

Karar Ağaçları ve Bayes Ağları ile OECD Ülkelerindeki İntiharların Değerlendirilmesi

Şenol DEMİRCİ*, Duygu İÇEN**

Öz

İntiharlar, geçmişten günümüze farklı sıklıklarda ve sürekli karşılaşılan önemli bir sorundur. Bireyler, intihar girişiminde bulunarak kendi yaşamına kendi isteğiyle son vermeye çalışmaktadır. İntiharın birçok nedeni bulunmakla birlikte etkileri toplumdaki tüm fertler için yıkıcı düzeydedir. Bu çalışmada, aralarında Türkiye'nin de bulunduğu OECD üye ülkelerinin kişi başı gayri safi yurtiçi hasıla (GSYİH), işsizlik, alkol tüketimi, yıllık çalışma süresi, boşanma ve antidepresan kullanımı değişkenlerinin intihar değişkeni ile ilişkisinin incelenmesi amaçlanmıştır. İlişkiyi incelemek için veri madenciliği sınıflandırma tekniklerinden faydalanılmıştır. Veri madenciliği teknikleri arasında ise en yüksek doğru sınıflandırma değerlerini veren algoritmalarından yararlanılmıştır. Yapılan analizler neticesinde tüm değişkenlerin intihar değişkeni ile ilişkili olduğu belirlenmiştir. Buna göre işsizlik düzeyi, alkol tüketimi, yıllık çalışma süresi, boşanma hızı ve antidepresan kullanımı yüksek ve GSYİH'si düşük düzeyde olan ülkelerde intihar hızının yüksek veya orta düzeyde olabileceği belirlenmiştir. Bu özelliklerin belirlenmesi OECD ülkelerinin intihara yönelik alacakları önlemlerde ve geliştirecekleri sağlık ve sosyal politikalarda hangi özellikleri hedef alabileceğini gösterebilecektir.

Anahtar Sözcükler: Veri Madenciliği, Rastgele Orman, TAN, Sosyal Politika, Sınıflandırma

Evaluation of Suicides in OECD Countries with Decision Trees and Bayesian Networks

Abstract

Suicides are an important problem that has been encountered in different frequencies from past to present. Individuals are trying to end their own lives by attempting suicide. There are many causes of suicide but its effects are devastating for all individuals in the community. This study aimed to examine that OECD member countries including Turkey of the relationship between the variables of per capita gross domestic product (GDP), unemployment, alcohol consumption, annual working time, divorce rate, and antidepressant use with suicide variable. Data mining classification techniques were used to examine the relationship. Among the data mining techniques, we used the



Özgün Araştırma Makalesi (Original Research Article)

Geliş/Received: 04.07.2021

Kabul/Accepted: 14.12.2021

DOI: <https://dx.doi.org/10.17336/igusb.961070>

* Arş. Gör., Hacettepe Üniversitesi, İktisadi İdari Bilimler Fakültesi, Ankara, Türkiye,

E-posta: senoldemirci@gmail.com **ORCID** <https://orcid.org/0000-0001-8552-8151>

** Doç. Dr., Hacettepe Üniversitesi, Fen Fakültesi, Ankara, Türkiye,

E-posta: duyguicen@hacettepe.edu.tr **ORCID** <https://orcid.org/0000-0002-7940-5064>

algorithms giving the highest correct classification rates. As a result of the analyzes made, it was determined that all variables were associated with the suicide variable. Accordingly, it has been determined that the suicide rate may be high or moderate in countries with high unemployment level, alcohol consumption, annual working time, divorce rate, and antidepressant use and low GDP. Determining these features will be able to show what characteristics OECD countries will target in the measures they will take for suicide and the health and social policies they will develop.

Keywords: Data Mining, Random Forest, TAN, Social Policy, Classification

Giriş

İnsanlık tarihi boyunca intiharlar, farklı toplum, ülke ve gruplarda farklı sıklıklarda gerçekleşmiştir. İntihar sadece toplum ve ruh sağlığıyla ilgili bir problem değildir ayrıca, intiharın ekonomik ve kültürel yönü de bulunmaktadır. Her yıl neredeyse 800.000 kişi intihar sonucunda yaşamını yitirmekte ve daha fazla kişi intihar girişiminde bulunmaktadır (World Health Organization [WHO], 2018). İntihar, doğrudan ya da dolaylı olarak birey tarafından gerçekleştirilen, olumlu ya da olumsuz davranışlarından kaynaklanan ve ölüme sonuçlanan eylemler olarak tanımlanmaktadır. İntihar, birçok farklı nedenden kaynaklı olup yine birçok farklı şekilde gerçekleşmektedir (Durkheim, 2005). Bireyler, intihar girişiminde bulunarak kendi yaşamına kendi isteğiyle son vermeye çalışmaktadır. İntiharlar genellikle psikolojik bir rahatsızlığın sonucunda gerçekleşmektedir. Bununla birlikte, kişi bazen yaşamını tehdit eden fiziksel rahatsızlıklar, stres, uyuşturucu kullanımı, ekonomik sorunlar ve ailevi sorunlar gibi nedenlerden de intihar eyleminde bulunmaktadır (Piotrowski & Hartmann, 2019).

Birçok çalışmada, ülkelerin gelişmişlik düzeyi, işsizlik, alkol tüketimi, yalnız yaşama/boşanma, antidepresan kullanımı ve ruhsal rahatsızlıklar gibi durumların intihar davranışı ile ilişkili olduğu tespit edilmiştir (Stack, 1990; Barth vd., 2011; Kamat, Edgar, Niblock, McDowell, & Kelly, 2014; Laanani, Ghosn, Jouglu, & Rey, 2015; Khazaei Armanmehr, Nematollahi, Rezaeian, & Khazaei, 2017). Bu çalışmada, Organisation for Economic Co-operation and Development (OECD) üye ülkelerinin kişi başı GSYİH, işsizlik, alkol tüketimi, yıllık çalışma süresi, boşanma ve antidepresan kullanımı değişkenlerinin intihar değişkeni ile ilişkisinin incelenmesi ve ayrıca söz konusu değişkenler ile ulusal düzeyde intihar risk profilinin/özelliklerinin ortaya çıkarılması amaçlanmıştır. Bu çalışmada bağımsız değişkenlerin bağımlı değişken üzerindeki etkisi tahmin edilmek istendiği için veri madenciliğinin sınıflandırma yöntemlerinden faydalanılmıştır. Literatürde en fazla kullanılan sınıflandırma yöntemleri Karar Ağaçları, Bayes Ağları, Yapay Sinir Ağları, Destek Vektör Makineleri, K-en Yakın Komşu ve regresyondur (Koyuncugil & Özgülbaş, 2009). Bu çalışmada yer alan bağımlı değişken kategorik olduğu için sınıflandırma yöntemlerinden Karar Ağaçları ve Bayes Ağları kullanılmıştır. Aşağıda Karar Ağaçları ve Bayes Ağları daha detaylı incelenecektir.

Literatür Taraması

Literatürde veri madenciliği yöntemleri kullanılarak yürütülen çok sayıda çalışma bulunmaktadır. Çalış, Durmaz ve Gencer (2018), uçak seferlerindeki rötaları etkileyen faktörleri C5.0 algoritması ile incelemiştir. Çalış, Kayapınar ve Çetinyokuş (2014) tarafından C5.0 algoritması kullanılarak, farklı demografik özellikteki bireylerin bilgisayar ve internet güvenliği konusunda cevaplarından yararlanılarak çıkarımlarda bulunulmuştur. Doğan ve Türkoğlu (2008), demir eksikliği anemisi teşhisinde bulunmak

için Karar Ağaçlarını kullanmıştır. Bazila Banu ve Thirumalaikolundusubramanian (2018), Bayes Ağlarından en yüksek performansı gösteren TAN algoritmasını kullanarak göğüs kanseri hastalarını sınıflandırmıştır. Dogru ve Subasi (2018), trafik kazalarını tespit edebilmek için Rastgele Orman algoritmasından faydalanmıştır. C4.5, k-Ortalamalar, Destek Vektör Makineleri, Apriori, Beklenti Maksimizasyonu, PageRank, AdaBoost, K-en Yakın Komşu, Naive Bayes ve CART gibi algoritmalar veri madenciliğinde en fazla kullanılan algoritmalar olup eğitim, sağlık, bankacılık ve finansman, pazarlama, biyoenformatik, telekomünikasyon ve tarım gibi birçok alanda yaygın şekilde kullanılmaktadır (Wu vd., 2008; Smita ve Sharma, 2014).

Literatürde birçok alanda veri madenciliği yöntemleri kullanılmasına karşın, intihar ve veri madenciliği yöntemlerinin bir arada ele alındığı sınırlı sayıda çalışma bulunmaktadır. İlgen vd. (2009), Karar Ağaçları yöntemini kullanarak intihar açısından risk teşkil edebilecek bireysel özellikleri saptamıştır. Bae, Lee ve Lee (2015), CHAID algoritmasını kullanarak Güney Kore'de yaşayan ergen bireylerin intihar risk profilini ortaya çıkarmıştır. Berrouiguet vd. (2019), Karar Ağaçları yöntemini kullanarak intihar girişiminde bulunan bireyler ile ilişkili olabilecek faktörleri belirlemiştir. Baca-Garcia vd. (2007), Rastgele Orman algoritmasını kullanarak ailede intihar girişimi öyküsü ile ilişkili olabilecek değişkenleri tespit etmiştir. Amini, Ahmadiania, Poorolajal ve Amiri (2016), Lojistik Regresyon, Destek Vektör Makineleri, Karar Ağaçları ve Yapay Sinir Ağları yöntemleriyle intihar açısından yüksek risk grubunda yer alan bireylerin özelliklerini ortaya çıkarmıştır. Söz konusu çalışmaların ortak noktası, anketlerden elde edilen veriler ya da klinik kayıtlar ile yürütülmüş olmalarıdır. Bu sebepten ulusal düzeyde veriler kullanarak intihar ve veri madenciliği yöntemlerinin birlikte ele alındığı çalışmalara da ihtiyaç bulunmaktadır.

Veri Madenciliği ve Sınıflandırma

Veri madenciliği, elde edilen verilerden anlamlı bilgiler çıkaran ve bulan, veri içerisindeki gizli örüntüleri ve bağlantıları keşfeden, çeşitli değişkenler arasındaki ilişkiyi tespit eden ve bu sayede karar vermeye yardımcı olan bir yaklaşım olarak tanımlanmaktadır (Rygielski Wang, & Yen, 2002). Veri madenciliği, veri toplama, veri temizleme, model oluşturma, model testi ve uygulama gibi aşamaları içeren bir süreçtir (Seyrek & Ata, 2010). Veri madenciliği farklı alanlarda ve farklı amaçlarda kullanılmakta olup çoğunlukla sınıflandırma, kümeleme ve birliktelik işlevi bulunmaktadır. Sınıflandırma yöntemlerinin tahmin etme; kümeleme ve birliktelik yöntemlerinin ise tanımlama işlevi bulunmaktadır (Savaş, Topaloğlu, & Yılmaz, 2012).

Veri madenciliği içerisinde diğer yöntemlere kıyasla sınıflandırma yöntemleri daha fazla kullanılmaktadır. Sınıflandırma işlemi en basit şekliyle, bağımsız değişken değerleri belli iken bağımlı değişken değerini tahmin etme işlemidir. Karar Ağaçları, Bayes Ağları, Yapay Sinir Ağları ve Destek Vektör Makineleri gibi yöntemler en fazla kullanılan sınıflandırma yöntemleridir (Zaki, Meira Jr, & Meira, 2014). Birliktelik yöntemleri, birden çok tercihte ya da işlemde bulunan gözlemlerin, tercihleri ya da işlemleri arasındaki birliktelik yapılarını oluşturmaya yaramaktadır. Apriori algoritması, birliktelik yönteminde en yaygın kullanılan algoritmadır. Bir başka veri madenciliği işlevi olan kümelemede gözlem birimleri, benzerliklerine ya da benzeşmezliklerine göre gruplara ayrılmaktadır. Benzerliklerin bulunması için gözlemler arasındaki uzaklıklardan faydalanılmaktadır. Uzaklık değerlerine göre kümeler oluşturulmaktadır. Kümeler arasındaki mesafelerin ölçülmesinde uzaklık ölçüleri kullanılmaktadır. Kümelerin kendi içerisinde homojen, kümeler arasında ise heterojen olması sağlanmaktadır. En yaygın kullanılan kümeleme yaklaşımları Bölünmeli, Hiyerarşik, Yoğunluğa Dayalı, Grid Tabanlı, Model Tabanlı, Örüntü Tabanlı ve Kısıt Tabanlı'dır (Altunkaynak, 2017).

Karar Ağaçları

Karar Ağaçları, sınıflandırma ile ilgili problemlerin çözümünde kullanılan, ağaç görünümlü tahmin edici modellerdir. Oluşturulan veri tabanlarına kolay bir şekilde uyum sağlayabilen karar ağaçları; modelin kurgulanma ve yorumlanmasının kolay olması, kurulum maliyetlerinin az olması ve güvenilir olmasından günümüzde çok sık kullanılmaktadır (Akpınar, 2000). Karar Ağacı, ağaç tabanlı bir sınıflandırma modeli oluşturan parametrik olmayan bir yöntemdir. Karar Ağacında bağımlı değişkenin tahmin değeri bağımsız değişkenlerin değerleri üzerine dayanmaktadır. Karar Ağacının asıl amacı, bağımsız değişkenler aracılığıyla bağımlı değişken için tahmini bir model oluşturmaktır (Shi, 2013). Bir Karar Ağacı, düğüm, dal ve yapraklardan oluşmaktadır. İlk dallandırmanın olduğu yer kök düğümdür. Dalların son bulduğu düğümler ise yapraklardır. Karar Ağacındaki düğüm ve yaprakların içerisinde genellikle bağımlı değişken dağılım bilgisi yer alırken, dallar üzerinde ise dallandırma yapılan bağımsız değişkenin düzeyi ve sınıf aralığı bulunur. Karar Ağaçlarından yola çıkılarak düğümlere ilişkin yorumlar oluşturulabileceği gibi karar kuralları da oluşturulabilmektedir (Altunkaynak, 2017).

Sınıflandırma problemlerinin çözümünü sağlayan Karar Ağaçları için bazı algoritmaların geliştirildiği bilinmektedir. En yaygın kullanılan algoritmalar, CHAID (Otomatik Ki-Kare İlişki Belirleyicisi), C&RT (Sınıflandırma ve Regresyon Ağaçları), QUEST (Hızlı, Yansız, Etkili İstatistiksel Ağaçlar), Rastgele Orman, C4.5 ve C5.0'dır. C5.0 algoritması, C4.5 algoritmasının daha gelişmiş sürümüdür. C5.0 algoritması C4.5 gibi, ağaç ve kural tabanlı olup kök algoritmalarının çoğu benzerdir. C5.0 algoritmasında daha küçük ağaçlar geliştirilebilmekte ve kategorik değişkenlerden yararlanılmaktadır (Kuhn & Johnson, 2013). 1980 yılında Kaas tarafından geliştirilen CHAID algoritması, geniş veri kümeleri için kullanılan bir algoritmadır. Söz konusu algoritmada verilerin daha homojen alt gruplara bölünmesi sağlanır. C&RT algoritması, 1984 yılında, Breiman tarafından sınıflama ve regresyon amacıyla geliştirilmiştir. C&RT algoritmasında, verilerin iki alt düğüme sürekli bir şekilde ayrılmasını sağlayabilmek için bağımsız tüm değişkenler kullanılır (Ture, Tokatlı, & Kurt, 2009). Rastgele Orman algoritması, çok sayıda ağaç çıktısını tek bir sonuca erişmek için birleştirmektedir (IBM, 2020). Bahsi geçen algoritma kategorik değişkenler üzerinde kullanılabilen ve oluşturulan ağaçlar için budama işlemine gerek duyulmamaktadır. Ayrıca orijinal veri setini bölmeden modeli test edebilmektedir (Akman, Genç, & Ankaralı, 2011). Karar ağaçları için geliştirilen bu modeller arasında birtakım farklılıklar vardır. Bu farklılıklardan biri ağaçların üretim şeklidir. Örneğin, CHAID algoritmasında ağaç şekli, ikili olmayan çoklu ağaçlar şeklindeyken, C4.5 algoritmasında ise ağaçlar ikilidir (Ture, Tokatlı, & Kurt, 2009). C4.5 tek değişkenli bir karar ağacı iken, C&RT çok değişkenlidir. Rastgele Orman, CHAID ve C&RT algoritmasında sürekli ve kategorik tüm değişken türleriyle ilgilenebilmekteyken; QUEST algoritmasında değişken türlerinin bir kısmıyla ilgilenebilmektedir (Akman vd., 2011; Ratner, 1998).

Bayes Ağları

Bayes Ağları, olasılıksal grafik modelleri ailesine ait olup inanç ağları (belief networks) olarak da bilinmektedir. Olasılıksal grafik modelleri, ilgilenilen problemin kesin olmayan tanım kümesi hakkındaki bilgiyi betimlemesi amacıyla kullanılmaktadır (Ben-Gal, 2007). Bayes Ağlarında değişkenler düğümler, değişkenler arası olasılıksal bağımlılık ilişkileri ise yönlü oklar aracılığıyla gösterilmektedir. Genellikle Bayes Ağları, değişkenler arası ilişkilerin gösteriminin yapıldığı grafik ve değişkenlere ait koşullu

olasılık tabloları olmak üzere iki parçadan oluşmaktadır. Grafikselsel parça ağı yapısını ortaya çıkarmaktadır. Ağdaki iki düğümün birbirine ok ile bağlandığı durumda okun başlangıcındaki düğüm ebeveyn (parent) düğüm, okun bitişinde bulunan düğüm ise çocuk (child) düğüm olarak adlandırılır (Çinicioğlu, Atalay, & Yorulmaz, 2013). Naive Bayes, Tree Augmented Naive Bayes (TAN), Markov Blanket (MB), BN Augmented Naive Bayes (BAN), Bayesian Multinet ve Genel Bayes Ağı (GBN) gibi sınıflayıcılar Bayes Ağları yapısını belirlemek amacıyla kullanılmaktadır (Cheng & Greiner, 2001).

Yöntem

Araştırmanın Amacı ve Değişkenler

Bu çalışmada, OECD ülkelerinin intihar hızları ile ilişkili olabilecek sosyo-ekonomik faktörlerin belirlenmesi ve ulusal düzeyde risk profillerinin/özelliklerinin ortaya çıkarılması amaçlanmıştır. Çalışmanın yöntemi nicel araştırma yöntemlerinden, veri madenciliği teknikleridir.

Çalışmanın bağımlı/hedef değişkeni ülkelerin her 100.000 ölüm içerisindeki intihar hızıdır. İntihar verisi düşük, orta ve yüksek olarak kategorilere ayrılmıştır ve analizler bu şekilde gerçekleştirilmiştir. Çalışmada yararlanılan bağımsız değişkenler, intihar üzerinde etkisi olduğu farklı çalışmalarca belirlenen kişi başı GSYİH, işsizlik, alkol tüketimi, yıllık çalışma süresi, boşanma hızı ve antidepresan kullanımınıdır (Stack, 1990; Andrés, Halicioğlu, & Yamamura, 2011; Barth vd., 2011; Kamat vd., 2014; Laanani vd., 2015; Khazaei vd., 2017). Bağımsız değişkenlerin tamamı nicel yapıdadır.

Örnekleme

Çalışmaya OECD'ye üye 36 ülke (Amerika Birleşik Devletleri, Almanya, Avustralya, Avusturya, Belçika, Çek Cumhuriyeti, Danimarka, Estonya, Finlandiya, Fransa, Hollanda, İngiltere, İrlanda, İspanya, İsrail, İsveç, İsviçre, İtalya, İzlanda, Japonya, Kanada, Güney Kore, Letonya, Litvanya, Lüksemburg, Macaristan, Meksika, Norveç, Polonya, Portekiz, Slovakya, Slovenya, Şili, Türkiye, Yeni Zelanda, Yunanistan) ve OECD ile iş birliği yapan 1 ülke (Rusya), toplamda 37 ülke dahil edilmiştir. OECD üyesi ülkeler Dünya GSYİH'sinin yüzde 63'ünü, dünya ticaretinin dörtte üçünü ve dünya nüfusunun yüzde 18'ini oluşturmaktadır (U.S. Mission to the Organization For Economic Cooperation & Development [USOEC], 2019).

Veri Kaynağı

Araştırmada OECD üye ve ortak ülkelerin (<https://data.oecd.org>) verilerinden yararlanılmıştır. Araştırma kapsamında yararlanılan bağımlı değişkenler, bağımsız değişkenler, açıklamalar, alındıkları veri tabanı ve yıl Tablo 1'de verilmiştir.

	Değişken Adı	Açıklama	Veri Tabanı (Yıl)
--	---------------------	-----------------	--------------------------

Bağımlı Değişken	İntihar_hızı	Her 100.000 ölüm içerisindeki intihar hızı	OECD (2015)
Bağımsız Değişken	GSYİH_kişibaşı	Kişi başına düşen GSYİH \$ (satın alma gücü paritesi-PPP)	OECD (2015)
	İşsizlik	İş gücü içerisindeki işsizlik oranı	OECD (2015)
	Alkol_tüketimi	15 yaş ve üzeri bireylerin bir yıl içerisindeki kişi başı alkol tüketimi (Litre)	OECD (2015)
	Yıllık çalışma_süresi	Yıllık ortalama çalışma süresi (Yılda toplam çalışılan saat/Yılda ortalama çalışan kişi sayısı)	OECD (2015)
	Kaba boşanma_hızı	Yılda 1000 kişi başına boşanma hızı	OECD (2015)
	Antidepresan	1000 kişi başına düşen günlük doz	OECD (2015)

Tablo 1. Çalışma Kapsamında Yararlanılan Değişkenlere İlişkin Bilgiler

Verilerin Analizi

Bu çalışmada, Karar Ağaçlarının uygulanmasında C&RT, C5.0 ve Rastgele Orman algoritmaları; Bayes Ağların uygulanmasında Tree Augmented Naive Bayes (TAN) ve Markov Blanket (MB) ağ yapıları kullanılmış ve IBM SPSS Modeller v18.2 veri madenciliği yazılımından yararlanılmıştır.

Algoritmaların performansının ölçülmesinde genel doğruluk oranından faydalanılmıştır. Genel doğruluk oranı, konfüzyon matrisinden yararlanılarak hesaplanmaktadır. Tablo 2'de örnek bir konfüzyon matrisi verilmiştir. Genel doğruluk oranı, $TP+TN/N$ formülü ile hesaplanmaktadır. Konfüzyon matrisine göre (Altunkaynak, 2017):

- TP: Doğru pozitiflerin sayısı
- TN: Doğru negatiflerin sayısı
- FP: Yanlış pozitiflerin sayısı
- FN: Yanlış negatiflerin sayısı

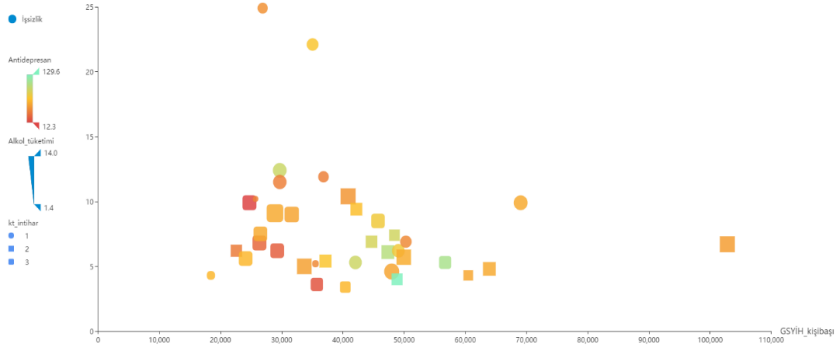
		Gerçek durum	
		Doğru (+)	Yanlış (-)
Algoritmanın tahmini	Doğru (+)	TP	FP
	Yanlış (-)	FN	TN

Tablo 2. Konfüzyon Matrisi

Algoritmaların yürütülmesi sonucunda intihar hızı için en önemli bağımsız değişkenin tespit edilmesinde tahmin değişkeni önem (predictor importance) değerinden faydalanılmıştır. Tahmin değişkeni önem değeri, modelin tahmin edilmesinde her bir değişkenin göreceli önemini belirtmek için kullanılmaktadır. Değerler göreceli olduğundan, tüm değişkenlere ait tahmin değişkeni önem değerlerinin toplamı 1,0'ı vermektedir. Söz konusu değer, algoritmanın performansının ölçülmesinde kullanılan genel doğruluk oranı ile ilişkili değildir. Daha ziyade her değişkenin intihar hızını tahmin etmedeki önemi ile ilgilidir (IBM, 2021).

Bulgular

Şekil 1'de ülkelerin işsizlik, kişi başı GSYİH, antidepresan kullanımı, alkol tüketimi ve intihar hızına göre tanımlayıcı bilgileri yer almaktadır. X ekseninde kişi başı GSYİH ve Y ekseninde işsizlik oranı yer almaktadır. İntihar hızı kategorilere 1=Düşük ($0 < \text{intihar hızı} \leq 10,6$), 2=Orta ($10,6 < \text{intihar hızı} \leq 13,1$) ve 3=Yüksek ($13,1 < \text{intihar hızı} \leq 29,0$) ayrılmış olup her kategori farklı şekilde gösterilmektedir. Antidepresan kullanımı Şekil 1'de renk skalasında verilmiştir. Alkol tüketimi arttıkça şekillerin boyutları da büyümektedir.



Şekil 1. OECD Ülkelerinin Tanımlayıcı Bilgileri

Şekil 1 incelendiğinde ülkelerin çoğunluğunun GSYİH'sinin 20.000\$ ile 50.000\$ arasında, işsizlik oranlarının %5 ile %15 arasında yer aldığı ve bu aralıkta yer alan ülkelerin intihar hızı, antidepresan kullanımı ve alkol tüketiminin ortalama düzeyde ve yüksek olduğu görülmektedir.

Tablo 3'de çalışma kapsamındaki değişkenlere ve ülkelere göre OECD tanımlayıcı istatistikleri yer almaktadır. OECD ülkelerinin her 100.000 ölüm içerisindeki intihar hızı ortalaması 12,3, kişi başı GSYİH'sı 40797,6\$, iş gücü içerisindeki işsizlik oranı 7,9, 15 yaş üstü kişi başı alkol tüketimi 9,0 litre, yıllık çalışma süresi 1705,5 saat, yılda 1000 kişi başına kaba boşanma hızı 2,0 ve 1000 kişi başına günlük antidepresan dozu 61,7'dir.

1. Değişkenler	2. Ortalama±Standart Sapma
3. İntihar hızı	4. 12,3±5,3
5. Kişi başı GSYİH	6. 40797,6±16068,9
7. İşsizlik oranı	8. 7,9±4,4
9. Alkol tüketimi	10. 9,0±2,5
11. Yıllık çalışma süresi	12. 1705,5±196,1
13. Kaba boşanma hızı	14. 2,0±0,7
15. Antidepresan kullanımı	16. 61,7±25,4

Tablo 3. OECD Ülkelerinin Tanımlayıcı İstatistikleri

Yukarıdaki bilgiler ışığında OECD ülkelerindeki intihar hızını tahmin etmek için daha önce bahsedilen yöntemler IBM SPSS Modeler v18.2 versiyonu kullanılarak sonuçlar elde edilmiştir (analiz için oluşturulan akış Ek 1'de verilmiştir).

Karar Ağaçları Sonuçları

Bağımlı değişken üç kategoriden oluştuğu için bu çalışmada model başarı ölçütleri olarak genel doğruluk oranından yararlanılmıştır. Tablo 4'de genel doğruluk oranları yer almaktadır.

Algoritmalar	Genel Doğruluk Oranı
C&RT	0,7297
C5.0	0,8919
Rastgele Orman	0,8919

Tablo 4. Karar Ağaçları Algoritmalarından Elde Edilen Genel Doğruluk Oranları

Tablo 4'de yer alan genel doğruluk oranlarına göre Karar Ağaçları algoritmaları içinde en iyi algoritmaların C5.0 ve Rastgele Orman olduğu belirlenmiştir. Bu doğrultuda en yüksek doğru sınıflandırma değerlerini veren bu algoritmaların doğrulama matrisleri ve bu matrislerden elde edilen yorumlar aşağıda verilmiştir.

		TAHMİN			TOPLAM
		1=Düşük	2=Orta	3=Yüksek	
GERÇEK	1=Düşük	13	0	0	13
	2=Orta	2	10	1	13
	3=Yüksek	0	1	10	12
TOPLAM		15	11	11	37

Tablo 5. C5.0 Algoritmasının Doğrulama Matrisi

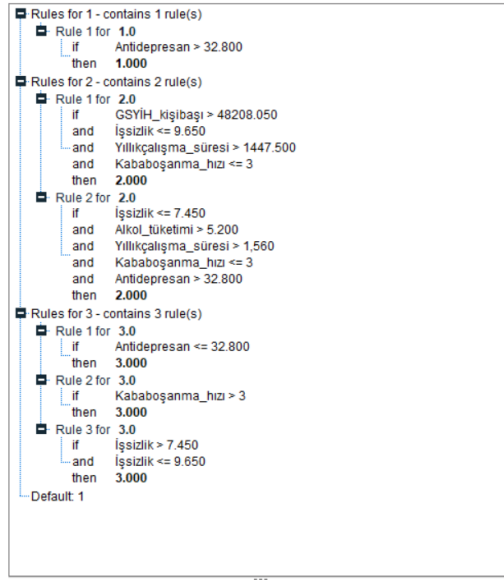
Tablo 5, C5.0 algoritması için yapılan doğrulama matrisi sonuçlarını göstermektedir. Doğrulama matrisi gerçek ve tahmin sonuçlarına göre ülkelerin %35,13'ünün düşük, %27,02'sinin orta ve %27,02'sinin yüksek intihar kategorilerinde sınıflandırıldığı belirlenmiştir. Sonuç itibariyle ülkelerin C5.0 algoritması toplam doğru sınıflandırma yüzdesi %89,1 olarak tespit edilmiştir.

		TAHMİN			TOPLAM
		1=Düşük	2=Orta	3=Yüksek	
GERÇEK	1=Düşük	13	0	0	13
	2=Orta	1	10	2	13
	3=Yüksek	0	1	10	12
TOPLAM		14	11	12	37

Tablo 6. Rastgele Orman Algoritmasının Doğrulama Matrisi

Tablo 6, Rastgele Orman algoritması için yapılan doğrulama matrisi sonuçlarını göstermektedir. Doğrulama matrisi gerçek ve tahmin sonuçlarına göre ülkelerin %35,13'ünün düşük, %27,02'sinin orta ve %27,02'sinin yüksek intihar kategorilerinde sınıflandırıldığı belirlenmiştir. Sonuç itibariyle ülkelerin Rastgele Orman algoritması toplam doğru sınıflandırma yüzdesi %89,1 olarak tespit edilmiştir.

Genel doğruluk oranı en yüksek C5.0 ve Rastgele Orman algoritmaları olduğu için kural çıkarımı yapılırken bu algoritmalarından elde edilen sonuçlar kullanılmıştır. Ayrıca C5.0 algoritmasının Modeller 18 akışındaki algoritma seçenekleri kullanılarak kurallar (rule set) oluşturulmuştur.



Şekil 2. C5.0 Algoritması Kurallar Seti Çıktısı

Şekil 2’de C5.0 algoritması kullanılarak elde edilen kurallar yer almaktadır. C5.0 algoritması tahmin değişkeni önem değerlerine göre intihar hızı için en önemli bağımsız değişkenin işsizlik olduğu belirlenmiştir. C5.0 algoritması ile üretilen kurallar şunlardır:

Kural 1: Eğer bir OECD ülkesi düşük intihar (0 ile $\leq 10,6$ arası) hızına sahip ise 1000 kişiye düşen antidepresan dozunun 32,8 dozdan fazla olması beklenmektedir.

Kural 2.1: Eğer bir OECD ülkesi orta düzeyde ($10,6 >$ ile $\leq 13,1$ arası) intihar hızına sahip ise kişi başı GSYİH’si 48208,05\$’dan fazla, işsizlik oranı %9,65’den ve daha az, yıllık çalışma süresi 1447,5 saatten fazla ve yılda 1000 kişi başına boşanma hızı 3 ve daha az olması beklenmektedir.

Kural 2.2: Eğer bir OECD ülkesinde işsizlik oranı %7,45 ve daha az, alkol tüketimi 5,2 litreden fazla, yıllık çalışma süresi 1560 saatten fazla, yılda 1000 kişi başına boşanma hızı 3 ve daha az, 1000 kişiye düşen antidepresan dozu 32,8’den fazla ise o ülkede intihar hızının orta düzeyde ($10,6 >$ ile $\leq 13,1$ arası) olması beklenmektedir.

Kural 3: Eğer bir OECD ülkesinde 1000 kişiye düşen antidepresan dozu 32,8 ve daha az, yılda 1000 kişi başına boşanma sayısı 3’den fazla ve işsizlik oranı %7,45’ ile $\leq 9,65$ arasında ise intihar hızının yüksek düzeyde ($13,1 >$ ile $\leq 29,0$ arası) olması beklenmektedir.

Karar Kuralları	İntihar Hızı Kategorisi
$(4,3 < \text{işsizlik oranı} \leq 7,4) \& (\text{boşanma hızı} \leq 2,1) \& (7,2 < \text{alkol tüketimi} \leq 10,4)$	Düşük
$(\text{antidepresan kullanımı} \leq 65,0) \& (\text{işsizlik oranı} > 5,0) \& (\text{kişi başı GSYİH} \leq 35054,49) \& (\text{alkol tüketimi} > 10,4)$	Yüksek
$(\text{kişi başı GSYİH} \leq 37158,23) \& (\text{alkol tüketimi} \leq 10,4)$	Düşük

(kişi başı GSYİH>35054,49) & (alkol tüketimi>10,4)	Düşük
(antidepresan kullanımı>65,0) & (işsizlik oranı≤9,0) & (kişi başı GSYİH≤47351,05) & (boşanma hızı >1,7)	Orta

Tablo 7. Rastgele Orman Algoritması Karar Kuralları Çıktısı

Tablo 7, Karar Ağaçları algoritmalarından olan Rastgele Orman ile elde edilen karar kurallarını göstermektedir. C5.0 algoritmasında olduğu gibi Rastgele Orman algoritmasında da tahmin değişkeni önem değerlerine göre intihar bağımlı değişkeni için en önemli bağımsız değişken işsizliktir. Rastgele Orman ile üretilen karar kuralları şunlardır:

Kural 1: İşsizlik oranı %4,3'ten fazla ve %7,4 ve daha az, yılda 1000 kişi başına boşanma hızı 2,1 ve daha az ve alkol tüketimi 7,2 litreden fazla ve 10,3 litre ve daha az olan OECD ülkelerinin intihar hızının düşük olması beklenmektedir.

Kural 2: 1000 kişiye düşen antidepresan dozu 65,0 ve daha az, işsizlik oranı %5,0'dan büyük, kişi başı GSYİH'si 35054,49\$ ve daha az ve alkol tüketimi 10,4 litreden fazla olan OECD ülkelerinin intihar hızının yüksek olması beklenmektedir.

Kural 3: Kişi başı GSYİH'si 37158,23\$ ve daha az ve alkol tüketimi 10,4 litre ve daha az olan OECD ülkelerinin intihar hızının düşük olması beklenmektedir.

Kural 4: Kişi başı GSYİH'si 35054,49\$'dan ve alkol tüketimi 10,4 litreden fazla olan OECD ülkelerinin intihar hızının düşük olması beklenmektedir.

Kural 5: 1000 kişiye düşen antidepresan dozu 65,0'dan fazla, yılda 1000 kişi başına boşanma hızı 1,7'den fazla, kişi başı GSYİH'si 47351,05\$ ve daha az ve işsizlik oranı %9,0 ve daha az olan OECD ülkelerinin intihar hızının orta düzeyde olması beklenmektedir.

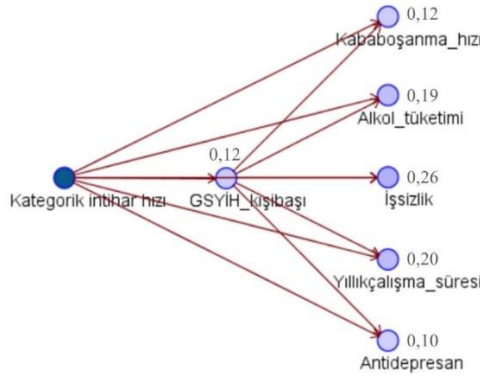
Bayes Ağları Sonuçları

SPSS IBM Modeller ile intihar hızı değeri Bayes Ağları ile tahmin edilmiştir. Bayes Ağlarından, TAN ve Markov Blanket ağ yapılarının vermiş oldukları genel doğruluk oranları Tablo 8'de verilmiştir.

TAN	0,8378
Markov Blanket	0,7838

Tablo 8. Bayes Ağları ile Elde Edilen Genel Doğruluk Oranları

TAN ağ yapısı en yüksek doğruluk oranını verdiği için intihar hızı ile ilgili yorumlar bu ağ yapısı üzerinden yapılacaktır. Şekil 3'te TAN ağ yapısına göre değişkenlerin önem değerleri verilmiştir. Buna göre önem düzeyi en yüksek değişken işsizlik oranıdır. Sonrasında ise sırasıyla yıllık çalışma süresi, alkol tüketimi, kaba boşanma hızı, kişi başı GSYİH ve antidepresan gelmektedir.



Şekil 3. TAN Ağ Yapısı Çıktısı

Şekil 3'te nedensellik ilişkilerine göre oluşturulmuş intihar hızı düğümü ve ilişkili düğümlerine ait ağ yapısı görülmektedir. Kişi başı GSYİH düğümü kaba boşanma hızı, alkol tüketimi, işsizlik oranı, yıllık çalışma süresi ve antidepresan düğümlerinin ebeveyn düğümüken GSYİH aynı zamanda intihar hızının çocuk düğümüdür. Bu ağ yapısına göre hesaplanan koşullu olasılıklar aşağıda verilmiştir.

Ebeveyn Düğüm	Çocuk Düğüm			
	Kişi Başı GSYİH			
Kategorik İntihar Hızı	≤35.313,89	>35.313,89- ≤52.189,74	>52.189,74- ≤77.503,51	≥77.503,51
1=Düşük	0,46	0,46	0,08	0,00
2=Orta	0,15	0,62	0,15	0,08
3=Yüksek	0,64	0,27	0,09	0,00

Tablo 9. Kişi Başı GSYİH Düğümü Hesaplanan Koşullu Olasılıklar

Tablo 9'da kişi başı GSYİH düğümüne ait koşullu olasılıklar yer almaktadır. Düşük intihar hızına sahip olan ülkelerin kişi başı GSYİH'sinin 35.313,89\$' dan az olması olasılığı %46'dır. Orta düzeyde intihar hızına sahip olan ülkelerin kişi başı GSYİH'sinin 35.313,90\$ ile 52.189,74\$ arasında olması olasılığı %62'dir. Yüksek intihar hızına sahip ülkelerin GSYİH'sinin 35.313,89\$'dan az olması olasılığı %64'tür.

Ebeveyn Düğümler	Kişi Başı GSYİH	Kategorik İntihar Hızı	Çocuk Düğüm			
			İşsizlik Oranları			
			≤7,7	7,7-≤12	12-≤18,45	>18,45
≤35.313,89	1=Düşük		0,17	0,33	0,17	0,33
≤35.313,89	2=Orta		1,00	0,00	0,00	0,00
≤35.313,89	3=Yüksek		0,57	0,43	0,00	0,00
>35.313,89-≤52.189,74	1=Düşük		0,83	0,17	0,00	0,00
>35.313,89-≤52.189,74	2=Orta		0,75	0,25	0,00	0,00
>35.313,89-≤52.189,74	3=Yüksek		0,67	0,33	0,00	0,00
>52.189,74-≤77.503,51	1=Düşük		0,00	1,00	0,00	0,00
>52.189,74-≤77.503,51	2=Orta		1,00	0,00	0,00	0,00
>52.189,74-≤77.503,51	3=Yüksek		1,00	0,00	0,00	0,00
>77.503,51	2=Orta		1,00	0,00	0,00	0,00

Tablo 10. İşsizlik Oranları Değişkeni için Hesaplanan Koşullu Olasılıklar

Tablo 10 incelendiğinde kişi başı GSYİH'sı 35.313,89\$'dan az ve intihar hızının orta düzeyde olduğu ülkelerde, işsizlik oranının 7,7'den daha az olması olasılığı %100'dür. Kişi başı GSYİH'sı 52.189,74\$'dan fazla olup 77.503,51\$'dan küçük olan ve intihar hızı oranının yüksek olduğu ülkelerde işsizlik oranının 7,7'den küçük olması olasılığı %100'dür.

Ebeveyn Dügümler		Çocuk Dügüm				
Kişi Başı GSYİH	Kategorik İntihar Hızı	Antidepresan Kullanımı (Doz)				
		≤35,7	>35,7-≤59,2	>59,2-≤82,6	>82,6-≤106,1	>106,1
≤35.313,89	1=Düşük	0,0	0,5	0,3	0,1	0,0
≤35.313,89	2=Orta	0,0	1,0	0,0	0,0	0,0
≤35.313,89	3=Yüksek	0,4	0,4	0,1	0,0	0,0
>35.313,89-≤52.189,74	1=Düşük	0,0	0,6	0,1	0,1	0,0
>35.313,89-≤52.189,74	2=Orta	0,0	0,1	0,3	0,3	0,1
>35.313,89-≤52.189,74	3=Yüksek	0,3	0,0	0,6	0,0	0,0
>52.189,74-≤77.503,51	1=Düşük	0,0	1,0	0,0	0,0	0,0
>52.189,74-≤77.503,51	2=Orta	0,0	1,0	0,0	0,0	0,0
>52.189,74-≤77.503,51	3=Yüksek	0,0	0,0	0,0	0,0	1,0
>77.503,51	2=Orta	0,0	1,0	0,0	0,0	0,0

Tablo 11. Antidepresan Değişkeni için Hesaplanan Koşullu Olasılıklar

Tablo 11'e bakıldığında kişi başı GSYİH'sı 52.189,74\$'dan fazla, 77.503,51\$'e eşit veya az olan ve intihar hızı yüksek olan ülke vatandaşlarının antidepresan kullanımının 106,1'den büyük olma olasılığı %100'dür. Kişi başı GSYİH'sı 52.189,74\$'dan fazla, 77.503,51\$'e eşit veya fazla olan ve intihar hızı düşük olan ülke vatandaşlarının antidepresan kullanım oranının 35,7<antidepresan kullanımı (doz) ≤59,2 arasında olması olasılığı %100'dür.

Tartışma ve Sonuç

Dünya genelinde son yıllarda artışı dikkat çeken intihar vakalarının incelenmesini konu alan bu çalışmada; OECD ülkelerinden elde edilen intihar ve intihar ile ilişkili değişkenler arasındaki ilişki veri madenciliği teknikleri ile incelenmiştir. İnceleme yapılırken veri madenciliği yöntemlerinden Karar Ağaçları ve Bayes Ağları kullanılmıştır. Her algoritmanın en yüksek doğru sınıflandırmayı verdiği değerler üzerinden farklı yorumlar elde edilmiştir.

Karar Ağacı ve Bayes Ağları algoritmaları kullanılarak yürütülen bu çalışmada tüm değişkenlerin intihar hızı ile ilişkili olduğu saptanmıştır. İntihar ile ilişkili değişkenlerin belirlenmesine yönelik yürütülen ve veri madenciliği tekniklerinden farklı tekniklerin kullanıldığı çalışmalarda da benzer sonuçlar elde edilmiş olup; kişi başı GSYİH, işsizlik, alkol tüketimi, yıllık çalışma süresi, boşanma hızı ve antidepresan kullanımı değişkenlerinin intihar hızı değişkeni ile ilişkili ve üzerinde etkili olduğu bulunmuştur (Okada & Samreth, 2013; Kamat vd., 2014; Laanani vd., 2015; Khazaei vd., 2017; Demirci, Konca, Yetim, & İlgün, 2019).

Bu çalışmayı diğer çalışmalardan farklı kılan yön veri madenciliği tekniklerinden olan Karar Ağaçları ve Bayes Ağların kullanılmasıdır. Bu teknikler sayesinde OECD ülkelerinin karar kuralları ve koşullu olasılıkları geliştirilmiş, intihar hızı değişkeni ile ilişkili değişkenler belirlenmiş ve belirlenen değişkenler ile OECD ülkelerinin intihar açısından risk profilleri/özellikleri ortaya çıkarılmıştır. Bu profillerin/özelliklerin belirlenmesi OECD ülkelerinin intihara yönelik alacakları önlemlerde ve geliştirecekleri sağlık ve sosyal politikalarda yardımcı olabilecektir.

Sonuç itibarıyla yürütülen bu çalışmanın, yöntem bakımından ilk çalışmalardan olması sebebiyle ortaya çıkan sonuçların literatüre önemli katkılar sağlayacağı düşünülmektedir. Bu konu üzerinde gelecekte yürütülecek çalışmalarda ülkeler özelinde veri madenciliği yöntemlerinin kullanılması önerilmektedir. Bu sayede ülkelere özgü sorunların tespiti sağlanabilecek ve benzer şekilde ülkelere özgü önlemlerin geliştirilmesine imkan verebilecektir. Son olarak intihar üzerinde yürütülecek çalışmalarda, bu çalışmada yer alan değişkenlerden farklı sosyal, ekonomik ve demografik değişkenlerin de etkisinin değerlendirilmesi önerilmektedir.

KAYNAKÇA

AKMAN, M., GENÇ, Y., & ANKARALI, H. (2011). Random Forests Yöntemi ve Sağlık Alanında Bir Uygulama. *Türkiye Klinikleri Biyoistatistik*, 3(1), 36-48.

AKPINAR, H. (2000). Veri tabanlarında bilgi keşfi ve veri madenciliği. *İÜ İşletme Fakültesi Dergisi*, 29(1), 1-22.

ALTUNKAYNAK, B. (2017). *Veri Madenciliği Yöntemleri ve R Uygulamaları*. Ankara: Seçkin Yayıncılık.

AMINI, P., AHMADINIA, H., POOROLAJAL, J., & AMIRI, M. M. (2016). Evaluating the high risk groups for suicide: a comparison of logistic regression, support vector machine, decision tree and artificial neural network. *Iranian journal of public health*, 45(9), 1179-1187.

ANDRÉS, A. R., HALICIOGLU, F. & YAMAMURA, E. (2011). Socio-economic determinants of suicide in Japan. *The Journal of Socio-Economics*, 40(6), 723-731.

BACA-GARCIA, E., PEREZ-RODRIGUEZ, M. M., SAIZ-GONZALEZ, D., BASURTE-VILLAMOR, I., SAIZ-RUIZ, J., LEIVA-MURILLO, J. M., ... & DE LEON, J. (2007). Variables associated with familial suicide attempts in a sample of suicide attempters. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 31(6), 1312-1316.

BAE, S. M., LEE, S. A., & LEE, S. H. (2015). Prediction by data mining, of suicide attempts in Korean adolescents: a national study. *Neuropsychiatric disease and treatment*, 11, 2367-2375.

BARTH, A., SÖGNER, L., GNAMBS, T., KUNDI, M., REINER, A. & WINKER, R. (2011). Socioeconomic factors and suicide: an analysis of 18 industrialized countries for the years 1983 through 2007. *Journal of Occupational and Environmental Medicine*, 53(3), 313-317.

BAZILA BANU, A., & THIRUMALAIKOLUNDUSUBRAMANIAN, P. (2018). Comparison of Bayes classifiers for breast cancer classification. *Asian Pacific journal of cancer prevention: APJCP*, 19(10), 2917-2920.

BEN-GAL, I. (2007). Bayesian Network. *Encyclopedia of Statistics in Quality & Reliability*. Wiley & Sons.

BERROUIGUET, S., BILLOT, R., LARSEN, M. E., LOPEZ-CASTROMAN, J., JAUSSENT, I., WALTER, M., ... & COURTET, P. (2019). An approach for data mining of electronic health record data for suicide risk management: database analysis for clinical decision support. *JMIR mental health*, 6(5), 1-11.

CHENG, J. & GREINER, R. (2001). *Learning bayesian belief network classifiers: Algorithms and system*. Conference of the Canadian Society for Computational Studies of Intelligence. Berlin: Springer.

ÇALIŞ, A., DURMAZ, K. İ. & GENCER, C. Uçak Seferlerindeki Rötarlari Etkileyen Faktörlerin Analizi. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, 18. EYİ Özel Sayısı, 179-190.

ÇALIŞ, A., KAYAPINAR, S. & ÇETİNYOKUŞ, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları ile Bilgisayar ve İnternet Güvenliği Üzerine Bir Uygulama. *Endüstri Mühendisliği*, 25(3), 2-19

ÇİNİCIOĞLU, E. N., ATALAY, M. & YORULMAZ, H. (2013). Trafik kazaları analizi için bayes ağları modeli. *Bilişim Teknolojileri Dergisi*, 6(2), 41.

DEMIRCI, S., KONCA, M., YETİM, B. & İLGÜN, G. (2019). Effect of economic crisis on suicide cases: An ARDL bounds testing approach. *International Journal of Social Psychiatry*, 66(1), 34-40.

DOGAN, S. & TURKOGLU, I. (2008). Iron-deficiency anemia detection from hematology parameters by using decision trees. *International Journal of Science & Technology*, 3(1), 85-92.

DOGRU, N., & SUBASI, A. (2018). *Traffic accident detection using random forest classifier*. 15th learning and technology conference. IEEE.

DURKHEIM, E. (2005). *Suicide: A study in sociology*. Routledge.

IBM. (2020). *Random Forest*. Erişim Tarihi: 09.12.2021, <https://www.ibm.com/cloud/learn/random-forest>

IBM. (2021). Predictor Importance. Erişim Tarihi: 09.12.2021, https://www.ibm.com/docs/en/spss-modeler/18.1.0?topic=SS3RA7_18.1.0/modeler_mainhelp_client_ddita/clementine/idh_common_predictor_importance.html

ILGEN, M. A., DOWNING, K., ZIVIN, K., HOGGATT, K. J., KIM, H. M., GANOCZY, D., ... & VALENSTEIN, M. (2009). Exploratory data mining analysis identifying subgroups of patients with depression who are at high risk for suicide. *The Journal of clinical psychiatry*, 70(11), 1495-1500.

KAMAT, M. A., EDGAR, L., NIBLOCK, P., MCDOWELL, C. & KELLY, C. B. (2014). Association between antidepressant prescribing and suicide rates in OECD countries: an ecological study. *Pharmacopsychiatry*, 47(1), 18-21.

KHAZAEI, S., ARMANMEHR, V., NEMATOLLAHI, S., REZAEIAN, S. & KHAZAEI, S. (2017). Suicide rate in relation to the Human Development Index and other health related factors: A global ecological study from 91 countries. *Journal of Epidemiology and Global Health*, 7(2), 131-134.

KOYUNCUGİL, A. & ÖZGÜLBAŞ, N. (2009). Veri madenciliği: Tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. *Bilişim Teknolojileri Dergisi*, 2(2).

KUHN, M. & JOHNSON, K. (2013). *Applied predictive modeling*. New York: Springer.

LAANANI, M., GHOSN, W., JOUGLA, E. & REY, G. (2015). Impact of unemployment variations on suicide mortality in Western European countries (2000–2010). *Journal of Epidemiology and Community Health*, 69(2), 103-109.

OKADA, K. & SAMRETH, S. (2013). A study on the socio-economic determinants of suicide: Evidence from 13 European OECD countries. *The Journal of Socio-Economics*, 45, 78-85.

PIOTROWSKI, N. A. & HARTMANN, P. M. (2019). *Suicide*. Magill's Medical Guide (Online Edition).

RATNER, B. (1998). CHAID for interpreting a logistic regression model. *Journal of Targeting Measurement and Analysis For Marketing*, 6, 215-226.

RYGIELSKI, C., WANG, J.-C. & YEN, D. C. (2002). Data Mining Techniques for Customer Relationship Management. *Technology in Society*, 24(4), 483-502.

SAVAŞ, S., TOPALOĞLU, N. & YILMAZ, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.

SMITA, SHARMA, P. (2014). Use of data mining in various field: A survey paper. *IOSR Journal of Computer Engineering*, 16(3), 18-21.

SEYREK, İ. H. & ATA, H. A. (2010). Veri Zarflama Analizi ve Veri Madenciliği ile Mevduat Bankalarında Etkinlik Ölçümü. *Journal of BRSA Banking & Financial Markets*, 4(2).

SHI, G. (2013). *Data mining and knowledge discovery for geoscientists*. Elsevier.

STACK, S. (1990). New micro-level data on the impact of divorce on suicide, 1959-1980: A test of two theories. *Journal of Marriage and the Family*, 119-127.

TURE, M., TOKATLI, F. & KURT, I. (2009). Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4. 5 and ID3) in determining recurrence-free survival of breast cancer patients. *Expert Systems with Applications*, 36(2), 2017-2026.

U.S. MISSION TO THE ORGANIZATION FOR ECONOMIC COOPERATION & DEVELOPMENT. (2019). *What is the OECD?*. Erişim Tarihi: 04.01.2020, <https://usoecd.usmission.gov/our-relationship/about-the-oecd/>

WORLD HEALTH ORGANIZATION. (2018). *Suicide*. Erişim Tarihi: 04.01.2020, <https://www.who.int/news-room/fact-sheets/detail/suicide>

WU, X., KUMAR, V., QUINLAN, J. R., GHOSH, J., YANG, Q., MOTODA, H., ... & STEINBERG, D. (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, 14(1), 1-37.

ZAKI, M. J., MEIRA JR, W. & MEIRA, W. (2014). *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.

Summary

Throughout human history, suicides have occurred at different rates in the societies, countries and groups. Suicide is not only a problem with society and mental health, it also has an economic and cultural aspect. Individuals deliberately end their own lives by attempting suicide. Suicides usually occur as a result of a psychological disorder. However, the person sometimes commits suicide due to life-threatening physical ailments, stress, drug use, economic problems, and family problems. In studies, it was determined that conditions such as the level of development of countries, unemployment, alcohol consumption, living alone/divorce, antidepressant use, gender and psychological diseases were associated with suicidal behavior. In the light of this information, it is aimed to determine the impact of GDP per capita, unemployment, alcohol consumption, annual working time, divorce and antidepressant use on suicide of the Organization for Economic Co-operation and Development (OECD) member states. In this study, classification methods of data mining were used because the effect of independent variables on the dependent variable was estimated. Among the data mining techniques, we used the algorithms giving the highest correct classification rates. According to the C5.0 algorithm, unemployment was the most important independent variable for estimating suicide rate. One of the decision rules created according to the C5.0 algorithm is as follows. If an OECD country has a moderate suicide rate (between 10 > and ≤13.1), its GDP per capita is expected to be more than \$48208.05, the unemployment rate is 9.65% or less, the annual working time is more than 1447.5 hours, and the divorce rate per 1,000 people per year is expected to be 3 and less. The decision rules created with Random Forest are as follows. OECD countries with an antidepressant dose of 65.0 and less per 1,000 people, unemployment rate greater than 5.0%, GDP per capita of \$35054.49 and less, and alcohol consumption of more than 10.4 liters, are expected to have a high suicide rate. The conditional probabilities created with the TAN network structure are as follows. Countries with high suicide rates have a 64% probability that their GDP is less than \$35,313.89. According to the data mining techniques used in the study, it was determined that all variables are effective variables in estimating suicide rate. In this study,

decision rules and conditional probabilities of OECD countries were improved and suicide risk groups were identified with indicators at the national level. Identifying these groups will be able to show which groups OECD countries will target in the measures they will take on suicide and in the health and social policies they will develop.

Ek-1.

IBM Modeler Programı ile Veri Madenciliği Teknikleri Uygulama Akışı

