



Tahmin performansını arttırmak için kolektif öğrenme yaklaşımı: Engelli bireylere yönelik deneysel çalışma

Ensemble learning approach for enhancing performance prediction: Experimental study for disabled people

Melda Kokoç^{1,*} , Fatih Kokoç² 

¹ Gazi Üniversitesi, Eğitim-Öğretim ve Dış İlişkiler Kurum Koordinatörlüğü, 06560, Ankara, Türkiye

² Gazi Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, 06680, Ankara, Türkiye

Özet

Fazla değişken söz konusu olduğunda elle sınıflama yapmak zaman ve emek isteyen bir süreç haline gelmektedir. Böyle bir duruma örnek olan engelli bireylerin öz bakım aktivitelerinde yaşadıkları sorunlara göre sınıflandırılması, uzman terapistler için zaman alıcı bir süreçtir. Bu çalışmanın amacı uzman terapistlere zaman kazandırması açısından fiziksel ve motor engelli bireylerin öz bakım aktivitelerinde yaşadıkları sorunları tahmin edebilecek bir modelin geliştirilmesidir. Tahmin sürecinde yedi farklı (destek vektör makineleri, yapay sinir ağları, C5.0, CART, QUEST, CHAID, bayes ağları) veri madenciliği algoritmasından yararlanılmıştır. Söz konusu algoritmalar, tek olarak ve farklı kolektif modeller oluşturularak uygulanmıştır. Tek ve kolektif olarak uygulanan modellerin deneysel sonuçları kıyaslandığında, iki veya daha fazla algoritmanın birleştirilmesi ile uygulanan kolektif öğrenme tekniğinin tahmin performansını yükselttiği görülmüştür.

Anahtar kelimeler: Fiziksel ve motor engel, Medikal veri, Öz bakım, ICF-CY, Veri madenciliği

1 Giriş

Engel, kişinin günlük faaliyetlerini gerçekleştirme kabiliyetini kısıtlayan çeşitli fiziksel ve zihinsel bozuklukları içermektedir. Dünya Sağlık Örgütü'ne göre engel, faaliyet ve katılım kısıtlamalarını kapsayan bir şemsiye terimdir. Faaliyet kısıtlaması, bir bireyin bir görevi veya eylemi yerine getirmesinde karşılaştığı bir zorluk; katılım kısıtlaması ise bireyin yaşam koşullarına katılımı ile ilgili yaşadığı bir problemdir [1]. Dünya Sağlık Örgütü'nün 2011 Dünya Engellilik Raporunda 18 yaş üzerindeki nüfusun %15,6'sının engelli olduğu, Türkiye İstatistik Kurumu'nun 2011 Nüfus ve Konut Araştırması sonuçlarında ise 3 yaş ve üzeri nüfusun %6,9'sının engelli olduğu belirtilmiştir [2, 3]. Oranlar, nüfusun dikkate değer bir kısmının engele sahip olduğunu göstermektedir.

Engellilik, farklı bakış açıları ile literatürde yer edinmiş bir konudur. Engelliliğin toplumsal yaşamdaki etkilerinin analiz edildiği çalışmalarda, genellikle motor ve sosyal iletişim becerilerin, çocukların günlük aktiviteler katılımları ile düşük fonksiyonlu otizm spektrum bozukluğu olan çocuklar üzerindeki etkisi incelenmiştir. Ayrıca otizm bozuklukları ile duygusal rahatsızlığın bir arada olduğu özgül öğrenme bozukluğuna sahip bireyler üzerinde analizler yapıldığı da görülmüştür [4, 5]. Bir başka bakış açısı engelliliği sınıflandırmak ve ölçmek için kullanılan

Abstract

In the case of too many variables, manual classification becomes a time- and labor-intensive process. For instance, classifying people with disabilities according to the problems they experience in self-care activities is a time-consuming process for specialist therapists. The aim of this study is to develop a model that can predict the problems experienced by physically and motor disabled individuals in self-care activities in order to reduce time spent by specialist therapist. Seven different data mining algorithms (support vector machines, artificial neural networks, C5.0, CART, QUEST, CHAID, bayesian networks) have been used in the estimation process. These algorithms have been applied individually and by forming different ensemble models. When experimental results of single and ensemble models were compared, it was seen that ensemble learning technique combined with two or more algorithms increased predictive performance.

Keywords: Data mining, Physical and motor disability, Medical data, Self-care, ICF-CY

İşlevsellik, Yetiyitimi ve Sağlığın Uluslararası Sınıflandırması (International Classification of Functioning, Disability and Health -ICF) çerçevesinin değerlendirilmesi üzerine odaklanmıştır [6, 7]. Bunlara ek olarak bu çalışmanın konusu olan engelli bireylere ait verilerden yararlanılarak bireylerin sınıflandırılması veya tanı teşhisi için farklı örüntü ve bilgilerin keşfedilmesine olanak sağlayan çeşitli veri madenciliği tekniklerini kullanan çalışmalar da literatürde yer almaktadır [8-10].

Toplumsal refahın artırılması hususunda engelli bireylerin yaşam koşullarının iyileştirilmesi önem arz etmektedir. Engelli bireyler için uygulanan tedavi ve rehabilitasyonlar engele neden olan hastalığı ya da genetik durumu düzeltmese de bu durumun neden olduğu semptomları ve yeti yitimini azaltabilmektedir [11]. Bu sayede söz konusu bireylerin topluma kazandırılabilmesi mümkün olmaktadır. Bu nedenle, bu bireylerin engellilik düzeyleri ve sınıfları belirlenerek gereken eğitimi almaları, sosyal ve eğitim haklarından yararlanabilmeleri için yönlendirilmeleri gerekmektedir [12]. Engel teşhisi ve sınıflandırması, uzman terapistler tarafından gerçekleştirilen karmaşık bir süreçtir. Bilimsel ve politik bağlamlarda, engellilik tanıları, genellikle sağlık koşullarının, tanıların veya etiyolojik faktörlerin örtüşmesiyle karakterize edilmektedir [7]. Engel sınıflandırması için çeşitli çerçeveler bulunmakla birlikte ICF uluslararası çerçeve olarak kabul

* Sorumlu yazar / Corresponding author, e-posta / e-mail: meldakokoc@gazi.edu.tr (M. Kokoç)

Geliş / Received: 26.02.2020 Kabul / Accepted: 07.09.2020 Yayınlanma / Published: 15.01.2021

doi: 10.28948/ngumuh.693303

edilmektedir [2]. ICF'nin, alt fonksiyonu olan ICF-CY, çocuk ve gençlerin zihinsel ve fiziksel çeşitli fonksiyonel sorunlarını belirlemek için ayrıntılı ve kapsamlı bir sınıflandırma çerçevesi sunmaktadır. ICF-CY, dört seviyeyi göz önünde bulundurarak her bir sağlık bileşenini kodlamaktadır. Seviye-1, vücut fonksiyonları, vücut yapıları, faaliyetler ve katılım ile çevresel faktörler olarak adlandırılmış dört kategoriden oluşan genel faaliyetleri içermektedir. Seviye-1 kodları, rakamların ve harflerin bir kombinasyonu olarak tanımlanmaktadır. Bu çalışmanın konusu olan öz bakım, ICF-CY'de, seviye 1'deki faaliyetler ve katılım bileşeninin bir alt bölümüdür ve "d5" bileşen kodu ile başlamaktadır. Sonraki seviyeler; seviye 2 (10 ile 99 arasında bir değere sahip), seviye 3 (0 ile 9 arasında bir değere sahip) ve seviye 4 (0 ile 9 arasında bir değere sahip) bir önceki seviyeyi daha ayrıntılı olarak açıklamak için kullanılmaktadır [13].

Öz bakım sorunlarının sınıflandırılması, uygun tedavi yaklaşımlarının seçilmesinde önemli bir faktördür. Öz bakım problemlerinin çeşitliliği, karmaşıklığı ve mesleki terapistlerin eksikliği, mevcut terapistlere ICF-CY kriterlerine uygun sınıflandırmada destek sağlayacak bir modelin geliştirilmesini gerektirmiştir. Bu çalışmanın amacı, fiziksel ve motor engelli bireylerin öz bakım sorunlarının yüksek doğruluk ile tahmin edilmesi için kolektif öğrenme yöntemine dayanan sınıflandırıcı geliştirilmesidir. Kolektif yöntemleri kullanma eğiliminin artması, kolektif yaklaşımların bireysel veri madenciliği tekniklerinin performansını geliştirdiğini göstermektedir [14]. Kolektif yöntemler günümüzde yaygın olarak kullanılmasına rağmen, bu tarz bir çalışmaya öz bakım aktivitelerinin ICF-CY'ye göre sınıflandırılmasında rastlanmamıştır.

Bu çalışmanın ikinci bölümünde kullanılan veri madenciliği algoritmaları kısaca açıklanmış, algoritmaların performansının ölçülmesi için kullanılan veri seti tanıtılarak analiz edilmiştir. Tanıtılan veri setine ilişkin çeşitli veri madenciliği algoritmaları bireysel ve kolektif olarak uygulanarak sonuçlar değerlendirilmiştir. Çalışmanın sonuç bölümünde ise yapılan uygulama özetlenerek gelecek çalışmalarda yapılması hedeflenen uygulamalar sunulmuştur.

2 Materyal ve metot

Verinin bilgiye dönüştürülmesi ve büyük miktarda heterojen verinin yönetimi, veri madenciliği gibi önemli bir araştırma alanının gelişmesine neden olmuştur [15]. Veri madenciliği, tahmin edilemeyen ilişkiyi ve verilerde saklı kalıpları keşfetmek için, çok sayıda gözlemsel veri setinden bilginin çıkarılması ve kullanıcılar için anlaşılabilir, kullanışlı hale getirilmesi için verilerin özetlenmesidir [16, 17].

Veri madenciliği teknikleri, denetimsiz ve denetimli öğrenme teknikleri olarak sınıflandırılabilir. Denetimsiz öğrenme tekniği, değişken tarafından yönlendirilmeyen ve analizden önce bir hipotez oluşturmayan tekniklerdir. Bu tekniklerde, modeller elde edilen sonuçlara dayanarak üretilmektedir. Literatürde sıklıkla karşılaşılan denetimsiz tekniklerden birisi kümelemedir. Denetimli öğrenme tekniği, algoritmanın veri

setindeki alana özgü etiketler ile eğitilmesi ile gerçekleştirilmektedir. Hem tıbbi hem de klinik araştırmalarda kullanılan denetimli öğrenme teknikleri genellikle sınıflandırma, regresyon ve birliktelik kurallarıdır [15]. Bu çalışmada engelli bireylerin öz bakım aktivitelerinin ICF-CY'ye göre sınıflandırılabilmesi için denetimli öğrenme tekniklerinden olan C5.0, CART, QUEST, CHAID, Destek Vektör Makineleri (DVM), Yapay Sinir Ağları (YSA) ve Bayes Ağları (BA) kullanılmıştır. Kullanılan yöntemler ile ilgili kısa bilgilere ve gerçekleştirilen deneysel uygulamalara bu bölümde yer verilmiştir.

2.1 Tahmin modelleri

2.1.1 C5.0

Ross Quinlan, 1970'li yılların sonlarına doğru karar ağaçlarına dayalı olarak işlem yapan ID3 algoritmasını tanıtmıştır [18, 19]. ID3 metodunun eksikliklerinin giderilmesi amacıyla 1993 yılında algoritma genişletilmiş ve ID3 prensiplerine dayanan C4.5 algoritması geliştirilmiştir. Yapılan çalışmalar sonucunda, C4.5 algoritmasının ID3'e göre daha iyi performans gösterdiği görülmüştür [20, 21]. C4.5 algoritmasının da iyileştirilmesi ile temel olarak karar ağacı ve kural seti olmak üzere iki tip model üretme becerisine sahip olan C5.0 algoritması elde edilmiştir. Karar ağacı, algoritma tarafından bulunan budamaların basit ve düz bir tanımlamasıdır. Kural seti ise karar ağacından elde edilen mevcut bilginin basitleştirilmiş bir versiyonunu ifade etmektedir [22].

C5.0 algoritması sınıflama işlemini gerçekleştirirken entropi ve bilgi kazancından faydalanmaktadır. Hesaplanan ölçütler ile her değişken için belirli değerler dallara ayrılmakta olup bölümlene yapılabilecek herhangi bir düğüm kalmayınca kadar bu süreç sürmektedir. Bu doğrultuda, seçilen değişkenin her bir değeri dallara çevrilmekte ve diğer değişkenlerin de sürecin devamında dikkate alınması ile aynı işlem devam etmektedir [23].

C5.0 algoritmasının güçlü yanlarından bazıları; çok amaçlı bir sınıflandırıcı olarak çeşitli problemlerle çalışabilmesi, mevcut yöntemlere nazaran daha üretken olması, sadece en önemli ve anlamlı özelliklerden faydalanması, eğitim için hem nispeten küçük hem de çok büyük örneklerle çalışabilmesidir. Algoritmanın dezavantajlarından bazıları ise; özelliklerin çok sayıda seviyeye sahip olduğu bölümlere ayrılması sürecinde yanlış davranabilmesi, modelde aşırı uyum veya az uyum gibi problemlerin oluşabilmesi, karar mantığındaki büyük değişikliklerin eğitim verisinde küçük değişiklikler olarak çıktı verebilmesidir [24].

C5.0'ın en önemli özelliği boosting tekniğini kullanmasıdır. Boosting tekniği ile doğruluğun artırılması amaçlanmaktadır. Her ne kadar boosting tekniği kullanıldığında algoritmanın çalışma süresinin artma ihtimali bulunsun da birçok örnekte tahmin doğruluğunu artırdığı görülmektedir. C5.0 algoritması, veri madenciliğindeki en önemli zorluklardan biri olan ve bir sınıftaki veri sayısının diğer sınıflardakinden daha yüksek olmasını ifade eden veri dengesizliği ile başa çıkabilen etkili bir çözümdür.

Boosting tekniğinin algoritma adımları Pang ve Gong tarafından aşağıdaki gibi tanımlanmıştır [25]. Burada N toplam örnek sayısını, S örnek kümesini ve i ele alınan örneğin numarasını temsil etmek üzere;

$$i=1,2,\dots,N,$$

$$T = \text{karar ağaçlarının sayısı, } t=1,2,\dots,T,$$

C^t = t denemesinde öğrenme sisteminin oluşturduğu karar ağacı,

$$C^* = \text{nihai karar ağacı,}$$

$$p^t = w_i^t \text{ nin normalize faktörü,}$$

$$\beta^t = \text{ağırlık ayarlama faktörü,}$$

$$\theta_i^t = i \text{ örneği yanlış sınıflandırılmış ise 1 aksi takdirde 0.}$$

Adım 1: Değişkenlere ilk değerleri ata; $T=10$ (T değeri genellikle 10'dur) ve $t = 1$, bu durumda $w_i^1 = \frac{1}{n}$

Adım 2: $\sum_{i=0}^n(p_i^t) = 1$ kabul edildiğinde Denklem (1) değerini hesapla.

$$p_i^t = w_i^t / \sum_{i=0}^n(w_i^t) \quad (1)$$

Adım 3: Her bir örnek için ağırlık olarak p_i^t 'yi kullan ve bu dağılım altında C^t yi oluştur.

Adım 4: C^t 'nin hata oranını hesaplarken Denklem (2)'yi kullan.

$$\varepsilon^t = \sum_{i=0}^n(p_i^t \theta_i^t) \quad (2)$$

Adım 5: $\varepsilon^t < 0.5$ ise deneyi bitir, $T = t + 1$

$$\varepsilon^t = 0 \text{ ise deneyi bitir, } T = t$$

$$0 < \varepsilon^t < 0.5 \text{ ise adım 6'dan devam et.}$$

Adım 6: β^t 'yi Denklem (3) ile hesapla,

$$\beta^t = \varepsilon^t / (1 - \varepsilon^t) \quad (3)$$

Adım 7: Hata oranını dikkate alarak ağırlık ayarlamalarını aşağıdaki gibi hesapla.

$$\text{Örnek yanlış sınıflandırıldıysa } w_i^{t+1} = w_i^t \beta^t,$$

$$\text{Örnek doğru sınıflandırıldıysa } w_i^t$$

Adım 8: $t = T$ ise deneyi bitir.

Aksi durumda, $t = T + 1$ ve Adım 2'ye git ve bir sonraki deneye başla.

2.1.2 CART (classification and regression trees)

Beirman vd. [26] tarafından 1984 yılında geliştirilen CART algoritması, hedef değişkenin sürekli veya kategorik olma durumuna bağlı olarak regresyon modeli veya sınıflandırma modeli üretmektedir. Yani hem sınıflandırma hem de regresyon problemlerinde bir çözüm olarak kullanılabilir. Hedef değişken sürekli ise regresyon modeli üretilmektedir. Hedef değişkenin değerini tahmin etmek için regresyon modeli kullanıldığında, ağacın yaprak düğümüne düşen satırın ortalama hedef değişken değeri, tahmin edilen değerdir. Hedef değişken kategorik ise bir sınıflandırma modeli üretilmektedir. Hedef değişkenin değerini tahmin etmek için yaprak düğümüne ulaşana kadar

tahmin edici değişken değeri kullanılarak ağaç üzerinde ilerlenmekte ve bu düğüm için gösterilen kategori tahmin edilmektedir [27].

CART algoritması, bağımsız değişkenlere dayanarak homojenliği artan alt kümeleri bulabilmek için veriyi tekrar tekrar parçalara ayırmaktadır. Bağımlı veri, sonradan gelen bir dizi sağ ve sol alt düğümlere ayrıştırılmaktadır [27]. CART algoritması, Gini indeksini dallanma kriteri olarak kabul etmekte olup kuruluş süresince herhangi bir durma kuralı olmadan sürekli bölünerek büyümektedir. Yeni bir bölünmenin mümkün olmadığı durumda ise bu defa uçlardan köklere doğru giden bir budama süreci başlamaktadır. Her bir budama sonrasında bağımsız bir test verisi kullanılarak analiz yapılmakta ve en başarılı karar ağacı tespit edilmektedir [28].

2.1.3 Quest (quick, unbiased, and efficient statistical tree)

QUEST, Wei-Yin Loh (Wisconsin-Madison Üniversitesi, ABD) ve Yu-Shan Shih (National Chung Cheng Üniversitesi) tarafından 1997 yılında geliştirilen ikili karar ağacı yapısı kullanan bir sınıflandırma algoritmasıdır [29]. Bu algoritma, tek değişkenli ve doğrusal kombinasyonlu bölünmeleri desteklemektedir. Her bir bölünme sürecinde girdi değişkeni ile hedef değişken arasındaki ilişki, değişken türüne bağlı olarak Anova F-testi, Levene'nin testi veya Pearson'ın Ki-Kare testi ile hesaplanmaktadır [30]. CHAID ve CART algoritmalarından farklı olarak ağacın oluşturulma sürecinde değişken seçimi ile bölünmeyi aynı anda yapmamakta, her bir süreç ile ayrı ayrı ilgilenmektedir. Geliştirilmesinin temel amacı dallanma sürecindeki yanlış seçimin daha genel bir hale getirilmesi ve hesaplama maliyetlerinin minimize edilmesidir [31].

2.1.4 CHAID (chi-square automatic interaction detector)

CHAID 1980'li yıllarda Kass [32] tarafından geliştirilmiş sınıflandırma ve tahmin için kullanılabilen bir karar ağacı algoritmasıdır. Bu algoritma tarafından kategorik girdi ve hedefler kullanıldığından, hedef özellik ile her bir mevcut tahmin edici arasında ki-kare testi hesaplanmakta ve sonucunda örneklerin parçalara ayrılmasında en iyi tahmin ediciden faydalanılmaktadır. Süreç, herhangi bir anlamlı bölme işlemi kalmayınca kadar devam etmektedir [33], [34]. CHAID algoritmasının adımları aşağıda sunulmuştur:

Adım 1: Her bir sürekli tahmin edici için kategorik tahmin ediciler oluşturulur.

Adım 2: İki den fazla satır ile satır kategorileri kombine edilerek en uygun form elde edilir.

Adım 3: Kombine edilen kategoriler daha küçük parçalara bölünür.

Adım 4: Optimal kombinasyon tamamlanır ve açıklayıcı değişkenler ile kombine edilir. Her bir açıklayıcı değişkenin tablosunu azaltmak için ki-kare testini uyarlayabilmek için "Bonferroni" hesaplanır.

Adım 5: Adım 4'te hesaplanan en anlamlı değerler, düğümü kategorilere göre ayırmak için kullanılır.

CHAID algoritması veri yönetiminde iki farklı istatistiksel test kullanılmaktadır [35].

(i) Eğer bağımlı değişken kategorik ise Pearson ki-kare testi aşağıdaki gibi uygulanır.

n_{ij} = gözlemlenen hücre sıklığını,

m_{ij} = tahmin edilen beklenen hücre sıklığını ($x_n = i$,

$y_n = j$ için),

f_n = duruma bağlı sıklık ağırlığını,

x_n ve y_n = eğitim örneğini,

I = bağımsızlık derecesini,

D = ilgili veriyi temsil etmektedir.

Bu durumda X^2 , n_{ij} ve P-değeri sırasıyla Denklem (4), (5) ve (6) kullanılarak hesaplanır.

$$X^2 = \sum_{j=1}^J \sum_{i=1}^I \left(\frac{(n_{ij} - m_{ij})^2}{m_{ij}} \right) \quad (4)$$

$$n_{ij} = \sum_{n \in D} f_n I(x_n = i \cap y_n = j) \quad (5)$$

$$P = \Pr(x_d^e > x^2) \quad (6)$$

(ii) Ölçek bağımlı değişkenler için ise F testi kullanılır. F-değeri ve F-değeri için gerekli olan y_n , \bar{y} , N_f değerleri sırasıyla Denklem (7), (8), (9), (10) ile hesaplanır.

$$F = \frac{\sum_{i=1}^I \sum_{n \in D} w_n f_n I(X_n = i) (\bar{y} - y)^2 / (I-1)}{\sum_{i=1}^I \sum_{n \in D} w_n f_n I(X_n = i) (y_n - \bar{y}_i)^2 / (N_f - I)} \quad (7)$$

$$y_n = \frac{\sum_{n \in D} w_n f_n y_n I(x_n = i)}{\sum_{n \in D} w_n f_n I(x_n = i)} \quad (8)$$

$$\bar{y} = \frac{\sum_{n \in D} w_n f_n y_n I}{\sum_{n \in D} w_n f_n I} \quad (9)$$

$$N_f = \sum_{n \in D} f_n \quad (10)$$

$$P = P(F(I-1, N_f-1) > F) \quad (11)$$

2.1.5 Destek vektör makineleri

Destek Vektör Makineleri (DVM), Vapnik ve Chervonenkis [36] tarafından optimal ayırma hiper düzlemi kullanılarak geliştirilen makine öğrenmesinin yeni bir türü olarak ortaya çıkmıştır. DVM ile sınıflandırmanın amacı, eğitim verisi ile elde edilen bir karar fonksiyonu yardımıyla, genellikle $\{-1, +1\}$ gibi sınıf etiketleri ile gösterilen iki örnek sınıfını birbirinden ayırmaktır. Bununla birlikte, DVM'nin hedefi, en yakın noktalar arasındaki mesafeyi en üst seviyeye çıkararak hiper düzlemi bulmaktır. Hiper düzlemin iki tarafında Denklem (12) ve (13)'ü sağlayan noktalar

arasındaki düz mesafelerin toplamına sınır olarak isimlendirilmektedir. Burada $x \in \mathbb{R}^n$ boyutlu bir uzayı, w ağırlık vektörünü (hiper düzlemin normali) ve b eğilim değerini belirtmektedir. Bu limiti maksimum yapan hiper düzlem, optimum hiper düzlemi vermektedir. Optimum hiper düzlemi belirlemek için, bu düzleme paralel olan ve sınırları oluşturan iki hiper düzlem belirlenmelidir [37, 38].

$$w \times x_i - b = -1 \quad (12)$$

$$w \times x_i - b = +1 \quad (13)$$

Optimum hiper düzlem sınırını maksimize etmek için, ağırlık vektörünün minimize edilmesi gerekmektedir. Bu durumda, en uygun hiper düzlemin belirlenmesi aşağıdaki sınırlı optimizasyon probleminin çözümünü gerektirmektedir [39]. Buna bağlı olan sınırlamalar, eşitsizlikte gösterildiği şekilde ifade edilmektedir. Bu optimizasyon problemi Lagrange Eşitlikleri kullanılarak çözülebilmektedir. Lagrange Eşitlikleri kullanıldığında Denklem (16) elde edilmekte ve doğrusal olarak ayrılabilen iki sınıflı bir problem için karar fonksiyonu Denklem (17)'deki gibi yazılabilmektedir [40].

$$\min \left[\frac{1}{2} \|w\|^2 \right] \quad (14)$$

$$y_i(w \cdot x_i + b) - 1 \geq 0 \text{ ve } y_i \in \{1, -1\} \quad (15)$$

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^k a_i y_i (w \cdot x_i + b) + \sum_{i=1}^k a_i \quad (16)$$

$$f(x) = \text{sign} \left(\sum_{i=1}^k \lambda_i y_i (x \cdot x_i) + b \right) \quad (17)$$

Verilerin doğrusal olarak ayrılabilmesi durumunda, DVM, orijinal verileri bir haritalama yöntemiyle daha yüksek bir boyuta taşımaktadır. Burada, verileri sınıflandırmak için uygun olabilecek lineer ayırıcı hiper düzlem bulunmaya çalışılmaktadır. Sınırı maksimize etmek ve yanlış sınıflandırma hatalarını minimize etmek arasındaki denge, pozitif değerler alan ve C ile temsil edilen bir regresyon parametresi ($0 < C < \infty$) tanımlanarak kontrol edilebilmektedir [39]. Regresyon parametresi ve yapay değişken (ξ_i) kullanılarak doğrusal olmayan ayrılabilir veriler için optimizasyon problemi ve yapay değişken aşağıdaki gibi tanımlanmaktadır [39-41]:

$$\min \left[\frac{1}{2} \|w\|^2 + C \cdot \sum_{i=1}^r \zeta_i \right] \quad (18)$$

$$y_i(w \cdot \varphi(x_i) + b) - 1 \geq 1 - \zeta_i \text{ ve } \zeta_i \geq 0 \text{ ve } i=1, \dots, n \quad (19)$$

Optimizasyon probleminin çözümü için giriş alanındaki veriler, özellik alanı olarak tanımlanan yüksek boyutlu bir

alanda görüntülenmektedir. Böylece, veriler doğrusal olarak ayrılabilen ve sınıflar arasındaki hiper düzlem belirlenebilmektedir. Destek vektör makineleri, Denklem (20)'deki gibi ifade edilen bir çekirdek fonksiyonunun yardımıyla doğrusal olmayan dönüşümleri matematiksel olarak yapabilmektedir. Böylece verilerin yüksek boyutta doğrusal olarak ayrılması sağlanmakta ve çekirdek fonksiyonu kullanılarak doğrusal olarak ayrılamayan iki sınıflı bir problemin çözülmesi için karar fonksiyonu Denklem (21)'deki gibi yazılabilmektedir.

$$K(x_i, x_j) = \varphi(x) \cdot \varphi(x_j) \quad (20)$$

$$f(x) = \text{sign} \left(\sum_{i=1}^k a_i y_i \varphi(x) + \varphi(x_i) + b \right) \quad (21)$$

DVM tarafından yapılacak sınıflandırma işleminde kullanılacak çekirdek fonksiyonunun ve bu fonksiyonun optimum parametrelerinin belirlenmesi esastır. Parametrelerin kesin bir değeri yoktur ve her veri seti birbirinden farklı olduğu için istenen sınıflandırma doğruluğuna ulaşana kadar farklı değerler denenmektedir. Çekirdek fonksiyonları, doğrusal (Denklem (22)), polinom (Denklem (23)), radyal (Denklem (24)) ve sigmoid (Denklem (25)) olarak farklı temel fonksiyonlara sahiptir. Literatürde en sık kullanılan çekirdek fonksiyonları polinom ve radyal temel fonksiyonlardır [42]. Eşitliklerde yer alan γ , r ve d çekirdek parametreleridir.

$$K(x_i, x_j) = x_i^T \cdot x_j \quad (22)$$

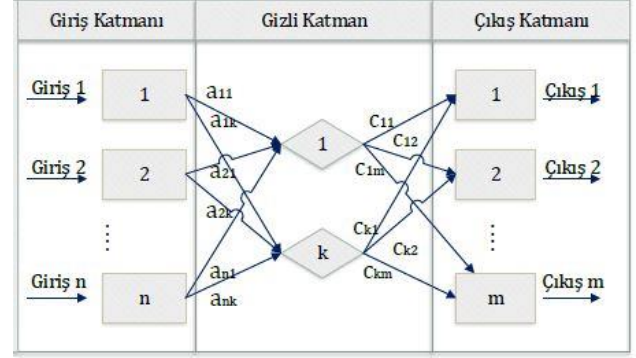
$$K(x_i, x_j) = (\gamma x_i^T \cdot x_j + r)^d, \gamma > 0 \quad (23)$$

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0 \quad (24)$$

$$K(x_i, x_j) = \tanh(\gamma x_i^T \cdot x_j + r) \quad (25)$$

2.1.6 Yapay sinir ağları

İnsan beyindeki sinir sisteminin çalışma prensiplerinden esinlenen YSA ile ilgili ilk çalışma 1943 yılında McCulloch ve arkadaşları [43] tarafından yapılmış, elektrik devreleri üzerinde temel seviyede denemeler gerçekleştirilmiş ve basit bir yapay sinir ağı oluşturulmuştur. Sonraki süreçte, Rosenblatt [44]'in tek katmanlı ve eğitilebilen perceptron ağı modelini geliştirmesiyle YSA ile ilgili çalışmalar gittikçe artmış ve çeşitli modeller önerilmiştir [45]. İnsan beyinin bir simülasyonu olarak düşünülebilen YSA modelleri, sınıflandırma, optimizasyon, ilişki kurabilme, eksik veri ile çalışabilme, örnek veriler sayesinde öğrenme becerisi kazanabilme gibi çeşitli fonksiyonları gerçekleştirebilmektedir [46]. Bu sebeple, YSA birçok problemin çözümünde araştırmacılar tarafından başvurulan yöntemlerden biri olmuştur. Bir YSA modeli girdi katmanı, gizli katman ve çıkış katmanından oluşmaktadır. Çok katmanlı bir YSA örneği Şekil 1'de sunulmuştur.



Şekil 1. Çok katmanlı YSA'nın genel yapısı

Giriş katmanında vektör olarak tanımlanmış ağırlık girdileri yer almaktadır. Gizli katmanda aktivasyon olarak isimlendirilen bir işlemci bulunmakta olup hücrelerdeki net girdi değeri karşılığında çıktı olarak elde edilebilecek değerin hesaplanmasını sağlamaktadır. Bazı YSA modellerinde aktivasyon fonksiyonunun türevlenebilir olması zorunluluğu bulunmaktadır.

YSA'da ağırlık istenen performansı ve çıktıyı verebilmesi için ağırlık eğitilmesi ve test edilmesi gerekmektedir. Ağırlık eğitilmesindeki en önemli amaç kullanılan öğrenme algoritmaları ile girdi ve çıktı verileri arasındaki en uygun ilişkinin elde edilebilmesi ve bunu sağlayacak ağırlıkların ortaya çıkarılmasıdır.

YSA'da bir öğrenme kuralı esas alınarak eşik fonksiyonu yardımıyla "w" ağırlık vektörünü, istenilen performansı sağlayacak şekilde ayarlayabilmek çok önemlidir. Bu noktada, performans ölçütü en küçükleme veya en büyükleme açısından ele alınmakta ve çoğunlukla karesel hata en küçüklenmesi amacıyla Denklem (26) kullanılarak araştırılmaktadır [46]:

$$\min F(w) = \int_A |f(x) - G(x, w)|^2 p(x) dv(x) \quad (26)$$

$G(x, w)$: Giriş ve çıkış fonksiyonu,

y : Çıkış vektörü,

x : Giriş vektörü,

w : Ağırlık vektörü,

$p(x)$: Olasılık yoğunluk fonksiyonu.

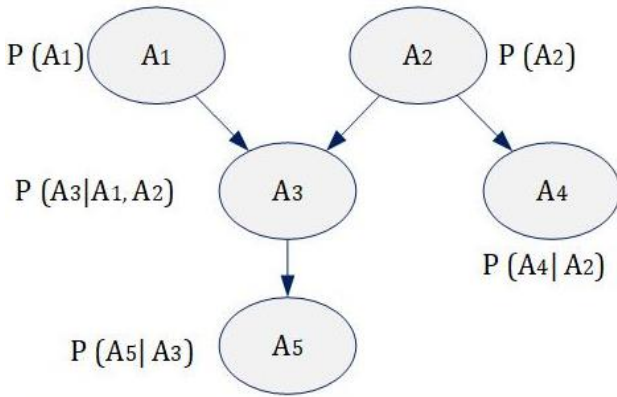
YSA modelleri, yapılarına göre ileri beslemeli ve geri beslemeli olmak üzere iki farklı şekilde tasarlanabilmektedir. Öğrenme algoritmalarına göre sınıflandırıldığında ise danışmanlı, danışmansız ve takviyeli olmak üzere üçe ayrılmaktadır. İleri beslemeli ağlar örüntü tanıma, geri beslemeli ağlar ise optimizasyon problemlerinde sıklıkla tercih edilmektedir [47].

2.1.7 Bayes ağları

Bayes ağları veri modelleme yöntemlerinden biri olup, olasılık hesaplamalarına dayanmaktadır. Bayes ağları, bayes teoremi esas alınarak ortaya çıkarılmıştır ve bayes olasılık kurallarını kullanan grafiksel modellemeye dayalı bir yöntemdir [48]. Bir bayes ağı, olasılıksal parametreler ile grafiksel yapılardan oluşmaktadır. Grafiksel yapısı,

düğümlemlerden ve oklardan oluşan tek yönlü çevrimsiz grafiğdir. Düğümler, deęişkenleri temsil ederken oklar deęişkenler arasındaki nedensel ilişkileri ifade etmektedir. İki deęişken birbirine bağlandığında okun ucundaki deęişken çocuk (child) deęişken, okun başlangıcındaki deęişken ise ebeveyn (parent) deęişken olarak adlandırılmaktadır [49]. Bayes ağında yer alacak deęişkenlerin sahip olabileceęi ebeveyn veya çocuk sayısı için herhangi bir kısıtlama bulunmamaktadır.

Şekil 2’de A1, A2, A3, A4, A5 deęişkenlerinden oluşan bir Bayes ağı örnek olarak sunulmuştur. Şekil 2, çocuk ve ebeveyn deęişkenleri açısından incelendiğinde; A1 ve A2 deęişkenlerinin A3 deęişkeninin ebeveyni, A3 deęişkeninin A5 deęişkeninin ebeveyni, A4 deęişkeninin ise A2 deęişkeninin çocuk deęişkeni olduęu görülmektedir. Ayrıca şekilde söz konusu deęişkenlerin sahip oldukları koşullu olasılık dağılımları da $P(A1)$, $P(A2)$, $P(A3|A1A2)$, $P(A4|A2)$, $P(A5|A3)$ olarak belirtilmiştir. Ağda yer alan herhangi bir deęişken ile başka bir deęişken arasında bir ok bulunmaması bu iki deęişken arasında olasılıksal bir ilişki bulunmadığını yani ağda marjinal olasılık dağılımı ile yer aldığı anlaşılmaktadır [50].



Şekil 2. Bayes ağı örneęi

Bir bayes ağında bulunan deęişkenlerin sahip oldukları koşullu olasılıkların çarpımı, ağı birleşik olasılık dağılımını göstermekte olup Denklem (27) ile hesaplanmaktadır. Burada N, ağda bulunan deęişkenlerin toplam sayısını ifade etmektedir [50].

$$P(X_1, \dots, X_n) = \prod_{i=1}^N P(X_i | E_{beveyn}(X_i)) \quad (27)$$

Tablo 1. Teşhis kategorileri [13].

Teşhis kategorisi	N	Teşhis kategorisi	N
Serebral Palsi		Pompe hastalığı	1
-Dipleji	26	Kafa travması	1
-Kuatripleji	12	Mukopolisakkaritler	1
-Hemipleji	4	Akçaağaç şurubu idrar hastalığı	1
Miyelomeningosel	8	Cam kemik hastalığı	1
Kas distrofisi	7	Konjenital el anomalisi	1
Çarpık ayak	3	Beyin sapı tümörü	1
Cücelik	2	Hidrocefali	1

“N” kategoride yer alan birey sayısını ifade etmektedir.

Bayes ağında çıkarım süreci, gözlem yapıldıkça ağı güncellenmesi, yani Bayes teoreminin art arda uygulanması olarak ifade edilebilmektedir.

2.2 Kolektif (ensemble) öğrenme

Kolektif öğrenme, aynı sınıflandırıcının farklı eğitim setleri ile eğitilmesi veya aynı eğitim verisinin farklı sınıflandırma algoritmaları ile değerlendirilmesi sonucunda elde edilen doğruluk oranlarının birleştirilmesidir. Kolektif öğrenme, genelleme yeteneğini geliştiren ve her birinin aşırı yüklenme riskini azaltan kombine bir model elde etmek için bireysel veya heterojen modellerin bir araya getirilmesi ile oluşmaktadır [51].

Dietterich [52]’in çalışmasında bir kolektif sınıflandırıcının tek bir sınıflandırıcıdan daha iyi olduğunu gösteren üç temel özellik şu şekilde sunulmuştur: (i) Sadece tek bir sınıflandırıcı seçildiğinde, kötü olanı seçme riski üstlenilmiş olmaktadır. (ii) Birçok öğrenme algoritması, yerel aramayı kullandığı için algoritma en uygun sonucu bulamayabilmektedir. Bu yüzden öğrenme algoritmasının birkaç kez çalıştırılması ve elde edilen modellerin birleştirilmesi, algoritmaların herhangi birinden daha iyi sonuç elde edilmesini sağlayabilmektedir. (iii) Çoğu veri madencilięi probleminde uygun bir sınıflandırıcı bulunamadığında çeşitli sınıflandırıcıları birleştirerek en uygun sınıflandırıcıya ulaşılabilmektedir

2.3 Veri seti

Bu çalışmada SCADI olarak isimlendirilen ve ICF-CY çerçevesine göre oluşturulan, fiziksel ve motor engelli çocukların ve genç bireylerin öz bakım aktivitelerini içeren veri seti kullanılmıştır. Veri seti, 14 Nisan 2018 tarihinde İran’daki İslam Azad Üniversitesi, Yazılım Mühendislięi Bölümü’nden Dr. Fatemi Bushehri tarafından araştırmacıların kullanması için çevrimiçi ortamda sunulmuştur [53]. SCADI, eğitim ve sağlık merkezlerinde 15 yıldan fazla iş deneyimi olan iki meslek terapisti tarafından toplanmıştır. SCADI kayıtlarının toplanması için 2016’dan 2017’ye kadar İran’ın Yazd şehrindeki üç eğitim ve sağlık merkezinde 70 çocuk ve genç birey incelenmiştir.

SCADI analiz edildiğinde, terapistler tarafından incelenen bireylerin yaşının 6 ile 22 arasında deęiştięi ve 41’inin kadın, 29’unun erkek olduęu görülmüştür. Söz konusu bireylerin sahip olduęu hastalıklara yönelik uzman terapist teşhisleri Tablo 1’de sunulmuştur.

Tablo 2. Öz bakım aktiviteleri.

Kategori	Aktiviteler (Kodlar)
Yıkama	Vücut parçalarını yıkama (d5100), Tüm vücudu yıkama (d5101), Kurulanma (d 5102)
Vücut parçalarının bakımı	Cilt bakımı (d5200), Diş bakımı (d5201), Saç bakımı (d5202), El tırnak bakımı (d5203), Ayak tırnak bakımı (d5204), Burun bakımı (d5205)
Tuvalet ihtiyacını karşılama	İdrar ihtiyacının belirtilmesi (d53000), İdrarın uygun şekilde yapılması (d53001), Defekasyon ihtiyacının belirtilmesi (d53010), Defekasyonun uygun şekilde yapılması (d53011), Regl bakımı (d5302)
Giyinme	Giyisi giyme (d5400), Giysi çıkarılma (d5401), Ayakkabı giyme (d5402), Ayakkabı çıkartma (d5403), Uygun kıyafeti seçme (d5404),
Yeme İçme	Yemek ihtiyacını belirtme (d5500), Uygun şekilde yemek yeme (d5501) İçme ihtiyacını belirtme (d5600), Biberon ile beslenme (d5602)
Sağlığın gözetimi	Fiziksel rahatlığın sağlanması (d5700), Diyet ve fitness yönetimi (d5701), İlaçları yönetmek ve sağlık tavsiyelerinin izlenmesi (d57020), Bakıcılar veya profesyonellerden tavsiye veya yardım isteme (d57021), Uyuşturucu veya alkolün kötüye kullanılmasıyla ilgili risklerden kaçınma (d57022)
Güvenliğini gözetme	Kişisel güvenliğinin gözetilmesi (d 571)

SCADI'de ICF-CY'ye dayalı öz bakım için 29 aktivite yer almaktadır. Her kategoride yer alan faaliyetler ve bu faaliyetlerin kodları Tablo 2'de sunulmuştur. SCADI oluşturulurken 29 aktivite için binary veri tipi kullanılmıştır. Yani, kişinin söz konusu öz bakım aktivitesinde engel varsa, veri setindeki ilişkili hücreye 1, aksi takdirde 0 değeri işlenmiştir.

SCADI oluşturulurken faaliyet ve katılım bileşenlerindeki bozulma derecesini açıklamak için Tablo 3'te yer alan 7 kod kullanılmıştır. Her bir öz bakım aktivitesi, sunulan 7 bozukluk türünden birine sahip olabilmektedir. Bu durumdan dolayı SCADI, 203 (29*7) tane öz bakım aktivitesine yaş ve cinsiyet değişkenlerinin de eklenmesiyle 205 değişken olarak hazırlanmıştır.

Tablo 3. Bozukluk derecesi

Kod	Açıklama	Değer (%)
xxx.0	Bozukluk yok	0-4
xxx.1	Hafif bozukluk	5-24
xxx.2	Orta dereceli bozukluk	25-49
xxx.3	Şiddetli bozukluk	50-95
xxx.4	Tam bozukluk	96-100
xxx.8	Belirtilmemiş	
xxx.9	Uygulanamaz	

SCADI'de tahmin edilmesi beklenen "sınıf" değişkeni, fiziksel ve motor engelli çocukların ve genç bireylerin öz bakım sorunlarıdır. Meslek terapistleri, Tablo 1'de belirtilen hastalıklara sahip olan bireyleri inceledikten sonra bireylerde teşhis edilen Tablo 2'deki öz bakım aktivitelerini, 7 kategoriye ayırarak inceledikleri bireyleri uygun sınıfa atamışlardır (Tablo 4).

Tablo 4. Hedef değişken (sınıflar).

Hedef	Açıklama	Sayı
Sınıf 1	Vücut parçalarına bakım sorunu	2
Sınıf 2	Tuvalet sorunu	7
Sınıf 3	Giyinme sorunu	1
Sınıf 4	Yıkama, vücut parçalarına bakım ve giyinme sorunu	12
Sınıf 5	Yıkama, vücut parçalarına bakım, tuvalet ve giyinme sorunu	3
Sınıf 6	Yeme, içme, yıkama, vücut parçalarına bakım, tuvalet, giyinme, birinin güvenliği ve sağlığı ile ilgilenebilme sorunu	29
Sınıf 7	Sorun yok	16

2.4 Deneysel çalışma

Engelli bireylerin öz bakım aktivitelerindeki problemlere ilişkin sınıf tahmini uygulamasının, yüksek doğruluk oranı ile gerçekleştirilebilmesi için çeşitli deneysel çalışmalar yapılmıştır. Söz konusu deneysel çalışmaların uygulanma sürecinde, kullanıcının istatistiksel ve veri madenciliği algoritmalarından faydalanmasını sağlayan ve görsel ara yüze sahip olan IBM SPSS Clementine modelleyiciden yararlanılmıştır. Teorik çerçevesi Şekil 3'te sunulan deneysel çalışma, veri setinin tanımlanması, özellik seçimi, eğitim ve test verisinin oluşturulması, algoritmaların uygulanması ve bulguların değerlendirilmesi olmak üzere temelde beş aşamadan oluşmaktadır.

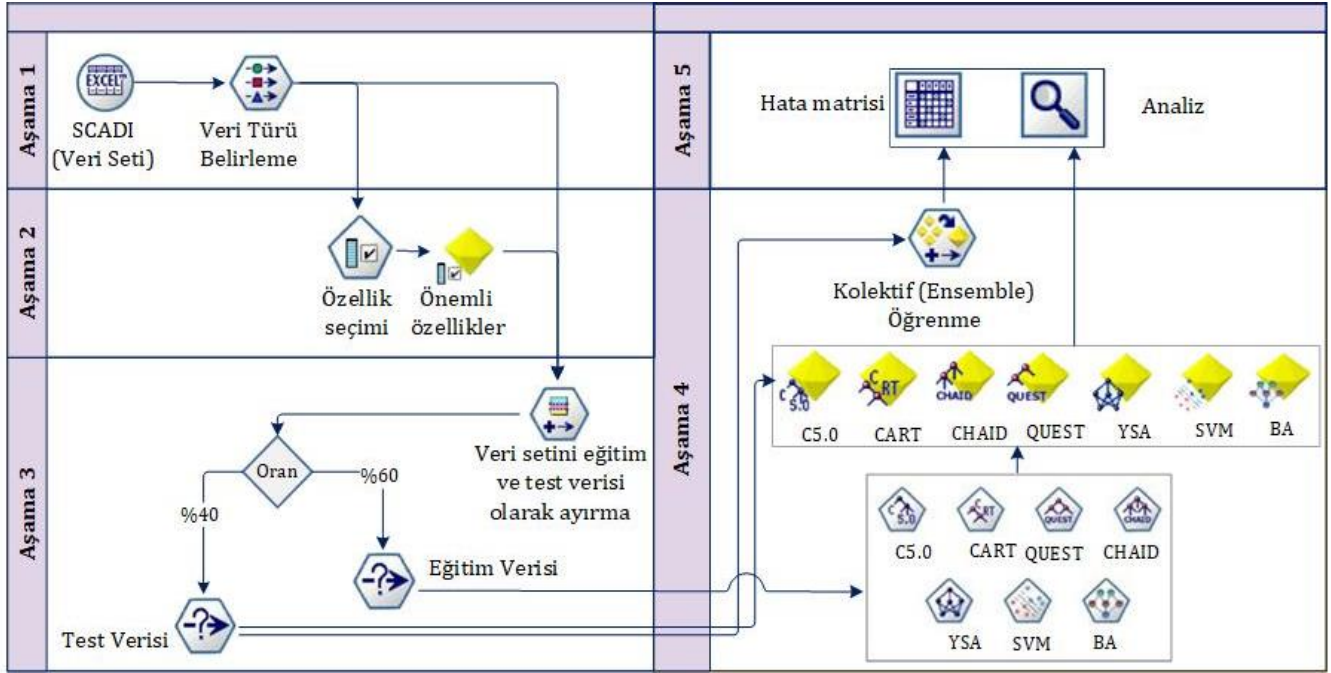
2.4.1 Veri setinin tanımlanması

SCADI'ye veri madenciliği algoritmasının uygulanması için veriler uygun formata dönüştürülerek modelleyiciye yüklenmiştir. Daha sonra veri setinde yer alan değişkenlerin tipleri (cinsiyet: flag, yaş: range, sınıf: flag ve d5 kodu ile başlayan öz bakım aktivitesi değişkenleri: flag) tanımlanmıştır.

2.4.2 Özellik seçimi

Sınıflandırma tahmininin yüksek doğruluk oranı ile gerçekleştirilebilmesi amacıyla ihtiyaç duyulan güvenilir modelleri oluşturmak için, özelliklerin mümkün olduğunca yararlı bilgi içermesi beklenmektedir. Bununla birlikte, veri seti hakkında genellikle çok az ön bilgi sahibi olunduğu için, hangi özelliklerin uygun veya birbirinden farklı olduğunu ayırt etmek zordur. Özellik seçimi, veri madenciliği yöntemlerinin uygulanmasında önemli bir aşamadır. Veri setleri genellikle verimli model yapısı için çok fazla değişkenle belirtilmektedir. Ancak ilgisiz ve gereksiz özellikler, sadece eğitim verimliliğini azaltmakla kalmaz, aynı zamanda boyut laneti/belası olarak adlandırılan probleme neden olarak modelin performansını da olumsuz yönde etkilemektedir [54, 55]. Bu nedenle mümkün olan en iyi sınıflandırma sonucunu verebilecek minimal özellik setinin belirlenmesi gerekmektedir.

SCADI analiz edildiğinde 19 değişkenin sadece 0 değerine sahip veri içerdiği belirlenmiştir. Bu durum veri setinde sonucu yanıtlanabilecek gereksiz verinin olduğunu ve veri madenciliği analizi yapılmadan önce özellik seçiminin



Şekil 3. Teorik çerçeve

yapılmasını gerektiğini göstermektedir. SPSS Clementine modelleyicide yer alan özellik seçimi düğümü (feature selection node) aracılığı ile gerçekleştirilen özellik seçimi sonucunda, değişkenlerin sınıf ile ilişkileri analiz edildiğinde 205 değişken içerisinde 80 değişken önemli, 4 değişken marjinal yani düşük önemli, 15 değişken ise önemsiz olarak belirlenmiştir. Geri kalan 106 değişken ise fazla sayıda tek değişken içerdiği için sınıflandırmada ayırt edici bir etkiye

neden olmayacağından veri setinden çıkartılmıştır. Diğer değişkenler, eğitim ve test verisi oluşturmak üzere ayrılmıştır. Önemli bulunan değişkenlerin sırası ve önem değeri Tablo 5'te sunulmuştur. Özellik seçiminin sınıf tahmini sonuçları üzerindeki etkisinin analiz edilmesi için deneysel çalışmalar özellik seçimi yapılmış ve yapılmamış veriler için ayrı olarak uygulanmıştır.

Tablo 5. Önemli değişkenler.

Sıra	Değişken	Değer	Sıra	Değişken	Değer	Sıra	Değişken	Değer	Sıra	Değişken	Değer
1	d 5400-4	100.00%	21	d 5100-4	100.00%	41	d 53011-1	100.00%	61	d 5602-0	99.88%
2	d 5402-4	100.00%	22	d 5602-2	100.00%	42	d 5201-1	99.99%	62	d 5403-1	99.85%
3	d 5101-4	100.00%	23	d 53001-2	100.00%	43	d 5401-2	99.99%	63	d 5700-3	99.79%
4	d 5102-4	100.00%	24	d 53011-2	100.00%	44	d 5501-1	99.99%	64	d 571-1	99.74%
5	d 5203-4	100.00%	25	d 5102-3	100.00%	45	d 5202-1	99.99%	65	d 5501-0	99.59%
6	d 5204-4	100.00%	26	d 5201-3	100.00%	46	d 5203-2	99.99%	66	d 5101-1	99.55%
7	d 53001-4	100.00%	27	d 5202-3	100.00%	47	d 5204-2	99.99%	67	d 5200-1	99.55%
8	d 53011-4	100.00%	28	d 5205-3	100.00%	48	d 5205-1	99.99%	68	d 5403-0	99.55%
9	d 5401-4	100.00%	29	d 5401-3	100.00%	49	d 5100-1	99.98%	69	d 5404-1	99.44%
10	d 5403-4	100.00%	30	d 5101-3	100.00%	50	d 5701-1	99.98%	70	d 5701-3	99.41%
11	d 5201-4	100.00%	31	d 5200-4	100.00%	51	d 5700-1	99.98%	71	d 5205-2	99.39%
12	d 5202-4	100.00%	32	d 5203-3	100.00%	52	d 5501-3	99.98%	72	d 5404-0	99.23%
13	d 5501-4	100.00%	33	d 5204-3	100.00%	53	d 53000-0	99.93%	73	d 5501-2	98.75%
14	d 5205-4	100.00%	34	d 5102-2	100.00%	54	d 53010-0	99.92%	74	d 57021-0	98.51%
15	d 5602-4	100.00%	35	d 5101-2	100.00%	55	d 571-0	99.91%	75	d 57020-3	98.32%
16	d 53000-4	100.00%	36	d 5100-2	100.00%	56	d 53000-1	99.90%	76	d 5700-4	97.64%
17	d 53010-4	100.00%	37	d 5200-3	100.00%	57	d 53010-1	99.90%	77	d 5500-0	96.54%
18	d 5400-3	100.00%	38	d 5400-2	100.00%	58	d 5602-1	99.89%	78	d 57020-1	96.27%
19	d 5402-3	100.00%	39	d 5402-2	100.00%	59	d 5200-2	99.89%	79	d 5701-2	95.50%
20	d 5403-2	100.00%	40	d 53001-1	100.00%	60	d 5202-2	99.89%	80	d 5404-4	95.32%

2.4.3 Eğitim ve test verisinin oluşturulması

Algoritmaların eğitilmesi ve eğitim performansının değerlendirilmesi için veri seti, eğitim ve test verisi olarak ikiye ayrılmıştır. Eğitim ve test verisinin, veri setinin ne kadarını kapsayacağına ilişkin kesin bir oran yoktur. Bu oran veri setinin büyüklüğüne ve değişken sayısına bağlı olarak değişmektedir. Kullanıcının farklı oranları deneyerek en uygun değere ulaşması gerekmektedir. SCADI için yapılan denemelerde, veri setinin %65 ve üzerindeki oranlara sahip olan eğitim setleri ile algoritmalarda aşırı uyum (overfitting), %60'ın altındaki oranlarda ise uyumsuzluk (underfitting) sorunları ile karşılaşmıştır. Yapılan denemeler sonucunda en iyi performansın %60 oranındaki eğitim setiyle sağlandığı görülmüştür.

Kolektif öğrenme yönteminin hem farklı algoritmalar ile hem de farklı eğitim verileri ile desteklenmesi için beş farklı eğitim verisi oluşturulmuştur. Eğitim verilerinin birbirinden farklı olabilmesi için rasgele sayı üreticini başlatmak için kullanılan sayıların birbirinden farklı olmasına dikkat edilmiştir.

2.4.4 Modelin uygulanması

Sınıf tahmini için yapılan deneysel çalışmalarda yedi farklı algoritma (C5.0, CART, CHAID, QUEST, YSA, SVM, BA) kullanılmıştır. En iyi sınıflandırma performansının belirlenebilmesi için söz konusu algoritmalar tek olarak ve kolektif olarak uygulanmıştır. Kolektif modeller oluşturulurken yüksek doğruluk oranına sahip olan algoritmalar farklı birliktelikler ile deneyerek kolektif olarak çalıştırılmıştır.

3 Bulgular ve tartışma

Literatürde sınıflandırma tahmini yapılan çalışmaların performansını ölçmek için farklı parametrelerin kullanıldığı görülmüştür. Bu bölümde yedi tahmin algoritmasının ve kolektif modellerin performansını ölçmek için literatürde sıklıkla kullanılan doğruluk oranı **Denklem (28)** ve hata oranı **Denklem (29)** kullanılmıştır. Sınıflandırma doğruluğu, doğru tahmin sayısının toplam tahmin sayısına bölünmesiyle elde edilmektedir. Hata oranı ise 1 eksi hedef değişkenin tahmininde yapılan hatayı üreten doğruluk oranıdır [57].

$$\text{Doğruluk Oranı (DO)} = \frac{GP + GN}{GP + GN + YP + YN} \quad (28)$$

$$\text{Hata Oranı (HO)} = 1 - DO \quad (29)$$

Doğruluk oranı hesaplanırken kullanılan parametreler **Şekil 4**'teki hata matrisinde sunulmuştur. Oranın hesaplanmasında kullanılan eşitliğin pay kısmı doğru sınıflandırılan örnek sayısını, payda kısmı ise sınıflandırılan tüm örnek sayısını kapsamaktadır.

		Tahmin Edilen	
		Pozitif	Negatif
Gerçekleşen	Pozitif	Gerçek Pozitif (GP)	Yanlış Pozitif (YP)
	Negatif	Yanlış Negatif (YN)	Gerçek Negatif (GN)

Şekil 4. Hata matrisi.

SCADI veri setini kullanarak sınıflandırma tahmini yapan çalışmaların [13,56,58] doğruluk oranları **Tablo 6**'da sunulmuştur. Bu çalışmalar arasında en yüksek doğruluk oranına (%88,57) Keleş ve Kılıç [56], K-En Yakın Komşu algoritmasını kullanarak ulaşmıştır.

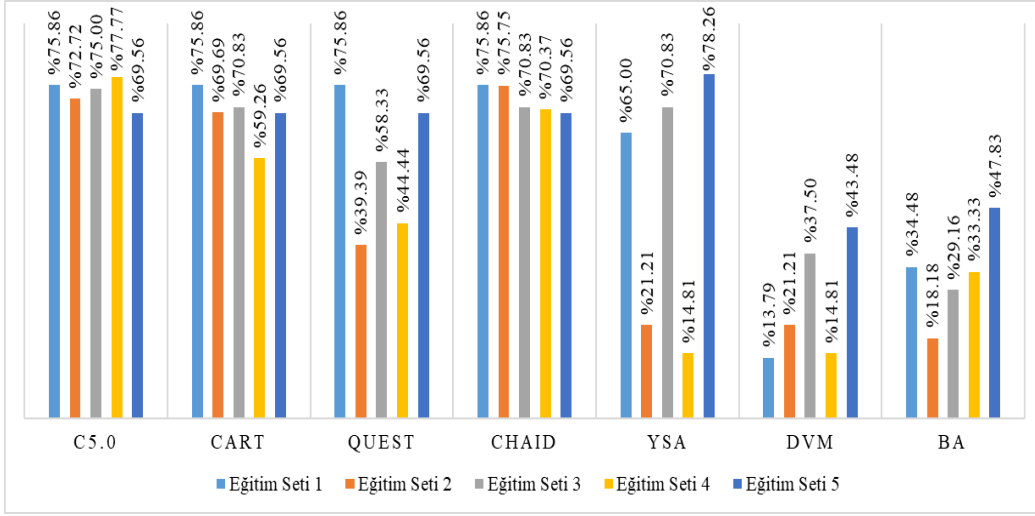
Tablo 6. SCADI ile tahmin analizi yapan çalışmalar.

	Zarchi vd. [13]	Keleş ve Kılıç [56]	Choudhury ve Greene [58]
Yapay Sinir Ağları	%83.10		
K- En Yakın Komşu		%88.57	
Rastgele Orman		%85.71	%87.75
Sıralı Minimal Optimizasyon		%87.14	
Karar Tablosu		%87.14	
Naïve Bayes		%87.14	%83.00
Destek Vektör Makineleri			%82.43
Lazy LWL			%78.57
Hoeffding Tree			%80.71

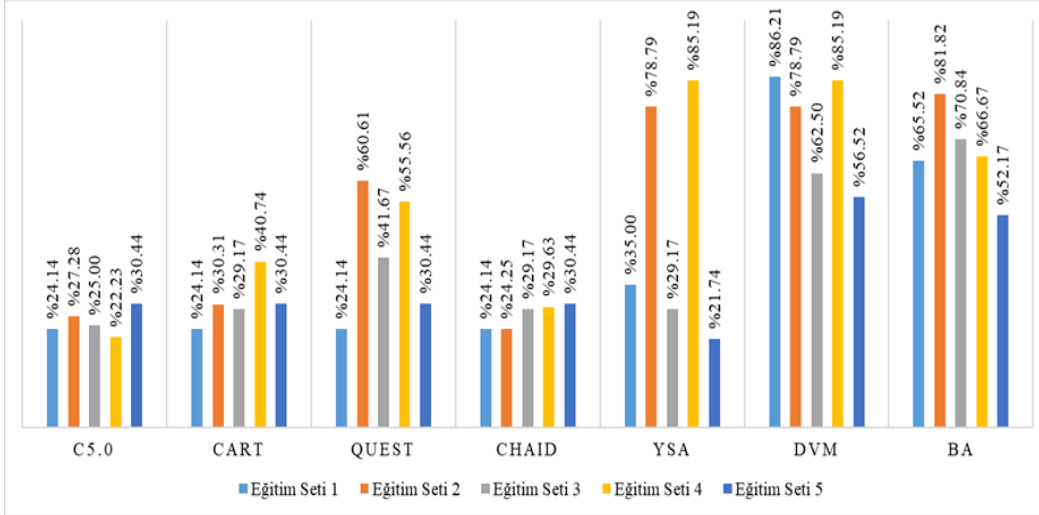
Yapılan ilk deneyde, özellik seçiminin tahmin doğruluğu üzerindeki etkisinin analiz edilmesi amaçlanmıştır. Bu amaçla özellik seçimi uygulanmadan önce veri setinden, rastgele sayılar kullanılarak birbirinden farklı beş eğitim seti elde edilmiştir. Çalışma kapsamında analiz edilecek algoritmaların her biri bireysel olarak söz konusu beş eğitim setine uygulanmıştır. **Şekil 5**'te sunulan doğruluk oranları incelendiğinde en yüksek doğruluk oranlarına eğitim seti 1 için C5.0 (%75.86), CART (%75.86), QUEST (%75.86), CHAID (%75.86) algoritmaları, eğitim seti 2 için CHAID (%75.75), eğitim seti 3 için C5.0 (%75.00), eğitim seti 4 için C5.0 (%77.77) ve eğitim seti 5 için YSA (%78.26) algoritmaları ile ulaşıldığı görülmüştür.

Daha sonra veri setine özellik seçimi uygulandıktan sonra beş farklı eğitim seti oluşturulmuş ve her bir algoritma bireysel olarak bu eğitim verilerine uygulanmıştır. En yüksek doğruluk oranlarına eğitim seti 1 için YSA (%86.20), eğitim seti 2 için CART (%85.71), eğitim seti 3 için C5.0 (%81.48), CART (%81.48), CHAID (%81.48), eğitim seti 4 için YSA (%85.71) ve eğitim seti 5 için C5.0 (%89.66), CART (%89.66), CHAID (%89.66) algoritmaları ile ulaşılmıştır. Özellik seçimi yapılmış veri seti için uygulanan bireysel algoritmaların performansları **Tablo 6**'da sunulan literatürdeki çalışmaların doğruluk oranları ile kıyaslandığında eğitim seti 5 için gerçekleştirilen deneyde daha yüksek doğruluk oranı elde edildiği görülmekle birlikte diğer eğitim setleri ile yapılan deneylerin performansı düşük çıkmıştır.

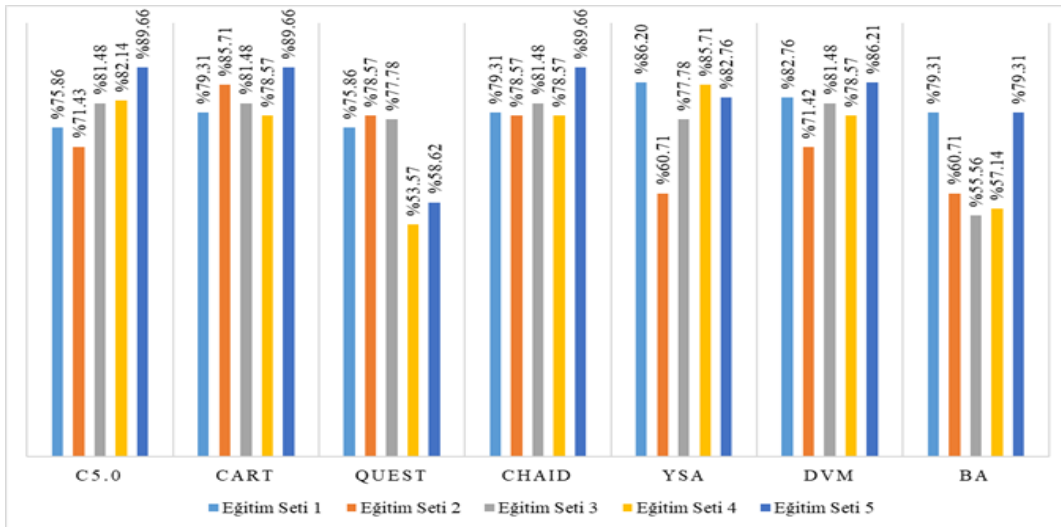
SCADI üzerinde özellik seçimi uygulanmasının deneysel sonuçlara etkisinin analiz edilmesi için her bir veri setine uygulanan yedi farklı algoritmanın doğruluk oranlarının aritmetik ortalaması alınarak **Şekil 7**'de sunulan grafik oluşturulmuştur. Grafik analiz edildiğinde özellik seçiminin eğitim setlerinin hepsi için %17.79 ile %28.50 arasında performans artışı sağladığı görülmüştür.



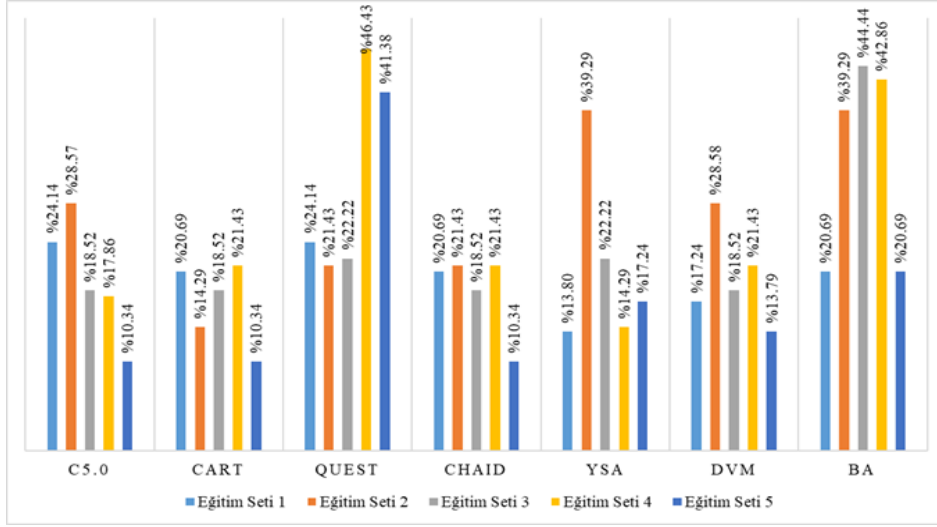
Şekil 5. Özellik seçimi yapılmadan hesaplanan doğruluk oranları.



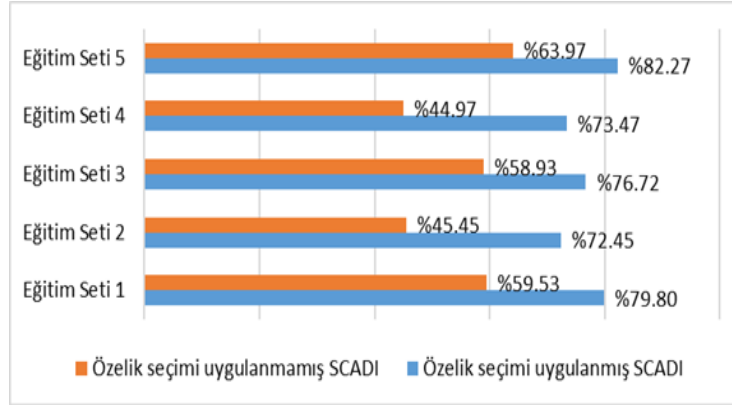
Şekil 6. Özellik seçimi yapılmadan hesaplanan hata oranları.



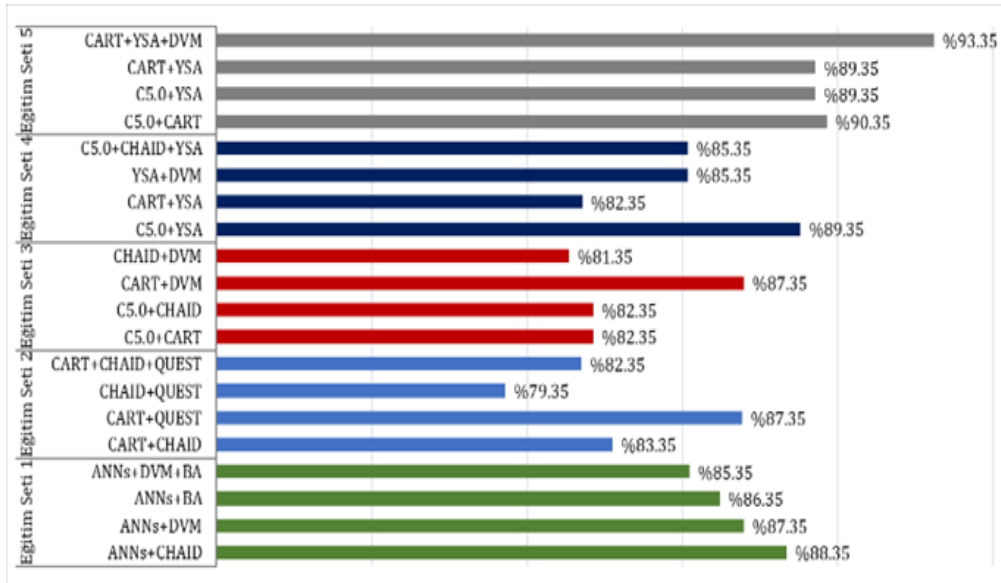
Şekil 7. Bireysel modellerin farklı eğitim verisine göre doğruluk oranları.



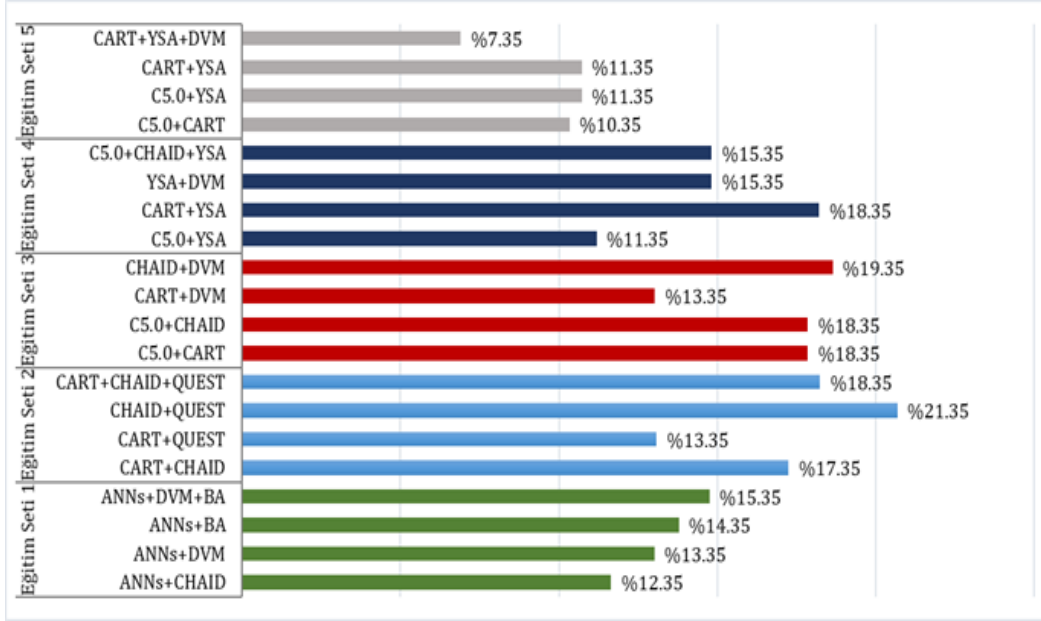
Şekil 8. Bireysel modellerin farklı eğitim verisine göre hata oranları.



Şekil 9. Özellik seçiminin sonuçlara etkisi.



Şekil 10. Kolektif modellerin farklı eğitim verisine göre doğruluk oranları.



Şekil 11. Kolektif modellerin farklı eğitim verisine göre hata oranları.

İkinci deneyde, kolektif öğrenme yönteminin algoritmaların performansını artırıp artırmadığı analiz edilmiştir. Bu amaçla, bireysel olarak uygulanan algoritmalar, çeşitli kombinasyonlar ile birlikte uygulanarak deneyler yapılmış ve performansları test edilmiştir. Yapılan deneyler arasında her bir veri seti için en iyi performansı gösteren dört kolektif modelin doğruluk oranları Şekil 10'da ve hata oranları Şekil 11'de sunulmuştur. Şekil 10'da sunulan doğruluk oranı grafiği analiz edildiğinde bütün eğitim setleri için doğruluk oranının, bireysel algoritmalar ile elde edilen doğruluk oranlarına göre arttığı görülmüştür. Doğruluk oranı artışları incelendiğinde en yüksek artış (%5.49) sağlayan kolektif model, eğitim seti 3 için hazırlanan CART+DVM modelidir. En yüksek doğruluk oranına (%93.10) sahip olan model ise eğitim seti 5 için geliştirilen CART+YSA+DVM modelidir

4 Sonuçlar

Teknolojinin gelişmesiyle bilgi oluşturma ve büyük miktarlarda heterojen verinin yönetimi veri madenciliğini önemli bir araştırma alanı haline getirmiştir. Verilerin yönetilerek kullanışlı bilgiler haline dönüştürülmesi için yaygın olarak veri madenciliği algoritmalarına başvurulmaktadır. Özellikle sağlık sektöründe yaygın olarak kullanılan sınıflandırma algoritması mevcut veriler ışığında eğitilerek, yeni veriler için öğrenilmek istenen bilgileri kısa sürede ve maliyetsiz olarak sunmaktadır.

Öz-bakım aktivitelerinin belirli bir standarda göre sınıflandırılması ve engelli bireylerin yaşadıkları sorunları teşhis ederek hangi sınıfa ait olduklarının belirlenmesi zahmetli bir süreçtir. Bu nedenle bu çalışmada uzman terapistlere bireylerin ICF-CY standartlarına göre yüksek doğrulukla sınıf tahmini yapan kolektif model önerilmiştir. Model önerilmeden önce beş farklı eğitim verisi üzerinden çeşitli deneysel çalışmalar gerçekleştirilerek kıyaslanmıştır.

SCADI veri setindeki örnek sayısı (70), değişken sayısına (205) göre oldukça az olması ve veri setinin çok sınıflı probleme yönelik oluşturulmasına rağmen her sınıf için örnek sayısının yeterli miktarda olmaması (örneğin; sınıf 1 için 2, sınıf 3 için 1 örnek, sınıf 5 için 3 örnek bulunmaktadır), sınıflandırma algoritmasının her sınıf için yüksek doğrulukla tahmin yapmasını zorlaştırmıştır ve yetersiz uyum problemi ile karşı karşıya kalınmasına neden olmuştur. Ancak bu zorluklara rağmen geliştirilen kolektif öğrenme modelleri ile gerçekleştirilen deneylerin ortalama doğruluk oranı (DO_{ort}) (%88.83) literatürde SCADI için yapılan çalışmalardan daha yüksek çıkmıştır (Denklem (30)).

$$DO = \frac{\%93,10 + \%88,80 + \%86,79 + \%86,92 + \%88,35}{5} = \%88,83 \quad (30)$$

Algoritmaların gösterdikleri performanslar analiz edildiğinde;

(i) Değişen eğitim verilerine göre modellerin öğrenme kapasitesinin de değiştiği,

(ii) Veri seti üzerinde yapılan özellik seçiminin tahmin performansını arttırdığı,

(iii) Veri madenciliği algoritmalarının kolektif olarak uygulanmasının bireysel olarak uygulanmasına kıyasla sonuçlarda iyileşme sağladığı görülmüştür.

Gelecek çalışmalarda bulanık mantık tekniğinden yararlanılarak öz bakım sorunlarının daha hassas sınıflandırmasını sağlayan algoritma geliştirilmesi planlanmaktadır. Ek olarak elde edilen karar kurallarının uzman görüşleri ile desteklenerek karar destek sistemine dönüştürülmesi ve bir web servis ile desteklenerek çevrimiçi ortamda kullanıma sunulması hedeflenmektedir.

Çıkar çatışması

Yazarlar çıkar çatışması olmadığını beyan etmektedir.

Benzerlik oranı (iThenticate): %7

Kaynaklar

- [1] World Health Organization, Disabilities, Accessed 20 November 2019. <https://www.who.int/topics/disabilities/en/>
- [2] World Health Organization (WHO). World report on disability, https://www.ncbi.nlm.nih.gov/books/nbk304079/pdf/bookshelf_nbk304079.pdf Accessed: 12 December 2019.
- [3] Türkiye İstatistik Kurumu (TÜİK). Nüfus ve konut araştırması, Erişim 31 Ocak 2013. <https://tuikweb.tuik.gov.tr/PreHaberBultenleri.do?id=15843>
- [4] J. Ditterline, D. Banner, T. Oakland, and D. Becton, Adaptive behavior profiles of students with disabilities., Journal of Applied School Psychology, 24(2), 191-208, 2008. <https://doi.org/10.1080/15377900802089973>
- [5] L. Rosenberg, A. Moran and O. Bart, The associations among motor ability, social-communication skills, and participation in daily life activities in children with low-functioning autism spectrum disorder, Journal of Occupational Therapy, Schools, & Early Intervention, 10(2), 137-46, 2017. <https://doi.org/10.1080/19411243.2017.1304842>
- [6] E. Björck-Åkesson et al., The International Classification of Functioning, Disability and Health and the version for children and youth as a tool in child habilitation/early childhood intervention–feasibility and usefulness as a common language and frame of reference for practice, Disability and Rehabilitation, 32 (1), 125-38, 2010.
- [7] R. J. Simeonsson et al., Applying the International Classification of Functioning, Disability and Health (ICF) to measure childhood disability, Disability and Rehabilitation, 25(11-12), 602-10, 2003. <https://doi.org/10.1080/0963828031000137117>
- [8] T.K. Wu, S.C. Huang, and Y.R. Meng, Evaluation of ANN and SVM classifiers as predictors to the diagnosis of students with learning disabilities, Expert Systems with Applications, 34(3), 1846-56, 2008. <https://doi.org/10.1016/j.eswa.2007.02.026>
- [9] T.K. Wu, Y.R. Meng, and S.C. Huang, Application of Artificial Neural Network to the Identification of Students with Learning Disabilities, in IC-AI , pp. 162-168, 2006.
- [10] N. Muangnak, W. Pukdee, and T. Hengsanunkun, Classification students with learning disabilities using Naïve Bayes Classifier and Decision Tree, in The 6th International Conference on Networked Computing and Advanced Information Management, pp. 189-192, 2010.
- [11] B. B. Ertürk, İşlevsellik, yetiyitimi ve sağlığın uluslararası sınıflandırılması, Erişim 20 Mayıs 2019. http://www.sosyalsiyaset.net/documents/yeti_yitimi_islevsellik.htm
- [12] N. Şahin, H. Altun, ve K. Bilge, Özürlü çocuk sağlık kurulu raporlarının değerlendirilmesi, Kocatepe Tıp Dergisi, 15(1), 48-53, 2014.
- [13] M. Zarchi, S. F. Bushehri, and M. Dehghanizadeh, SCADI: A standard dataset for self-care problems classification of children with physical and motor disability, International Journal of Medical Informatics, 114, 81-7, 2018. <https://doi.org/10.1016/j.ijmedinf.2018.03.003>
- [14] J. S. Chou and A.D. Pham Enhanced artificial intelligence for ensemble approach to predicting high performance concrete compressive strength, Construction and Building Materials, 49, 554-63, 2013. <https://doi.org/10.1016/j.conbuildmat.2013.08.078>
- [15] L. Jena and N. K. Kamila Distributed data mining classification algorithms for prediction of chronic-kidney-disease, Int. J. Emerg. Res. Manag. & Technology, 4(11), 110-8, 2015.
- [16] İ. Ertuğrul, A. Organ, ve A. Şavlı, Veri madenciliği uygulamasına ilişkin PAÜ hastanesinde hasta profilinin belirlenmesi, Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi, 19(2), 97-103, 2013. <https://doi.org/10.5505/pajes.2013.68077>
- [17] D. A. Adeniyi, Z. Wei and Y. Yongquan Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method, Applied Computing and Informatics, 12(1), 90-108, 2016. <https://doi.org/10.1016/j.aci.2014.10.001>
- [18] J. R. Quinlan, Discovering rules by induction from large collections of examples, Editör: Donald, M. Expert systems in The Micro Electronic Age, 168–201, Edinburgh, İskoçya, Edinburgh University Press, 1979.
- [19] J. R. Quinlan, Induction of decision trees, Machine Learning, 1(1), 81-106, 1986. <https://doi.org/10.1007/bf00116251>
- [20] J. R. Quinlan, Bagging, boosting, and C4. 5, in AAAI/IAAI, 1,725-30, 1996.
- [21] J. R. Quinlan, C4.5: Programs for Machine Learning San Francisco, CA, USA, Morgan Kaufmann Publishers Inc., 1993.
- [22] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, An electric energy consumer characterization framework based on data mining techniques, IEEE Transactions on Power Systems, 20(2), 596-602, 2005. <https://doi.org/10.1109/TPWRS.2005.846234>
- [23] M. B. Kılıçalan, Hanehalkı işgücü araştırma verileri ile veri madenciliği yöntemlerinin uygulanması ve modellerin karşılaştırılması, Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, Hacettepe Üniversitesi, Ankara, Türkiye, 2018.
- [24] B. Lantz, Machine learning with R. Packt Publishing Ltd, Birmingham, İngiltere, 2013.
- [25] S. L. Pang, and J. Z. Gong, C5. 0 classification algorithm and application on individual credit evaluation of banks, Systems Engineering-Theory & Practice, 29(12), 94-104, 2009. [https://doi.org/10.1016/S1874-8651\(10\)60092-0](https://doi.org/10.1016/S1874-8651(10)60092-0)
- [26] L. Breiman, J. H. Friedman and R.A. Olshen, and C.J. Stone Classification and regression trees. Belmont, CA: Wadsworth. International Group, 432, 151-66, 1984. <https://doi.org/10.1002/widm.8>

- [27] A. M. Elsayad and H. A. Elsalamony, Diagnosis of breast cancer using decision tree models and SVM, *International Journal of Computer Applications*, 83(5), 19-29, 2013.
- [28] E. Şatir, F. Azboy, A. Aydın, H. Arslan ve Ş. Hacıfendioğlu Veri indirgeme ve sınıflandırma teknikleri ile glokom hastalığı teşhisi, *El-Cezeri Journal of Science and Engineering*, 3(3), 485-97, 2016.
- [29] W. Y. Loh and Y. S. Shih, Split selection methods for classification trees, *Statistica Sinica*, pp. 815-40, 1997.
- [30] M. Coşkun and H. İ. Bülbül, Hane halkı internet hizmeti sahipliğini etkileyen faktörlerin karar ağaçları ile incelenmesi. *Türk Bilim Araştırma Vakfı*, 12(2), 1-17, 2019.
- [31] E. Ekici, Farklı sınıflandırma yöntemlerinin karşılaştırılması ve bir uygulama/An application on the comparison of various classification methods, *Yüksek Lisans Tezi, Fırat Üniversitesi, Elâzığ, Türkiye*, 2012.
- [32] G. V. Kass, An exploratory technique for investigating large quantities of categorical data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(2), 119-127, 1980. <https://doi.org/10.2307/2986296>
- [33] O. F. Althuwaynee, B. Pradhan, and S. Lee, A novel integrated model for assessing landslide susceptibility mapping using CHAID and AHP pair-wise comparison, *International Journal of Remote Sensing*, 37(5), 1190-209, 2016. <https://doi.org/10.1080/01431161.2016.1148282>
- [34] M. Chambers and T. W. Dinsmore, *Advanced Analytics Methodologies: Driving Business Value with Analytics*, Londra, Pearson Education, 2014.
- [35] M. Abdar, M. Zomorodi-Moghadam, R. Das, and I.H. Ting, Performance analysis of classification algorithms on early detection of liver disease, *Expert Systems with Applications*, 67, 239-51, 2017. <https://doi.org/10.1016/j.eswa.2016.08.065>
- [36] V. Vapnik and A. Chervonenkis, *Theory of pattern recognition*, ed: Nauka, Moscow, 1974.
- [37] C. J. Burges A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, 2(2), 121-67, 1998. <https://doi.org/10.1023/A:1009715923555>
- [38] W. S. Noble, What is a support vector machine? *Nature Biotechnology*, 24(12), 1565-7, 2006. <https://doi.org/10.1038/nbt1206-1565>
- [39] C. Cortes and V. Vapnik, Support-vector networks, *Machine Learning*, 20(3), 273-97, 1995. <https://doi.org/10.1007/BF00994018>
- [40] E. Osuna, R. Freund, and F. Girosit, Training support vector machines: an application to face detection, in *Proceedings of IEEE computer society conference on computer vision and pattern recognition*, pp. 130-6, 1997. <https://doi.org/10.1109/CVPR1997.609310>
- [41] C.C. Yeh, D.J. Chi, and M.F. Hsu, A hybrid approach of DEA, rough set and support vector machines for business failure prediction, *Expert Systems with Applications*, 37(2), 1535-41, 2010. <https://doi.org/10.1016/j.eswa.2009.06.088>
- [42] C.W. Hsu, C.C. Chang, and C.J. Lin, *A practical guide to support vector classification*, ed: Taipei, 2003.
- [43] W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys*, 5(4), 115-33, 1943. <https://doi.org/10.1007/bf02478259>
- [44] F. Rosenblatt, *The perceptron, a perceiving and recognizing automaton Project Para*, Cornell Aeronautical Laboratory, 1957.
- [45] H. Erdal, Makine öğrenmesi yöntemlerinin inşaat sektörüne katkısı: basınç dayanımı tahminlemesi, *Pamukkale University Journal of Engineering Sciences*, 21(3), 109-14, 2015. <https://doi.org/10.5505/pajes.2014.26121>
- [46] E. Aydemir, M. Karaatlı, G. Yılmaz, ve S. Aksoy, 112 Acil çağrı merkezine gelen çağrı sayılarını belirleyebilmek için bir yapay sinir ağı tahminleme modeli geliştirilmesi, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 20(5), 145-9, 2014. <https://doi.org/10.5505/pajes.2014.98608>
- [47] Ş. Hacıfendioğlu, Makine öğrenmesi yöntemleri ile glokom hastalığının teşhisi, *Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü, Konya*, 2012.
- [48] H. B. Kartal, Envanter sınıflandırmada yapay öğrenme yöntemlerinin kullanımı ve destek vektör makineleri ile bir uygulama, *Yüksek Lisans Tezi, Fen Bilimleri Enstitüsü, İstanbul Teknik Üniversitesi, İstanbul*, 2012.
- [49] M. Karabıyık ve B. Yet, Bayes ağları ile futbol analitiği: FutBA modeli, *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 25(1), 121-31, 2019.
- [50] E. N. Çinicioğlu, M. Atalay, ve H. Yorulmaz, Trafik kazaları analizi için bayes ağları modeli, *Bilişim Teknolojileri Dergisi*, 6(2), 41, 2013.
- [51] J. M. Moyano, E. L. Gibaja, K. J. Cios, and S. Ventura, Review of ensembles of multi-label classifiers: models, experimental study and prospects, *Information Fusion*, 44, 33-45, 2018. <https://doi.org/10.1016/j.inffus.2017.12.001>
- [52] T. G. Dietterich, Ensemble methods in machine learning, in *International Workshop On Multiple Classifier Systems*, 2000, Springer. https://doi.org/10.1007/3-540-45014-9_1
- [53] UCI Machine Learning Repository. SCADI Data Set, <https://archive.ics.uci.edu/ml/datasets/SCADI> Accessed: 06 December 2018.
- [54] I. A. Gheyas and L. S. Smith, Feature subset selection in large dimensionality domains, *Pattern Recognition*, 43(1), 5-13, 2010. <https://doi.org/10.1016/j.patcog.2009.06.009>
- [55] S. Gu, R. Cheng, and Y. Jin, Feature selection for high-dimensional classification using a competitive swarm optimizer, *Soft Computing*, 22(3), 811-22, 2018. <https://doi.org/10.1007/s00500-016-2385-6>
- [56] M. K. Keleş and Ü. Kılıç, Artificial bee colony algorithm for feature selection on scadi dataset, in *2018 3rd International Conference on Computer Science and Engineering (UBMK)*, pp. 463-6, 2018. <https://doi.org/10.1109/UBMK.2018.8566287>

- [57] H. B. F. David, H. B. F. and A. Suruliandi, Performance evaluation of ensemble classifiers on benchmark datasets, *Journal of Emerging Technologies and Innovative Research*, 6(2), 926-30, 2019.
- [58] A. Choudhury, Classification of functioning, disability, and health: ICF-CY self care (SCADI dataset) using

predictive analytics, in Choudhury A., *Classification of Functioning, Disability, and Health: ICF-CY Self Care.(SCADI DATASET) Using Predictive Analytics*. In: *Proceedings of the 2019 IISE Annual Conference*. Edited by HE Romeijn AS, R. Thomas. Orlando: IISE, 2019. <http://dx.doi.org/10.2139/ssrn.3307719>

