kirliliğinde artışlar görülmesi ülkelerin gelişmişliği için tek başına GSYİH'nin yeterli olmadığını göstermektedir (Erden Özsoy ve Tosunoğlu, 2017:286).

Dünyada var olan tüm ülkelerin geleceğine ve gelişimine yön veren gelecek nesillerdir. Bu nesillerdeki artış oranı ise işgücü piyasasını önemli derecede etkilemektedir. Yapılacak olan yatırımların ne seviyede ve hangi alanlara yönelik olması gerektiği doğrudan gelecek nesillerle bağlantılıdır (Akça ve Ela, 2012:224). Gelecek nesillerin niceliğini belirleyen faktörlerin başında ise ilk olarak doğum oranları, daha sonra ise ülkedeki ölüm oranları ve yaşam süreleri gibi faktörler gelmektedir.

Ülke gelişimini etkileyen unsurlardan biri olan pozitif enflasyon-büyüme ilişkisinin esasları, enflasyonda meydana gelen artışların yatırım portföyünün finansal sektörden reel sektöre doğru yönelmesine sebep olması, kişilerin reel tasarruf seviyelerini bulunduğu konumda tutmak için ellerinde daha yüksek miktarda para tutmaya çalışması, senyoraj geliri ya da enflasyon vergisinin yükselmesine neden olması gibi sebeplere ve enflasyonun

zorunlu tasarrufları arttırdığı görüşlerine dayanmaktadır (Bruno ve Easterly, 1998:9). Bu fikri doğru bulan birçok uygulamalı çalışmada, enflasyon oranının oldukça düşük veya belirli bir seviyenin altında bulunması durumunda aralarında doğrusal bir ilişkinin olduğu sonucuna ulaşılmaktadır (Yapraklı, 2007:288).

Ülkelerin gelişme ve büyüme kuramlarında fiziksel sermaye ve sermaye kullanımları, nüfus artışı ve sermayede meydana gelen artış ekonomik gelişimin esas öğeleri olarak görülmektedir. Ülkelerde büyüme ve gelişmenin içinde yer alan teknoloji gelişimi ise geleneksel büyüme kuramında kendini ön planda göstermektedir (Keskin, 2011:127). Yüksek teknolojinin ekonomik büyümeye olumlu etkisinden dolayı ileri teknolojiye sahip ülkeler diğer ülkelere bakılarak çok daha avantajlı ve öncelikli durumdadırlar. Bunun nedeni ise gelişmiş ülkelerde teknolojiyi geliştirmek amacıyla yapılan çeşitli yatırımlar zamanla ekonomik gelişimde en temel unsur haline gelmiştir. Teknolojide gün geçtikçe oluşan pozitif yöndeki değişimler, üretim aşamasından etkin bir rol oynayarak üretim faktörlerinin de daha doğru ve aktif kullanımını sağlamaktadır. Bu da ekonomik gelişimin hız kazandırıp ülkede yaşayan bireylerin yaşam memnuniyetlerini ve refah düzeylerini arttırmaktadır (Taban ve Şengür, 2014:356).

Bugüne kadar ülkelerin gelişmişlik seviyelerini etkileyen unsurlarla ilgili nitel ve nicel çalışmalar incelendiğinde tek bir faktörün ülke gelişmişliği için belirleyici bir rol taşımadığı, sosyal ve ekonomik öğelerin bir bütün olarak ele alınması gerektiği sonucuna varılmıştır. Bundan yola çıkarak yapılan bu çalışmada ise daha sağlıklı sonuçlara ulaşabilmek amacıyla gelişmişlik seviyelerine tek bir perspektiften yaklaşılmamış, sosyal ve ekonomik konular bir arada değerlendirilmiştir. Sağlık, nüfus, teknoloji ve ekonomik faktörler beraber ele alınarak bir veri madenciliği sınıflandırma uygulaması yapılmıştır.

Bu çalışmada ilk olarak veri madenciliği, veri madenciliği yöntemleri ve model

başarım ölçütleri açıklanmıştır. Çalışmanın üçüncü bölümünde gelişmişlik seviyeleri ile ilgili yapılmış çalışmalardan bahsedilmiş ve veri madenciliği sınıflandırma yöntemleri kullanılarak yapılan çalışmalara yer verilmiştir. Dördüncü bölümde ülkelerin gelişmişlik seviyelerinin sınıflandırılması üzerine bir uygulama yapılmıştır. Uygulamada veri madenciliğinde kullanılan Waikato Üniversitesinde Java programlama diliyle geliştirilmiş olan WEKA (Waikato Environment for Knowledge Analysis) programı tercih edilmiştir.

2. VERİ MADENCİLİĞİ

Günümüzde birçok kurumda çok sayıda veri üretilmektedir. Fakat bu verileri doğru ve verimli bilgiye dönüştürmekte sıkıntı yaşanmaktadır. Bu sebeplerden dolayı özel yöntemlere ihtiyaç vardır. Veri madenciliği bu ihtiyacı gidermek amacıyla oluşturulmuştur (Özkan, 2008:4).

Veri madenciliği, doğru kararı verebilmek amacıyla büyük veri setlerinden henüz bilinmeyen, elverişli ve mantığa uygun olan bilgiyi çıkarma işlemidir (Argüden ve Erşahin, 2008:15). Bu şekilde verilerin birbirleri ile olan bağlantısını gösterip geleceğe yönelik öngörülerde bulunabilmektedir. Veri madenciliği verilerinin tamamının üretildiği alandan gerekli metotlar kullanarak mevcutta bulunan veya gelecekte oluşabilecek bilinmeyen bilgileri açığa çıkarma sürecidir (Özkan, 2008:38). Veri madenciliği tek başına değil başka analiz yöntemleriyle kullanıldığı zaman anlamlı bir duruma gelmektedir. Veri madenciliği mevcut problemleri çözme ve geleceğe ışık tutma açısından önemli bir rol oynar (Argüden ve Erşahin, 2008:15). Veri madenciliği yaparken bilgi teknolojilerinden yararlanmak kaçınılmazdır, fakat en büyük unsurun insan faktörü olduğu göz ardı edilmemelidir (Asilkan, 2008:23).

2.1. Veri Madenciliğinde Kullanılan Modeller

Veri madenciliğinde modeller “Tahmin Edici Modeller” ve “Tanımlayıcı Modeller” olmak üzere ikiye ayrılır.

➤ Tahmin Edici Modeller

- Sınıflama ve Regresyon

➤ Tanımlayıcı Modeller

- Kümeleme
- Birliktelik Kuralı
- Diğer Metotlar

Tahmin edici modeller, geçmiş tecrübelerden kazanılan bilgilerin kullanımıyla, gelecekte meydana gelebilecek ihtimali olan durumlar için tahminde bulunabilmekte ve en iyi kararı verme aşamasında büyük ölçüde yardımcı olabilmektedir. Bu modellerde sonuçları ortada olan verilerden yola çıkılarak yeni bir model oluşturulması ve oluşturulan bu modellerden faydalanılarak sonuçları henüz bulunamamış veri kümelerinin sonuç değerleri hakkında tahminlerde bulunulması beklenir (Akpınar, 2006:5).

Tanımlayıcı modeller, öncesinde hiç bir hipoteze bağlı olmadan, veri kümesinde bulunan veriler arasında ne tür bağlar olduğunu anlamaya yardımcı olmaktadır. Analizleri yapan kişilerin veri tabanında çok geniş ölçüde bulunan bilgiyi incelemesi çok zordur bu yüzden bu görev veri madenciliği programlarına devredilir. Ortaya çıkan bilginin verimli ve nitelikli olması yapılan uygulamanın elverişliliğini ve gücünü arttırmaktadır (Güvenç, 2001). Tanımlayıcı modeller üç grupta incelenmektedir ve en çok kullanılanı ise kümeleme metodudur.

Bu çalışmada tahmin edici modeller arasında yer alan sınıflandırmayla ilgili bir uygulama yapılmıştır. Sınıflandırma ve regresyon modelleri, veri madenciliğinde en çok kullanılan yöntemdir. Elde bulunan verilerden yola çıkılarak öngörülerde bulunmak amacıyla kullanılmaktadır (Han ve Kamber, 2001). Bu iki model arasındaki fark ise; sınıflandırma, bulunan veriler önceden belli olan sınıflardan biriyle eşleşir

ve eğitim kümesi ile programa öğretilen sistem daha sonra test verilerinin kendi sınıflarını bulmasında yardımcı olur. Bankacılık, finans ve işletme gibi alanlarda sıklıkla kullanılır (Pryke, 1998). Sınıflamada aralarında herhangi bir yönden benzerlik olan değerlerin tahminini yaparken regresyon aynı devam eden değerlerin tahmininde bulunur. Regresyondaki amaç girdi ve çıktıyı bağdaştıracak bir model oluşturup en doğru tahminde bulunmaktır (Babaoğlu, 2015:25).

Veri madenciliği karar vermenin gerekli olduğu ve en iyi karar verme yönteminin arandığı her türlü alanda geniş bir kullanıma sahiptir. Bu alanlar sigortacılık, borsa, sanayi, iletişim, bankacılık, satış ve pazarlama, finans, elektronik ticaret, tıp ve

birçok endüstriyel çalışma alanı olabilmektedir.

2.2. Model Başarım Ölçütleri

Uygulanan algoritmaların başarı karşılaştırmaları incelenirken göz önüne alınan temel ölçütler vardır. Bunlar doğruluk-hata oranı, kesinlik ve F-ölçütüdür. Bu ölçütler modelin başarısını belirlemektedir. Yapılan sınıflandırmada belirlenen sınıflara doğru bir şekilde yerleştirilen örnek sayısı ve yanlış sınıfa yerleştirilen örnek sayısını temel almaktadır.

Yapılan uygulama neticesinde çıkan başarımların sonuçları karşılıklı matris yoluyla gösterilebilir. Tablo 1.'de sütunlar modelin tahminlemesini, satırlar ise test edilecek kümedeki örneklerin gerçek sayılarını göstermektedir.

Tablo 1: Hata Matrisi

		Öngörülen Sınıf	
		Sınıf=1	Sınıf=0
Doğru Sınıf	Sınıf=1	a	b
	Sınıf=0	c	d

a: TP (Gerçek Pozitif)

c: FN (Yanlış Negatif)

b: FP (Yanlış Pozitif)

d: TN (Gerçek Negatif)

Doğruluk-Hata Oranı: Kurulan modellerin başarısını değerlendirilmesinde bakılan en kolay ve en bilinen etken modelin doğruluk oranıdır. Doğru sınıflandırılmış örnek sayısını (TP +TN), toplam örnek sayısına (TP+TN+FP+FN) oranlayarak elde edilir. Hata oranı ise bu değer 1'e tamlayanıdır. Farklı bir deyişle yanlış sınıflandırılmış örnek sayısının (FP+FN), toplam örnek sayısına (TP+TN+FP+FN) oranıdır. 1 ve 2 numaralı eşitlikte görülmektedir.

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+FN+TN} \quad (1)$$

$$\text{Hata Oranı} = \frac{FP+FN}{TP+FP+FN+TN} \quad (2)$$

Duyarlılık: Doğru sınıflandırılmış pozitif örnek sayısının (TP), toplam pozitif örnek sayısına (TP+FN) oranı ile elde edilen değerdir. 3 numaralı eşitlikte görülmektedir.

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (3)$$

Kesinlik: Kesinlik, sınıfı 1 olarak tahminlenmiş True Pozitif örnek sayısının (TP), sınıfı 1 olarak tahminlenmiş tüm örnek sayısına (TP+FN) oranı ile elde edilen değerdir. 4 numaralı eşitlikte görülmektedir.

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (4)$$

F-Ölçütü: Uyguladığımız algoritmaların başarımlarını karşılaştırmamızda sadece

kesinlik ve duyarlılık ölçütleri yeterli değildir. Bu iki ölçütü bir arada hesaba almak daha doğru sonuçlar vermektedir. Bu amaçla F-ölçütü oluşturulmuştur. F-ölçütü, kesinlik ve duyarlılığın harmonik ortalaması olup, 5 numaralı eşitlikte görülmektedir.

$$F - \text{Ölçütü} = \frac{2 \times \text{Duyarlılık} \times \text{Kesinlik}}{\text{Duyarlılık} + \text{Kesinlik}} \quad (5)$$

Kappa İstatistiği: Sınıflandırma algoritmalarının doğru bilgi verip vermediğini araştırmak amacıyla yeni bir yöntem öne sürülmüştür. Yanlış bir şekilde meydana gelebilecek sınıflandırmaları karşılayacak niteliktedir. Cohen' in Kappa istatistiğine dayanan Kappa metodu, çok önemli olan sınıflandırma problemlerinin doğru olup olmadığını ölçmek amacıyla geliştirilmiştir (Ben-David, 2008). Kappa katsayısı sadece 0-1 aralığındaki değerleri alabilmektedir. '1' uyumun mükemmelliğini, '0' ise uyumun zayıf olduğunu ortaya koymaktadır (Eroğlu ve Palabaş, 2016: 514).

3. LİTERATÜR TARAMASI

Ülkelerin gelişmişlik düzeylerini ekonomi, sağlık, teknoloji, eğitim gibi birçok faktör etkilemektedir. Bu çalışma da birçok değişken dikkate alınarak ülkelerin gelişmişlik seviyeleri farklı algoritmalar yardımıyla tahmin edilmiştir. Yapılan literatür taraması iki farklı açıdan ele alınmıştır. Literatür taramasının ilk bölümünde ülkelerin gelişmişlik seviyeleri ile ilgili yapılmış çalışmalara, ikinci bölümünde ise veri madenciliği sınıflandırma yöntemleri kullanılarak yapılan çalışmalara yer verilmiştir.

Gelişmişlik seviyeleriyle ilgili yapılmış çalışmalar:

Göktaş Yılmaz (2005) çalışmasında Türkiye'de 2002 yılında gerçekleşen ekonomik büyümeye rağmen işsizlik oranlarının da artmasına bilimsel açıdan yaklaşarak bir yorum getirmeyi amaçlamıştır. Bunun amaç doğrultusunda ilk olarak teorik yaklaşımda bulunmuş

sonrasında ekonometrik analizler kullanmıştır.

Türkecul (2007) çalışmasında, Türkiye'de uzun bir süredir devam etmekte olan yüksek ve kronik enflasyonun tarım sektöründeki gelişmeleri hangi yönde etkilediğini zaman serisi verilerinden faydalanılarak araştırılmıştır. Bu uygulama 1981- 2005 dönemleri kullanılmıştır.

Erkekoğlu (2007) çalışmasında Türkiye'nin Romanya, Bulgaristan ve yirmi beş AB ülkesine göre göreceli gelişme düzeyini ele almıştır. AB ülkesi olan yirmi beş ülkenin yanında Türkiye, Romanya ve Bulgaristan da aday ülke olarak yapılan analize dahil edilmiştir. Yapılan karşılaştırma için sosyo-ekonomik gelişim göstergesi olan birçok unsur çalışmaya dahil edilmiştir.

Yumuşak ve Yıldırım (2009) çalışmalarında Türkiye'de sağlık endeksleri ve ekonomik büyüme arasında nasıl bir bağlantı olduğunu incelemişlerdir. Yaptıkları uygulamada Türkiye'nin 1980-2005 dönemlerindeki sağlık harcamaları, doğuştan yaşam beklentisi ve GSMH serilerini ele almışlardır.

Çestepe vd. (2012) çalışmalarında hizmet ticaretinin gelişmişlik seviyelerine göre ayrılan ülkelerin büyümesini hangi yönde etkilediğini görmek amacıyla deneysel bir araştırmasını yapmışlardır. 1980-2007 dönemleri arasındaki verileri kullanarak panel veri test ve tahmin metodlarından yararlanılmıştır.

Akça ve Ela (2012) çalışmalarında Türkiye'de kaba doğum hızına etki eden faktörler üzerinde ilk çalışmalarını gerçekleştirdikten sonra doğurganlık oranında meydana gelen ivmenin, aktif çalışan nüfusun yaş özelliğini etkilemesi açısından incelemişlerdir.

Koç (2013) çalışmasında beşeri sermayenin AB ülkesi üzerindeki ekonomik büyümeyi nasıl etkilediğini göstermek amacıyla yapılmıştır. 2012 AB ülkesi verileri kullanılmış yatay-kesit analiz yönteminden yararlanılmıştır.

Stec vd. (2014) çalışmalarında 2005 ve 2010 yılları arasında farklı faktörleri dikkate alarak AB’de sosyo-ekonomik gelişme düzeyini değerlendirmişlerdir.

Koşar Taş ve Örk Özel (2017) çalışmalarında Türkiye ve AB üyesi ülkelerini sosyo-ekonomik göstergeler açısından gelişmişlik seviyelerine göre kıyaslamışlardır. Çok değişkenli istatistik metotlarından olan faktör analizini kullanmışlardır.

Konak (2018) çalışmasında seçilmiş OECD üyesi ülkelerinin ve Türkiye’nin ileri düzey teknoloji barındıran ürün ihracatının gelişmişliğini, ilerleyişini ve ekonomik büyümeye etkisini araştırmayı amaçlamış, bu amaçla 1992-2016 yılları arası Dünya Bankası verilerinden faydalanmıştır.

Veri madenciliği sınıflandırma yöntemleri kullanılarak yapılan çalışmalar:

Danacı, Çelik ve Akkaya (2010) yaptıkları çalışmada Xcyt örüntü tanıma programından faydalanılarak doku ile ilgili veriler elde etmiş ve WEKA programından yararlanarak meme kanseri hücrelerini tahmin etmeyi ve bu hücrelere teşhis koymayı amaçlamışlardır.

Coşkun ve Baykal (2011) çalışmalarında WEKA veri madenciliği programı ile SEER veri kaynağı kullanmış ve sınıflandırma yöntemlerine ait algoritmalarından faydalanarak modeller oluşturmuştur. Ortaya çıkan modelleri karşılaştırarak algoritmalar arasında hangisinin tahmin etmekte daha başarılı olduğunu araştırmışlardır.

Phusavat, Ketsarapong, Ranjan, ve Lin (2011) yaptıkları çalışmada Yüksek Öğretim Komisyonunun (CHE) mevcut üniversite sınıflamasını iyileştirmek ve Tayland Yüksek Öğretim Sınıflandırma Modelini (THEC - modeli) geliştirmek amacıyla yapmışlardır.

Alan (2012) yaptığı bu çalışmada Cumhuriyet Üniversitesi Sosyal Bilimler Enstitüsü öğrencilerinden topladığı verilerle öğrenciler ile ilgili bazı özellikler tespit

etmeyi amaçlayan bir uygulama gerçekleştirmiştir. WEKA sınıflandırma algoritmalarını kullanarak en iyi sınıflandırma yapan algoritma araştırılmıştır.

Kaşıkçı ve Gökçen (2013) yaptıkları çalışmada metin sınıflandırmada çoğunlukla kullanılan başarı oranı yüksek olan algoritmaları incelemişlerdir. Sonrasında elektronik ticaret sitelerinin meydana getirdiği bir eğitim kümesi oluşturularak veriler belli birkaç algoritmayla eğitilmişlerdir. Bu algoritmalarının kullanımı için WEKA programından yararlanılmıştır.

Özel ve Topsakal (2014) yaptıkları çalışmada beton basınç dayanımını belirlemek amacıyla WEKA programında bulunan algoritmaların tamamını kullanarak modeller oluşturmuş ve korelasyon katsayısı en yüksek olan dokuz algoritmayı değerlendirmişlerdir.

Eroğlu ve Palabaş (2016) kronik böbrek hastalıklarını teşhis etmede WEKA sınıflandırma metotlarını ve kullandıkları algoritmaların başarılarını karşılaştırmak amacıyla bir çalışma gerçekleştirmişlerdir. Altı ayrı sınıflandırma metodu ve üç ayrı algoritma kullanmışlardır.

Şahin ve Kılıç (2016) yaptıkları çalışmada WEKA sınıflandırma yöntemlerini kullanmışlardır. Metin sınıflandırma amacıyla kullanılan yöntemler tıbbi dokümanlarda uygulanmış ve sınıflandırmadaki başarıları değerlendirmeye alınmıştır. Çok sınıflı ve çok etiketli olmak üzere iki veri seti kullanılmıştır.

Kaya vd. (2018) yaptıkları çalışmada WEKA programından yararlanarak hipermetropi ve miyopi göz kırma kusuru olan ve olmayan kişileri sınıflandırmayı amaçlamışlardır. Bu amaç doğrultusunda kişilerin sağ ve sol gözünden yatay ve dikey Elektrookülogram (EOG) sinyal verileri alınarak verilerde uygulanan veri madenciliği metotları ile sınıflandırmadaki başarı arttırılmaya çalışılmıştır.

Özmen vd. (2018) yaptıkları çalışmada kalp hastalığı verilerinden yararlanmış ve sınıflandırma yöntemlerinin iyi ve kötü yönlerini göz önüne alarak başarıları oranlarına göre bir karşılaştırma yapmışlardır. Çalışmada WEKA programı sınıflandırma yöntemlerinden yararlanılmıştır.

Yapılacak olan bu çalışmada ise yukarıdaki çalışmalardan farklı olarak ülkelerin gelişmişlik seviyeleri sınıflandırılacak ve kullanılan sınıflandırma algoritmalarının başarı düzeyleri karşılaştırılacaktır.

4. UYGULAMA

Uygulama bölümünde, yapılan çalışmanın amacı ve kısıtları anlatılarak sonrasında yöntem kısmına geçilmiştir.

4.1. Çalışmanın Amacı

Bu çalışma ülkelerin gelişmişlik seviyelerinin tahmininde kullanılan sınıflandırma algoritmalarının başarısını ölçmek ve performanslarını değerlendirmek için yapılmıştır. Araştırmada, 104 ülkenin 2016 yılı verileri dikkate alınmıştır. Araştırmanın yapıldığı süreçte 2017 yılı ve sonrasına ait veriler bulunmadığı için 2016 yılı uygulamaya dahil edilmiştir. Araştırmada doğum oranları, yetişkin ölüm oranları, bebek ölüm oranları, doğumda beklenen yaşam süreleri, yetersiz beslenme yaygınlıkları, ileri teknoloji ihracatları, sanayi (inşaat dahil) katma değerleri, nüfus artışları, kentsel nüfusları, kişi başına gayri safi milli hasılları (GSMH), gayri safi yurtiçi hasılları, toplam işsizlikleri (toplam işgücünün yüzdesi), enflasyonları (GSYH deflatörü) ve brüt sermaye oluşumları dikkate alınarak bir veri seti oluşturulmuştur. Gelişmişlik seviyelerini

belirlemede önemli olan faktörlerden biri de ülkelerin eğitim seviyeleridir. Fakat seçilen ülkelerin 2016 yılı eğitim verilerinde bulunan eksikliklerden dolayı çalışmaya dahil edilememiştir. Ülkeler ise gelişmiş ülkeler, gelişmekte olan ülkeler ve gelişmemiş ülkeler olmak üzere üç sınıfa ayrılmıştır.

4.2. Araştırmada Kullanılan Veriler ve Değişkenler

Bu çalışmada kullanılan veriler ülkeler ve gelişmişlik seviyeleri Uluslararası Para Fonu (IMF) kaynağından, ülkelere ait olan değişkenler ise Dünya Bankası Veri kaynağından alınmıştır. Çalışmada bazı ülkeler için gerekli olan değişken verilerinin bulunmadığından dolayı 104 ülke değerlendirmeye alınmıştır. Bu ülkelerin 32'si gelişmiş ülkeler, 57'i gelişmekte olan ülkeler ve 15'i gelişmemiş ülkelerdir. Gelişmişlik seviyelerini gösteren ülkelere ait brüt sermaye oluşumu, doğumda beklenen yaşam süreleri, yetersiz beslenme yaygınlıkları, ileri teknoloji ihracatları, sanayi (inşaat dahil) katma değerleri, nüfus artışları, kentsel nüfusları, kişi başına gayri safi milli hasılları (GSMH), gayri safi yurtiçi hasılları, toplam işsizlikleri (toplam işgücünün yüzdesi), enflasyonları (GSYH deflatörü) ve brüt sermaye oluşumları gibi değişkenler çalışmaya dahil edilmiştir.

Veri madenciliği analizi için kullanılan WEKA Programı için arff formatında veriler hazır hale getirilmiştir. Çalışmada kullanılan değişkenlerin tamamı sayısal değişkenlerdir. Ülkelerin gelişmişlik seviyeleri ise nominal değişken olarak ele alınmıştır.

Tablo 2: Ülkelerin Gelişmişlik Seviyelerine Göre Sınıflandırılması

Gelişmiş Ülkeler		Gelişmekte Olan Ülkeler				Gelişmemiş Ülkeler
Kanada	Litvanya	Meksika	Pakistan	Kolombiya	Tayland	Nepal
Almanya	İsviçre	Brezilya	Malezya	Ekvador	Vietnam	Kamboçya
Fransa	Avustralya	Mısır	Hırvatistan	Bolivya	Filipinler	Etiyopya
İngiltere	Çek	Şili	Polonya	Paraguay	Brunei Sultanlığı	Senegal
Amerika	Kıbrıs	Türkiye	Gürcistan	Uruguay	Sri Lanka	Sierra Leone
Hollanda	Slovenya	Ukrayna	Ermenistan	Jamaika	Lübnan	Burkina Faso
İtalya	İzlanda	Bosna-Hersek	Kazakistan	Fas	Kuveyt	Togo
Belçika	Yunanistan	Romanya	Dominik	Cezayir	Ürdün	Nijerya
İspanya	Lüksemburg	Bulgaristan	Peru	Gana	Birleşik Arap Emirlikleri	Benin
Portekiz	İrlanda	Macaristan	Nijerya	Kamerun		Madagaskar
Norveç	Slovakya	Hindistan	Guatemala	Güney Afrika		Mali
İsveç	İsrail	Suudi Arabistan	El Salvador	Namibya		Moritanya
Danimarka	Japonya	Rusya	Honduras	Botsvana		Uganda
Finlandiya	Avusturya	Arnavutluk	Nikaragua	Zimbabve		Lao PDR
Estonya	Kore	Moğolistan	Kosta Rika	Ermenistan		Ruanda
Letonya	Yeni Zelanda	Endonezya	Panama	Çin		

4.3. Bulgular ve Değerlendirme

WEKA programı kullanılarak uygulanan algoritmaların sonuçları Tablo 3.'de verilmiştir. Karşılaştırılacak algoritmaların

parametreleri kendi varsayılan değerleri olacak şekilde bırakılmıştır.

Tablo 3.'de kullanılan sınıflandırma algoritmalarının başarı oranları gösterilmiştir.

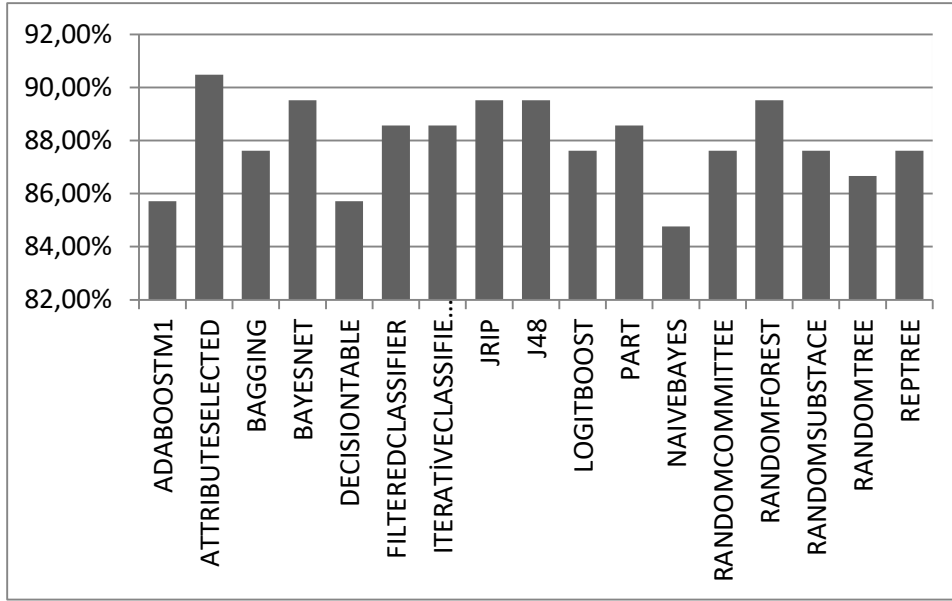
Tablo 3: Algoritmaların Model Başarım Ölçütleri

	<i>Doğruluk-Hata Oranı (%)</i>	<i>Kappa İstatistiği</i>	<i>Duyarlılık</i>	<i>Kesinlik</i>	<i>F-Ölçütü</i>
BAYESNET	89.52	0.8244	0.901	0.895	0.896
NAIVEBAYES	84.76	0.7421	0.852	0.848	0.849
ITERATIVECLASSIFIEROPTIMIZER	88.57	0.8014	0.885	0.886	0.885
ADABOOSTM1	85.71	0.7472	0.855	0.857	0.854
ATTRIBUTESELECTEDCLASSIFIER	90.48	0.8367	0.907	0.905	0.906
BAGGING	87.62	0.7857	0.875	0.876	0.875
FILTEREDCLASSIFIER	88.57	0.8014	0.885	0.886	0.885
LOGITBOOST	87.62	0.7891	0.880	0.876	0.878
RANDOMCOMMITTEE	87.62	0.7857	0.875	0.876	0.875
RANDOMSUBSPACE	87.62	0.7803	0.875	0.876	0.873
DECISIONTABLE	85.71	0.7534	0.858	0.857	0.855
JRIP	89.52	0.8201	0.894	0.895	0.894
PART	88.57	0.8066	0.893	0.886	0.888
J48	89.52	0.8191	0.896	0.895	0.895
RANDOMFOREST	89.52	0.8191	0.896	0.895	0.895
RANDOMTREE	86.67	0.7682	0.865	0.867	0.866
REPTREE	87.62	0.7857	0.875	0.876	0.875

Tablo 3.'de yer alan uygulama sonuçlarına göre algoritmaların performansları incelendiğinde Doğruluk Oranı(Correctly Classified Instances) yani doğru sınıflandırma oranı açısından bakıldığında Attribute Selected Classifier, Bayesnet, JRIP, J48 ve Random Forest algoritmaları

%89.52 ile birinci sırada yer almışlardır. Uygulanan diğer algoritmalar ise %84.76 ile %88.57 aralığında sonuçlar vermiştir. Şekil 1.'de ise sınıflandırma algoritmalarının doğruluk oranlarını gösteren grafik görülmektedir.

Şekil 1: Sınıflandırma Algoritmalarının Doğruluk Oranları



Kappa istatistiği yapılan tahminin doğruluk ölçüsünü gösterir. Bu istatistiğin sonuçları incelendiğinde ise kappa istatistiği değerleri en yüksek olan algoritmalar sırasıyla 0.8367 ile Attribute Selected Classifier, 0.8244 ile Bayesnet, 0.8201 ile JRIP, 0.8191 ile J48 ve Random Forest, 0.8066 ile PART, 0.8014 ile Iterative Classifier Optimizer ve Filtered Classifier algoritmalarıdır. Diğer algoritmalar ise 0.7421 ile 0.7857 aralığında yer almıştır.

Duyarlılık, Kesinlik ve F ölçütü değerlerine bakıldığı zaman en iyi sonucu 0.907 duyarlılık ölçütü, 0.905 kesinlik ölçütü ve 0.906 F-ölçütü ile Attribute Selected Classifier algoritması vermektedir. İkinci sırada yer alan algoritma 0.901 duyarlılık ölçütü, 0.895 kesinlik ölçütü ve 0.896 F-ölçütü ile Bayesnet' tir. Üçüncü sırada 0.896 duyarlılık ölçütü, 0.895 kesinlik

ölçütü ve 0.895 F-ölçütü ile J48 ve Random Forest'tir. Dördüncü sırada ise 0.894 duyarlılık ölçütü, 0.895 kesinlik ölçütü ve 0.894 F-ölçütü ile JRIP yer almaktadır. Diğer algoritmaların duyarlılık ölçütleri 0.852 ile 0,893, kesinlik ölçütleri 0.848 ile 0.886 ve F-ölçütleri ise 0.849 ile 0.888 aralığındadır. Ölçütlerin başarı oranları 1' e yaklaştıkça sonuç başarısı artış göstermektedir. Bu değerler dikkate alındığında uygulanan algoritmaların model başarımlık ölçütleri yüksektir.

5. SONUÇ

Ülkeler geçmişten bugüne kadar hem sosyal hem ekonomik açıdan çevresel faktörlerin etkilemesinden dolayı bir döngü içindedir. Ülkelerin gelişimleri zamana bağlı olarak ilerleme ve gerileme göstermektedir. Ülkelerde gelişmişlik seviyesini belirleyen sağlık, teknoloji, ekonomi, nüfus ve eğitim gibi pek çok unsur vardır. Ülke gelişmişliğini belirlerken bunları tek başına bir ölçüt olarak almak doğru değildir. Şimdiye kadar yapılan çalışmalarda da görüldüğü gibi bunların tek başına ele alınması tartışma konusu olmuş ve ülke gelişmişliğini belirlemede yeterli olmadığı görülmüştür. Çünkü ülke gelişimi bu faktörlerin hepsini bir bütün olarak kapsamaktadır.

Bu çalışma ülkelerin gelişmişlik seviyelerinin tahmininde kullanılan sınıflandırma algoritmalarının başarısını ölçmek ve performanslarını değerlendirmek için yapılmıştır. Bu amaçla verilere tam

olarak ulaşılabilen 104 ülke uygulamaya dahil edilmiştir. Gelişmişlik seviyeleri analiz edilen ülkelere ait brüt sermaye oluşumu, doğumda beklenen yaşam süresi, yetersiz beslenme yaygınlığı, ileri teknoloji ihracatı, sanayi (inşaat dahil) katma değeri, nüfus artışı, kentsel nüfus, kişi başına gayri safi milli hasıla (GSMH), gayri safi yurtiçi hasıla, toplam işsizlik (toplam işgücünün yüzdesi), enflasyon (GSYH deflatörü) ve brüt sermaye oluşumu değişkenleri dikkate alınmıştır.

Çalışmanın sonucunda model başarımları ölçütleri Doğruluk Oranı (Correctly Classified Instances), Kappa istatistiği, Duyarlılık (Recall), Kesinlik (Precision) ve F ölçütü dikkate alınarak algoritmalara değerlendirilmiştir. Değerlendirmenin sonucunda model başarımları ölçütleri açısından en başarılı algoritmanın Attribute Selected Classifier algoritması olduğu görülmüştür. Daha sonra sırayla Bayesnet, JRIP ve J48 algoritmaları almıştır.

KAYNAKÇA

- AKÇA, H. ve Ela, M. (2012). “Türkiye’de Eğitim, Doğurganlık ve İşsizlik İlişkisinin Analizi”, Maliye Dergisi, 163, 223-242.
- AKPINAR, H. (2006). “Veri Tabanlarındaki Bilgi Keşfi ve Veri Madenciliği”, İstanbul Üniversitesi İşletme Fakültesi Dergisi”, 29(1), 1-22.
- ALAN, M.A. (2012). “Veri Madenciliği ve Lisansüstü Öğrenci Verileri Üzerine Bir Uygulama”, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, 33, 165-174.
- ARGÜDEN, Y. ve Erşahin, B. (2008). Veri Madenciliği: Veriden Bilgiye, Masraftan Değere. Alkim Kağıt Sanayi ve Ticaret A.Ş., İstanbul.
- ASILKAN, Ö. (2008). Veri Madenciliği Kullanılarak İkinci El Otomobil Pazarında Fiyat Tahmini, Doktora Tezi, Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Antalya.
- BABAOĞLU, A. (2015). Veri Madenciliği Yöntemleri ve Bir Uygulama, Yüksek Lisans Tezi, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- BEN-DAVID, A. (2008). “Comparison of classification accuracy using Cohen’s Weighted Kappa”, Expert Systems with Applications: An International Journal, 34(2), 825-832.
- BRUNO, M. ve Easterly, W. (1998). “Inflation Crises and Long-run Growth”, Journal of Monetary Economics, 41(1), 3-26.
- ÇEŞTEPE, H., Vergil, H. ve Ergun, H. (2012). “Hizmet Ticaretinin Büyümeye Etkisi: Gelişmiş ve Gelişmekte Olan Ülkeler Üzerine Bir Panel Veri Analizi”, Business and Economics Research Journal, 3(4), 91-105.
- COŞKUN, C. ve Baykal, A. (2011). “Veri Madenciliğinde Sınıflandırma

- Algoritmalarının Bir Örnek Üzerinde Karşılaştırılması”, Akademik Bilişim’11 - XIII. Akademik Bilişim Konferansı Bildirileri, 1-8.
11. DANACI, M., Çelik, M. ve Akkaya, E. (2010). “Veri Madenciliği Yöntemleri Kullanarak Meme Kanseri Hücrelerinin Tahmin ve Teşhisi”, Akıllı Sistemlerde Yenilikler ve Uygulama Sempozyumu, 9-12.
 12. ERDEN Özsoy, C. ve Tosunoğlu, T.B. (2017). “GSYH’ nin Ötesi: Ekonomik Gelişmenin Ölçümünde Alternatif Metrikler”, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 26(1), 285-301.
 13. ERKEKOĞLU, H. (2007). “AB’ye Tam Üyelik Sürecinde Türkiye’nin Üye Ülkeler Karşısındaki Görelî Gelişme Düzeyi: Çok Değişkenli İstatistiksel Bir Analiz”, Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 14(2), 28-50.
 14. EROĞLU, K. ve Palabaş, T. (2016). “Kronik Böbrek Hastalığı Tespitinde Farklı Sınıflandırma Yöntemleri ve Farklı Topluluk Algoritmalarının Birlikte Kullanımının Sınıflandırma Performansına Etkisi”, Elektrik-Elektronik Mühendisliği Odası, 512-516.
 15. GÖKTAŞ Yılmaz, Ö. (2005). “Türkiye Ekonomisinde Büyüme ile İşsizlik Oranları Arasındaki Nedensellik İlişkisi”, İstanbul Üniversitesi İktisat Fakültesi Ekonometri ve İstatistik Dergisi, 2, 63-76.
 16. GÜVENÇ, E. (2001). Yüksek Öğretimde Öğrenci Performansının Veri Madenciliği Teknikleri ile Belirlenmesi, Yüksek Lisans Tezi, Endüstri Mühendisliği ABD, Fen Bilimleri Enstitüsü, Boğaziçi Üniversitesi, İstanbul.
 17. HAN, J. ve Kamber, M. (2001). Data Mining Concepts and Techniques, Morgan Kaufmann Publishers, Tokyo.
 18. KAŞIKÇI, T. ve Gökçen, H. (2013). “Metin Madenciliği İle E-Ticaret Sitelerinin Belirlenmesi”, Bilişim Teknolojileri Dergisi, 7(1), 25-32.
 19. KAYA, C., ErKaymaz, O., Ayar, O. ve Özer, M. (2018). “Elektrookülogram (EOG) Sinyallerinden Göz Kırama Kusurlarının Veri Madenciliği Teknikleri Kullanılarak Sınıflandırılması”, 26th IEEE Signal Processing and Communications Applications Conference, SIU 2018, 1-4.
 20. KESKİN, A. (2011). “Ekonomik Kalkınmada Beşeri Sermayenin Rolü ve Türkiye”, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, 25(3-4), 125-153.
 21. KOÇ, A. (2013). “Beşeri Sermaye ve Ekonomik Büyüme İlişkisi: Yatay Kesit Analizi ile AB Ülkeleri Üzerine Bir Değerlendirme”, Maliye Dergisi, 165, 241-258.
 22. KONAK, A. (2018). “Yüksek Teknoloji İçeren Ürün İhracatının İhracat Hacmi ve Ekonomik Büyüme Üzerine Etkisi; Seçilmiş OECD Ülkeleri ve Türkiye Örneği”, Yönetim, Ekonomi, Edebiyat, İslami ve Politik Bilimler Dergisi, 3(2), 56-80.
 23. KOŞAR Taş, Ç. ve Örk Özel, S. (2017). “Faktör Analizi Yöntemi ile Türkiye ve Avrupa Birliği Üyesi Ülkelerin Sosyo-Ekonomik Göstergeler Bakımından Gelişmişlik Düzeylerinin Karşılaştırılması”, Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 26(3), 60-77.
 24. KUBAR, Y. (2016). “Az Gelişmiş ve Gelişmekte Olan Ülkelerin Kalkınma Göstergeleri ile Ekonomik Büyüme Arasındaki İlişki: Bir Panel Veri Analizi (1995-2010) ”, Ardahan Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 4, 65-99.
 25. ÖZEL, C. ve Topsakal, A. (2014). “Veri Madenciliği Kullanarak Beton Basınç Dayanımının Belirlenmesi”,

- Cumhuriyet Üniversitesi Fen Fakültesi Fen Bilimleri Dergisi(CFD), 35(1), 43-57.
26. ÖZKAN, Y. (2008). Veri Madenciliği Yöntemleri. Papatya Yayıncılık Eğitim, İstanbul.
 27. ÖZMEN, Ö., Khdr, A. ve Avcı, E. (2018). “Sınıflandırıcıların Kalp Hastalığı Verileri Üzerine Performans Karşılaştırması”, Fırat Üniversitesi Mühendislik Bilimleri Dergisi, 30(3), 153-159.
 28. PRYKE, A. N. (1998) Data Mining Using Genetic Algorithms and Interactive Visualization, Ph. D. Thesis, Faculty of Science, University of Birmingham, Birmingham.
 29. PHUSAVAT, K., Ketsarapong, S., Ranjan, J. Ve Lin, B. (2011). “Developing a University Classification Model From Performance Indicators”, Performance Measurement and Metrics, 12(3), 183-213.
 30. STEC, M., Filip, P., Grzebyk, M.ve Pierscieniak, A. (2014). “Socio-Economic Development in the EU Member States-Concept and Classification”, Inzinerine Ekonomika-Engineering Economics, 25(5), 504–512.
 31. ŞAHİN, D.Ö. ve Kılıç, E. (2016). “Tıbbi Dokümanların Ayrıştırılmasında Kullanılan Sınıflandırma Algoritmalarının Karşılaştırılması”, International Conference on Computer Science and Engineering, 20-23.
 32. TABAN, S. ve Şengür, M. (2014). “Türkiye’ de Ar-Ge ve Ekonomik Büyüme”, Abant İzzet Baysal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 14(1), 355-376.
 33. TÜRKEKUL, B. (2007). “Türkiye’de Enflasyon- Büyüme İlişkisi: Tarım Sektörü İtibariyle Ekonometrik Bir Analiz”, Ege Üniversitesi Ziraat Fakültesi Dergisi, 44(1), 163-175.
 34. YAPRAKLI, S. (2007). “Enflasyon ve Ekonomik Büyüme Arasındaki İlişki: Türkiye İçin Eş-Bütünleşme ve Nedensellik Analizi”, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 10(2), 287-301.
 35. YUMUŞAK, İ.G. ve Yıldırım, D.Ç. (2009). “Sağlık Harcamaları İktisadi Büyüme İlişkisi Üzerine Ekonometrik Büyüme Üzerine Bir İnceleme”, Bilgi Ekonomisi ve Yönetimi Dergisi, 4(1), 57-70.