

BULAŞICI OLMAYAN HASTALIKLAR BAKIMINDAN FARKLI ÜLKELERİN MAKİNA ÖĞRENMESİ YÖNTEMLERİ KULLANILARAK SINIFLANDIRILMASI

*Songül ÇINAROĞLU **
*Keziban AVCI ***

Özet: Bu çalışmanın amacı Dünya Sağlık Örgütü (DSÖ)'ne üye olan 193 ülkeyi gelir gruplarına göre Bulaşıcı Olmayan Hastalıklar (BOH) ile ilgili veriler bakımından sınıflandırmaktır. Ülkelerin sınıflandırılmasında veri madenciliği yöntemleri içerisinde danışmanlı öğrenme yöntemleri arasında bulunan destek vektör makinesi ve random forest yöntemleri kullanılmıştır. Analizlerde bir açık kaynak kodlu yazılım olan Orange programından yararlanılmıştır. Analizler sonucunda random forest yöntemi kullanılarak elde edilen performans sonuçlarının destek vektör makinesine göre daha yüksek olduğu görülmüştür. Araştırma sonuçlarının küresel sağlık yöneticilerine Bulaşıcı Olmayan Hastalıklar (BOH) ile mücadele etmek konusunda ve etkin politikalar üretmede faydalı olacağı düşünülmektedir.

Anahtar Kelimeler: Bulaşıcı Olmayan Hastalıklar (BOH), Sağlık Bakım Göstergeleri, Makina Öğrenmesi

Classification of Different Countries in Terms of Noncommunicable Diseases Using Machine Learning Techniques

Abstract: The aim of this study is to classify 193 countries which are members of World Health Organization (WHO) in terms of Non Communicable Diseases (NCDs). Support vector machine and random forest methods used for classification which are one of supervised data mining methods. An open source programme Orange used for analysis. At the end of the analysis it was seen that random forest classification performance results were better than support vector machine classification performance results. The results of this study is useful for global health care managers for fighting against Noncommunicable Diseases and producing effective policies.

Keywords: Noncommunicable Diseases (NCDs), Health Care Indicators, Machine Learning

1. GİRİŞ

Bireylerin eğitim düzeyinin yükselmesi, ekonomik yapının iyileşmesi, besin ve konut ihtiyacının daha kolay karşılanması, temiz suya ulaşılabilirlik, bulaşıcı hastalıklardaki azalma ve aşılama oranlarındaki artış gibi nedenlerle ortalama yaşam süresinin uzaması beraberinde Bulaşıcı olmayan hastalıkların (BOH) prevalansını ve hastalık yükünü artırmaktadır (World Health Organization [WHO], 2008). Bu durum politik aktörlerin büyük ölçüde ilgisini çeken bir konudur. Tüm toplum genelinde bakıldığında bulaşıcı olmayan hastalıklar önemli bir sorun olmaya devam etmektedir. Uzun yıllar içerisinde bulaşıcı olan hastalıkların tedavisi konusunda edinilen

* Hacettepe Üniversitesi İktisadi ve İdari Bil. Fakültesi Sağlık İdaresi Bölümü, Beytepe Yerleşkesi, 06800 Ankara.

** Türkiye Yüksek İhtisas Eğitim ve Araştırma Hastanesi, Atatürk Bulvarı, Kızılay Sokak No:4, Sıhhiye, Ankara.

İletişim Yazarı: S. Çınaroğlu (cinaroglu @ hacettepe.edu.tr)

tecrübeler sonucunda bu hastalıklar ile etkili bir şekilde mücadele edilmesi gerektiği kabul edilmiştir (Alleyne ve diğ., 2011)

BOH'a yol açan temel risk faktörleri çevreden kaynaklanan risk faktörleri ve bireysel risk faktörleri olarak iki temel grupta incelenebilir. Buna göre çevresel risk faktörlerinin biyolojik çeşitliliğin gittikçe azalması, iklim değişikliği, doğal kaynakların giderek tükenmesi gibi faktörlerden oluştuğu görülür iken, (Bosquet ve diğ., 2011) bireysel risk faktörlerinin sigara kullanımı, tuz kullanımı, fazla yağlı gıdalar tüketmek, yetersiz fiziksel aktivite, zararlı alkol tüketimi gibi tüm dünyada benzer nedenler olduğu belirlenmiştir (Beaglehole ve diğ., 2011). Dünya nüfusunun büyük bir bölümünün fiziksel aktivite açısından oldukça yetersiz durumda olduğu düşünüldüğünde bu hastalıklar büyük bir halk sağlığı sorunu haline gelmektedir. (Lee ve diğ., 2012).

Nitekim Dünya Sağlık Örgütü (DSÖ) BOH'ın küresel ölümlerin %63'ünün (2008 yılı için 57 milyon ölüm) temel sebebi olduğunu belirtmektedir. BOH hemen hemen tüm ülkelerde temel ölüm nedenleri arasında yer almakla birlikte farklı gelir düzeyine sahip ülkelerde prevalans ve mortalite hızları da değişmektedir. BOH'ın genel görünümüne bakıldığında gelir düzeyi yüksek ülkeler için bu hastalıkların başlıca ölüm nedeni olduğu bilinmektedir. Gelir düzeyi düşük olan ülkelerde 60 yaş altı prematüre ölümlerin oranı yüksek gelir düzeyine sahip ülkelere üç kat daha fazladır. Öte yandan bu hastalıklar düşük ve orta gelir grubunda yer alan ülkeler açısından daha ağır bir yük oluşturabilmektedir (WHO, 2011).Orta ve düşük gelir grubunda yer alan ülkelerde özellikle genç nüfus arasında bu hastalıkların görülme sıklığında görülen artış bu hastalıkların tedavisi konusunda politika geliştirmeyi gerektirmektedir (WB, 2011).

BOH ile mücadele konusunda tüm dünya genelinde atılan küresel adımlar incelendiğinde DSÖ başta olmak üzere pek çok uluslararası kuruluşun bu hastalıklar ile mücadele amaçlı adımlar attıkları görülmektedir. Tablo 1'de bulaşıcı olmayan hastalıklar ile mücadele konusunda atılan küresel adımlar kronolojik olarak sıralanmıştır.

Tablo 1. Bulaşıcı Olmayan Hastalıklar İle Mücadele Amacıyla Atılan Küresel Adımlar (Bousquet ve diğ. (2011))

Mart 2008	61. Dünya Sağlık Asamblesinde Dünya Sağlık Örgütü 2008 ile 2013 yıllarını kapsayan bir süreliğine tüm dünya genelinde bulaşıcı olmayan hastalıkların önlenmesi ve kontrolü için bir öncelik politikası belirlemiştir. Bu öncelik alanları içerisinde mücadele stratejisi geliştirilen başlıca hastalıklar arasında; kalp ve damar hastalıkları, kanser, solunum yolu hastalıkları ve diyabet yer almaktadır.
Mayıs 2010	Birleşmiş Milletler Genel Asamblesi toplantısında 21. yüzyılda sürdürülebilir gelişme sağlayabilmek için bulaşıcı olmayan hastalıklar ile mücadelenin en temel öncelik olduğu kabul edilmiştir.
Aralık 2010	Avrupa Birliği Konseyi bulaşıcı olmayan hastalıklar ile mücadele amacıyla halk sağlığı ve sağlık sistemlerinin iyileştirilmesi amacıyla nüfusa dayalı ve hasta odaklı yenilikçi politikalar geliştirmiştir.
2010	Amerika Birleşik Devletleri Hastalıklar ile Mücadele ve Önleme Merkezi (US Center for Disease Control and Prevention) yaşlı yetişkinlerin bulaşıcı olmayan hastalıklardan ve bunların yol açacağı komplikasyonlardan korunmasının yaşamsal öneme sahip olduğunu belirtmiştir
19 Eylül 2011	Birleşmiş Milletler bulaşıcı olmayan hastalıklar ile ilgili genel bir sempozyum düzenlenmiştir.

2. YÖNTEM

2.1. Araştırmanın Amacı

Bu çalışmanın amacı DSÖ'ye üye olan 193 ülkeyi DSÖ tarafından 2011 yılında yayınlanan BOH ülke profili verileri açısından veri madenciliği yöntemlerini kullanarak gelir gruplarına göre sınıflandırmaktır.

2.2. Çalışma Evreni ve Ele Alınan Göstergeler

Araştırma evrenini 2011 yılı itibarıyla Dünya Sağlık Örgütü'ne üye olan toplam 193 ülke oluşturmaktadır. Veriler Dünya Sağlık Örgütü tarafından 2011 yılında yayınlanan 'Noncommunicable Diseases Country Profiles' isimli yayından alınmıştır. Dünya Sağlık Örgütü'nün bu yayınından üye ülkelerde BOH'a bağlı mortalite hızı, davranışsal risk faktörleri, metabolik risk faktörleri ve ülkelerin bu hastalıklar ile mücadele amacıyla uyguladıkları politikalar gibi bilgilere ulaşılabilmektedir. Bu çalışmada söz konusu göstergeler bulaşıcı olmayan hastalıklara bağlı mortalite hızı ve BOH ile mücadele amaçlı uygulanan politikalar olmak üzere iki farklı grupta ele alınmış ve bu gruplarda yer alan değişkenlere ilişkin bilgiler aşağıdaki tabloda (Bknz. Tablo 2) sunulmuştur.

Tablo 2. Çalışma Kapsamına Alınan Göstergeler (WHO (World Health Organization), (2011))

Ülke Nüfusu	Genel Göstergeler
Gayrisafi Yurtiçi Harcamalar İçerisinde Sağlık Harcamalarının Payı	
Her 1000 kişide bulaşıcı olmayan hastalıklara bağlı toplam ölüm hızı	
Tüm bulaşıcı olmayan hastalıklardan ölümler içerisinde 60 yaş altı ölüm	
Her 100.000 kişide yaşa göre ayarlanmış tüm bulaşıcı olmayan hastalıklardan ölüm hızı	Bulaşıcı Olmayan Hastalıklara Bağlı Mortalite Hızı
Her 100.000 kişide yaşa göre ayarlanmış kanser ölüm hızı	
Her 100.000 kişide yaşa göre ayarlanmış kronik solunum yolu hastalıklarına bağlı ölüm hızı	
Her 100.000 kişide yaşa göre ayarlanmış kardiyovasküler hastalıklar ve diyabete bağlı ölüm hızı	
Ülke genelinde Sağlık Bakanlığı bünyesinde BOH ile mücadele amaçlı bir birim olup olmadığı	
Ülke genelinde BOH tedavisi ve kontrolü için yeterli fon kaynaklarının ayırıp ayrılmadığı	
Ülke genelinde BOH'ın önlenmesi ve sağlığın geliştirilmesi için yeterli fon kaynaklarının ayrılıp ayrılmadığı	BOH ile Mücadele Amaçlı Uygulanan Politikalar
Ülke genelinde BOH'ın denetimi ve değerlendirilmesi	
Ulusal Sağlık Sonuçlarının Raporlanmasında BOH'a bağlı ölüm oranlarının sunulup sunulmadığı	
Ulusal Sağlık Sonuçlarının Raporlanmasında BOH'a bağlı morbidite oranlarının sunulup sunulmadığı	
Ulusal Sağlık Sonuçlarının Raporlanmasında BOH'a bağlı risk faktörlerinin sunulup sunulmadığı	

Tablo 2. (devamı) Çalışma Kapsamına Alınan Göstergeler (WHO (World Health Organization), (2011))

Ulusal Sağlık Sonuçlarının Raporlanmasında kanser ile ilgili kayıtların tutulup tutulmadığı	BOH ile Mücadele Amaçlı Uygulanan Politikalar (devamı)
Ülke genelinde kardiyovasküler hastalıklar ile mücadele konusunda bir program / eylem planı izlenip izlenmeme durumu	
Ülke genelinde kanser ile mücadele konusunda bir program / eylem planı izlenip izlenmeme durumu	
Ülke genelinde kronik solunum yolu hastalıkları ile mücadele konusunda bir program / eylem planı / bir politika izlenip izlenmeme durumu	
Ülke genelinde diyabet hastalıkları ile mücadele konusunda bir program / eylem planı / bir politika izlenip izlenmeme durumu	
Ülke genelinde alkolden kaynaklanan hastalıkları ile mücadele konusunda bir program / eylem planı / bir politika izlenip izlenmeme	
Ülke genelinde düzensiz beslenme/aşırı kilodan kaynaklanan hastalıkları ile mücadele konusunda bir program / eylem planı / bir politika izlenip	
Ülke genelinde fiziksel hareketsizlikten kaynaklanan hastalıkları ile mücadele konusunda bir program / eylem planı / bir politika izlenip	
Ülke genelinde sigara içmeden kaynaklanan hastalıkları ile mücadele konusunda bir program / eylem planı / bir politika izlenip izlenmeme	
Sigara ile mücadelede DSÖ tarafından geliştirilen MPOWER mücadele adımlarının yerine getirilmesi durumu	
Gelir Grubu	Hedef Kategorisi

2.2. Verilerin Analizi

193 ülkeye ait Tablo 2'de gösterilen veriler her bir ülkede erkek ve kadın nüfusa özel olarak değerlendirilmek amacı ile bilgisayar ortamına aktarılmıştır. Hedef kategori olarak gelir grubu belirlenmiş ve gelir grubu açısından ülkeler düşük ve yüksek olmak üzere iki farklı gruba ayrılmıştır. Veriler Orange veri madenciliği programı kullanılarak analiz edilmiştir. Orange ücretsiz şekilde ulaşılabilen açık kaynak kodlu yazılımlardan birisidir.

Ülkelerin BOH açısından gelir grubuna göre sınıflandırılmasında veri madenciliği yöntemleri arasında bulunan Destek Vektör Makinası ve Random Forest yöntemleri kullanılmış ve her iki yöntemin performans sonuçlarının karşılaştırılmasına yer verilmiştir. Her iki yöntemin uygulanmasında da 5 katlı çapraz geçerlilik yöntemi kullanıldığında performans sonuçlarının yeterli olduğu görülmüş ve bu nedenle 5 katlı çapraz geçerlilik tercih edilmiştir.

Çapraz geçerlilik yöntemi danışmanlı öğrenme analizlerinde sıklıkla kullanılan bir yöntemdir. Genel olarak '10 parça çapraz geçerlilik' yöntemi tercih edilmektedir. Bu yöntemde veri kümesi rastgele 10 eşit parçaya ayrılır. İlk aşamada birinci parça test veri seti olarak bırakılır, geriye kalan dokuz parça eğitim seti olarak kullanılır. İkinci aşamada ise ikinci parça test veri seti olarak kullanılır, geriye kalan dokuz parça ile model bulunmaya çalışılır. Bu süreç 10 parça ayrı ayrı test seti olarak kullanılıncaya kadar devam eder. Çapraz geçerlilik sonunda her parçadan elde edilen doğru sınıflama ya da ROC eğrisi altındaki alan gibi performans değerlerinin ortalaması alınır. Bu sonuç ilgili yöntemin genel performansını ifade eder (Coşgun ve Karaağaoğlu, 2011).

Random forest (RF) binlerce karar ağacından meydana gelen bir yapıdır. Random forest yönteminde CART (Classification and Regression Tree) algoritması ile ağaçlar oluşturulur. CART algoritması kullanıldığında veri setinin hangi değişkenden başlayarak dallara ayrılacağına

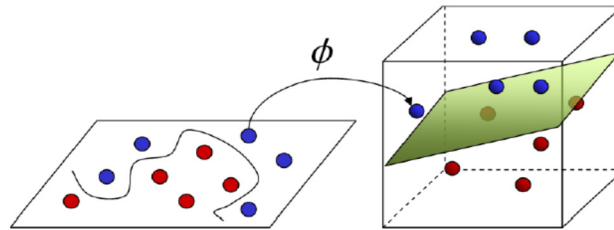
‘bilgi kazancı’ kullanarak karar verilir ancak dallara ayırmak için seçilen değişkenin uygun test kriteri (cut-off değeri) “Gini katsayısı” ile tespit edilir (Akman ve diğ., 2011). Random forest yönteminde her bir ağaç için bootstrap yöntemi ile veri setinden örneklem seçilir ve seçilen verilerin 2/3’ü ağaç oluşturmak için kullanarak bir sınıflama yapılır. Bu sınıflama bir ‘oy’ (vote) alır. Random forest algoritması ‘forest’ içindeki tüm ağaçlardan en çok oy alanı seçer ve onun sınıflamasını kullanır. Düşük hata oranına sahip ağaç daha iyi sınıflayıcıdır. Random forest yönteminde hata oranı iki şeye bağlıdır. Bunlardan birisi; iki ağaç arasındaki korelasyondur, korelasyon arttıkça hata oranı artar. İkincisi; her ağacın kendi hata oranıdır. Random forest yönteminin avantajları arasında; en iyi performansa ulaşabilmek için istenildiği kadar ağaç üretilebilmesi yer alır. Ayrıca hızlı bir algoritmadır. Elde edilen random forest diğer veri setlerinde kullanılmak üzere saklanabilir. Bu yöntem ile doğru sınıflama oranı eksik veriler olması durumunda da devam etmektedir. Random forest yönteminde bilgi ölçütü olarak kullanılan Gini indeksi eşitlik 1.de görüldüğü gibi hesaplanmaktadır (Coşgun ve Karaağaoğlu, 2011):

$$\text{Gini İndeksi} = 1 - \sum_{j=1}^n P_j^2 \quad (1)$$

Random forest yöntemi kategorik, sürekli veya her ikisinin de bulunduğu tüm veri setlerinde uygulanabilmektedir. Hem büyük hem de küçük boyutlu veri setlerinde bu yöntem doğru sonuçlar vermektedir. Orijinal veri setini eğitim ve test veri seti olarak ayırt etmeden de model test edilebilir (Akman ve diğ., 2011).

Random forest yönteminin kısıtlılıkları incelendiğinde; bu yöntemde tek bir karar ağacında olduğu gibi ortaya çıkan sonuç, ağaç yapısıyla görsel olarak görülemez. Model karmaşık olduğu için birçok karar ağacının değerlendirme sonucu, işlem adımları görülemeyecek (black box) şekilde verilir. Lojistik regresyon ve yapay sinir ağları yöntemlerindeki gibi oluşan sonuç için bir güven aralığı verilememektedir. Ayrıca random forest yöntemi için geliştirilen programlar, oluşan her ağacı sistem belleğinde tuttuğu için, çok fazla bellek kullanılmakta ve bu nedenle düşük bellekli bilgisayarlarda kullanılması zorlaşmaktadır (Akman ve diğ., 2011).

Destek vektör makinası yönteminde ise doğrusal olarak sınıflanabilen verileri birbirinden ayırt etmek için olası pek çok doğrusal fonksiyon içerisinde marjini en büyük olan belirlenir. Bu noktada sınıflandırılacak örnekler doğrusal bir düzlemlerle ayrıştırılabilecek seviyede olabileceği gibi Şekil 1’de görüldüğü gibi doğrusal bir çizgiyle ayrıştırılamayacak özellikte de olabilmektedir. Destek vektör makinası yöntemi herhangi bir şekilde doğrusal olarak sınıflanamayan verileri bir Kernel fonksiyonu (ϕ) yardımı ile daha yüksek boyutlu bir uzaya aktarır ve marjini en büyük olan hiper düzlemleri bulur. Veriler bu ayırt edici hiper düzleme göre sınıflara atanır [Bknz. Şekil 1]. Buna ek olarak veri setine yeni giren verilerin hangi sınıfta olduğu da kestirilebilir. Verileri çok boyutlu uzayda kategorilere ayırmak için farklı çekirdek fonksiyonları kullanılabilir. En sık kullanılan çekirdek fonksiyonları; doğrusal, çok terimli, radyal temel fonksiyonu (RTF) ve sigmoiddir (Coşgun ve Karaağaoğlu, 2011).



Doğrusal Olmayan Sınıflandırma Çekirdek Fonksiyonların 2 Boyutlu Uzaya Aktarımı

Şekil 1:

Destek Vektör Makinası Kullanılarak Doğrusal Olarak Sınıflanamayan Verilerin İki Boyutlu Uzaya Aktarımı (Boussemart (2011))

Random forest ve destek vektör makinası yöntemlerinin karşılaştırılması söz konusu olduğunda bu yöntemler arasında genellikle sınıflama doğruluğu bakımından karşılaştırma yapılmaktadır. Yapılan karşılaştırmalarda çoğu zaman destek vektör makinasının random forest yöntemine göre daha iyi bir performans sergilediği görülmektedir (Touw ve diğ., 2012).

3. BULGULAR

3.1. Tanımlayıcı Veriler

Gelir gruplarının sınıflandırılması ile ilgili elde edilen tanımlayıcı sonuçlar genel olarak incelendiğinde düşük gelir grubunda 97 (%50.3), yüksek gelir grubunda 96 (%49.7) ülkenin bulunduğu görülmektedir. Bu bilgilere göre hedef kategori olarak belirlenen gelir grubu değişkeni bakımından dengeli bir dağılımın olduğu söylenebilir (Bknz.. Tablo 3).

Tablo 3. Ülkelerin Farklı Gelir Gruplarına Göre Dağılımı

Gelir Grubu	Sayı	Yüzde
1 (Düşük Gelir Grubu)	97	%50.3
2 (Yüksek Gelir Grubu)	96	%49.7

3.2. Destek Vektör Makinesi (SVM-C) ve Random Forest Yöntemleri Kullanılarak Ülkelerin Gelir Gruplarına Göre Sınıflandırılması

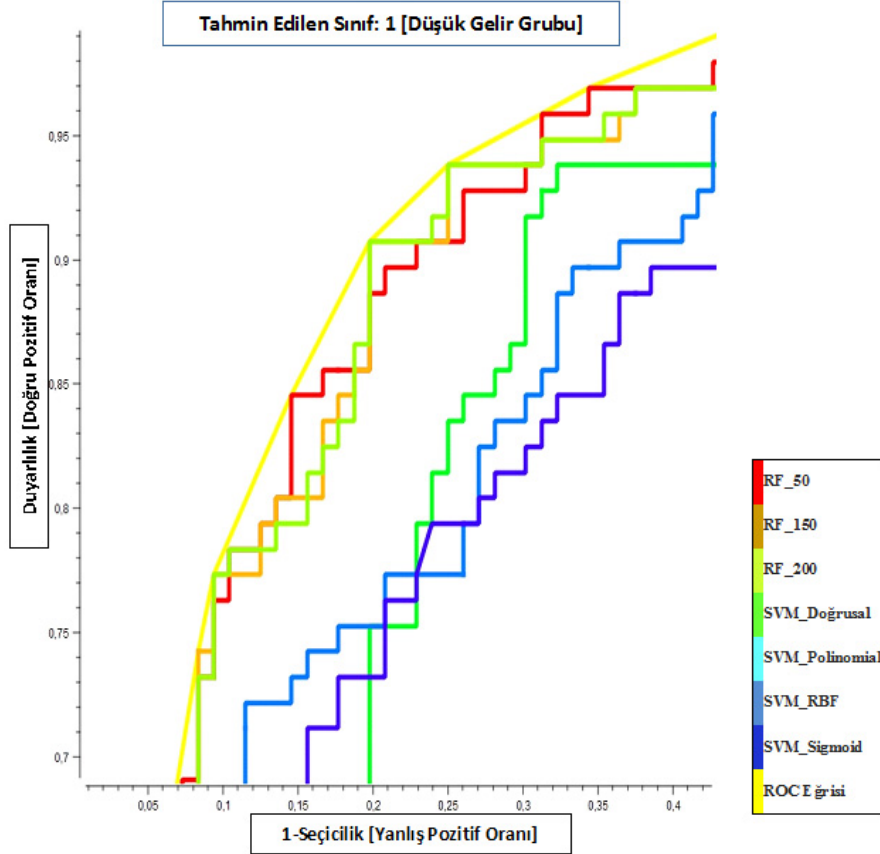
Tablo 4'de destek vektör makinesi ve random forest yöntemleri kullanılarak elde edilen performans sonuçlarına yer verilmiştir. Bu tabloda hedef kategori olarak düşük gelir grubu belirlendiğinde farklı sınıflama yöntemlerinin performans sonuçları görülmektedir. Destek vektör makinesi sınıflandırma yönteminde dört farklı çekirdek fonksiyon (doğrusal, çok terimli, radyal temel fonksiyonu (RBF) ve sigmoid) kullanıldığında elde edilen performans sonuçları ile 50, 150 ve 200 ağaç türetilerek oluşturulan random forest performans sonuçları karşılaştırıldığında 200 ağaç türetilerek oluşturulan Random Forest sonuçlarının ROC Eğrisi altında kalan alan (0.9276) bakımından en yüksek performansı sergilediği tespit edilmiştir.

Tablo 4. Hedef Kategori Olarak Düşük Gelir Grubu Belirlendiğinde Destek Vektör Makinesi (SVM-C) ve Random Forest Performans Sonuçlarının Karşılaştırılması

	Yöntem	Sınıflama Doğruluğu (CA)	Duyarlılık (Sens.)	Seçicilik (Spec.)	AUC
SVM (C)	SVM (C)_Linear	0.7668	0.7629	0.7708	0.8479
	SVM(C)_Polynomial	0.3993	0.1959	0.6042	0.1992
	SVM (C)_RBF	0.7718	0.7732	0.7708	0.8822
	SVM (C)_Sigmoid	0.7718	0.7835	0.7604	0.8685
RF	Random_Forest_50	0.8347	0.8247	0.8438	0.9260
	Random_Forest_150	0.8397	0.8351	0.8438	0.9248
	Random_Forest_200	0.8293	0.8144	0.8438	0.9276

Şekil 2'de destek vektör makinası ve Random Forest yöntemleri kullanılarak elde edilen ROC Eğrisi altında kalan alanlar gösterilmektedir. ROC Eğrisi altında kalan alan (AUC: Area Under ROC Curve) veri madenciliğine özel kullanılan alternatif performans karşılaştırma kriterlerinden birisidir. ROC eğrisi doğru pozitif oranına (duyarlılık) karşı yanlış pozitif oranların

(1-seçicilik) noktalanarak çizilmesiyle elde edilir. Bu eğrinin altında kalan alanın hesaplanması ile AUC elde edilir. AUC veri madenciliği literatüründe çok sık kullanılan sıralamaya dayalı (ranking) bir performans kriteridir. AUC, veri madenciliği sınıflayıcı modelinin hasta ve sağlıklı kişilerden rastgele seçilmiş iki kişiyi doğru tanımlayabilme olasılığı olarak ifade edilebilir. 0 ile 1 arasında değerler alabilir; 0.5 değeri rastgele bir tahmin olduğunu, 1'e yakın değerler modelin tahmin gücünün yüksek olduğunu gösterir. AUC değeri diğer kriterlerden farklı olarak göreceli sınıf dağılımından etkilenmez (Coşgun ve Karağaoğlu, 2011).



Şekil 2:

Hedef Kategori Olarak Düşük Gelir Grubu Belirlendiğinde Destek Vektör Makinesi (SVM-C) ve Random Forest Yöntemleri Kullanıldığında ROC Eğrisi Altında Kalan Alanlar

Hedef kategori olarak düşük gelir grubu belirlendiğinde ROC eğrisi altında kalan alan bakımından analiz sonuçları Şekil 2'de görülebilmektedir. Şekil 2 incelendiğinde dört farklı fonksiyona göre oluşturulan destek vektör makinası sonuçları ile 50, 150 ve 200 ağaç türetilerek elde edilen Random Forest performans sonuçları karşılaştırıldığında Random Forest performans sonuçlarının ROC eğrisine daha yakın oldukları görülmektedir. 200 ağaç türetilerek oluşturulan Random Forest sonuçlarının ise ROC eğrisine en yakın olduğu ve en yüksek performans sergilediği görülmektedir. Analizler sonucunda hedef kategori olarak düşük ya da yüksek gelir grubu seçildiğinde destek vektör makinası performans sonuçlarından sınıflama doğruluğu ve ROC Eğrisi altında kalan alan için sonuçların herhangi bir değişiklik göstermediği, duyarlılık ve seçiciliğin ise yer değiştirdiği görülmektedir.

4. SONUÇ

BOH ile mücadele tüm dünya genelinde pek çok ülkede sağlık ile ilgili kaynakların büyük kısmının harcanmasına neden olmaktadır. Bu nedenle BOH için 21. yüzyılın pandemisi olarak denilebilmektedir (Alberti, 2001). Çalışma sonuçları ülkelerin hedef kategori olarak yüksek ya da düşük gelir grubu değişkeni belirlendiğinde BOH ile ilgili veriler bakımından sınıflandırılmasında Random Forest yönteminin destek vektör makinasına göre daha yüksek performans sergilediğini göstermiştir.

Oluşturulan model sayesinde modele yeni giren herhangi bir ülkenin BOH ile ilgili verileri modele dâhil edildiğinde bu ülkenin hangi gelir grubuna girdiğinin tahmin edilmesi mümkün olabilmektedir. Bu durum farklı ülkelerde gelir düzeyi ile BOH ile mücadele konusunda sağlık yöneticileri için faydalı bilgiler sağlamaktadır. Ayrıca Random Forest yönteminde ağaç sayısı en fazla olduğunda en yüksek performans sonuçlarının elde edildiği de görülmektedir. Bu sonuçlar Random Forest yönteminde ağaç sayısının artırılmasının performans sonuçlarına nasıl yansıdığı konusunun sorgulanması gerekliliğine vurgu yapmaktadır. Bu açıdan bir değerlendirme yapıldığında bu konuda yapılacak bundan sonraki araştırmalarda, bu veri seti için ya da farklı veri setleri kullanıldığında Random Forest yönteminde ağaç sayısını artırmanın performans sonuçlarına nasıl yansıdığı incelenebilecektir.

Bu sonuçlardan yola çıkılarak bundan sonra yapılacak araştırmalarda modelin performansının incelenmesinde ele alınabilecek başka bir konu da kullanılan k-kat (k-fold) çapraz geçerlilik sayısının artırılması yolu ile performans sonuçlarının ele alınmasıdır. Ayrıca farklı makine öğrenmesi yöntemleri kullanılarak performans sonuçları karşılaştırılabilecektir. Farklı ülkelerin BOH ile ilgili verileri farklı sağlık göstergeleri de dikkate alınarak incelenebilecek ve farklı hedef kategoriler belirlenerek farklı sınıflamalar yapılabilecektir. Bu araştırmada elde edilen bilgilerin günümüzde başta gelişmiş ülkeler olmak üzere pek çok ülkede gün geçtikçe daha ciddi bir sorun haline gelen BOH ile mücadele konusunda katkı sağlayacağı düşünülmektedir.

KAYNAKLAR

1. Akman M., Genç Y. Ankaralı H. (2011). Random Forest Yöntemi ve Sağlık Alanında Bir Uygulama, *Türkiye Klinikleri J Biostat*, 3(1), s.36-48.
2. Alberti G. (2001). Noncommunicable Diseases: Tomorrow's Pandemic, *Bulletin of World Health Organization*, Editorial, 79(10), s.907.
3. Alleyne G., Basu S., Stuckler D. (2011). Who's Afraid of Noncommunicable Diseases? Raising Awareness of The Effects of Noncommunicable Diseases on Global Health, *Journal of Health Communication: International Perspectives*, 16, s.82-93.
4. Beaglehole R., Bonita R., Horton R., Adams C., Alleyne H., Asaria P., Baugh V. (2011). Priority Actions for The Non-Communicable Disease Crisis", *Health Policy*, vol.377, s.1438-1447.
5. Bousquet J., Anto J.M., Sterk P.J., Adcock I.M., Chung K.F. (2011). 'Systems Medicine and Integrated Care To Combat Chronic Noncommunicable Diseases', *Genome Medicine*, 3(43), s.1-12.
6. Boussemart Y. (2011). "Predictive Models Of Procedural Human Supervisory Control Behavior", *Doctoral Thesis*, Massachusetts Institute of Technology [MIT], Engineering Systems Division, s.33.
7. Cosgun E., Karaağaoğlu E. (2011). Veri Madenciliği Yöntemleriyle Mikrodizilim Gen İfade Analizi, *Hacettepe Tıp Dergisi*, 42, s.180-189.

8. Lee I.M., Eric J Shiroma E.J., Lobelo F., Puska P., Blair, Peter S.N, Katzmarzyk T., (2012). Effects of Physical Inactivity on Major Non-Communicable Diseases Worldwide: An Analysis of Burden of Disease and Life Expectancy, *Lancet*, 380, s.219-229.
9. Touw W.G., Bayjanov J.R., Overmars L., Backus L., Boekhorst J., Wels M., Van Hijum S.A.F.T. (2012). *Briefings In Bioinformatics*, July 10, 1-12.
10. WB (The World Bank) (2011). *The Growing Danger of Non-Communicable Diseases Acting Now to Reverse Course*, September 2011, Human Development Network.
11. WHO (World Health Organization) (2011). *Noncommunicable Diseases 2011 Country Profiles*. http://whqlibdoc.who.int/publications/2011/9789241502283_eng.pdf, Erişim Tarihi: 03.04.2013.
12. WHO (World Health Organization) (2008). *The World Health Report 2008, Primary Health Care- Now More Than Ever*. WHO, Geneva 2008; p. xiii. http://www.who.int/whr/2008/whr08_en.pdf, Erişim Tarihi: 03.07.2013.

Alınma Tarihi (Received) : 02.01.2015
Düzeltilme Tarihi (Revised) : 26.05.2015
Kabul Tarihi (Accepted) : 26.05.2015