

Çocuklarda Anemi Hastalığının Teşhisinde Topluluk Öğrenme Yöntemlerinin Kullanılması

Emin Farzaliyev^{1,*}, Qusay Saihood¹, Emrullah Sonuç¹

Bilgisayar Mühendisliği, Karabük Üniversitesi, Türkiye

*İletişimden sorumlu yazar: ferzeliyevemin99@gmail.com

Özet – Okul öncesi çocukların en yaygın hastalığı, özellikle gelişmekte olan ülkelerde, anemi olarak bilinmektedir. Anemi genellikle kötü beslenme ile ilişkili olup, demografik ve sosyal faktörlerle de yakından ilişkilidir. Önceki çalışmalarda, çocuklardaki anemi hastalığını tahmin etmek ve ilişkili faktörleri belirlemek için istatistiksel yöntemler kullanılmıştır. Ancak, bu yöntemlerin yetersiz olduğu sonucuna varılmıştır. Bu çalışmada, çocuklarda anemi hastalığını tahmin etmek için topluluk öğrenme (ensemble learning) tekniklerinin kullanımı araştırılmıştır. Anemi hastalığını tahmin etmek için Karar Ağacı, Destek Vektör Makineleri, Rastgele Orman, Lojistik Regresyon, En Yakın Komşu gibi çeşitli makine öğrenmesi algoritmaları test edilmiştir. Daha sonra bu sınıflandırıcılar, torbalama (bagging), artırma (boosting), istifleme (stacking) gibi öğrenme teknikleri ile modellenerek daha doğru ve güçlü bir tahmin modelinin oluşturulması amaçlanmıştır. Bu çalışmada kullanılan veri seti, Irak'ın Haditha Genel Hastanesi ve kliniklerinden toplanan 600 örneği içermektedir. Bu örneklerden 429'u anemi hastası iken, 171'i anemi hastası değildir. İlgili veri setinde her bir örneğe ait 31 özellik bulunmaktadır. Veri seti üzerinde farklı topluluk tekniklerinin performansları değerlendirilmiştir. Sonuçlara göre topluluk öğrenme teknikleri bireysel sınıflandırıcılara göre daha az doğrulukla tahminde bulunmuştur. Ayrıca, topluluk öğrenme teknikleri arasında artırma yönteminin en yüksek doğruluğa (%91) eriştiği görülmüştür. Çalışma, topluluk öğrenme tekniklerinin çocuklarda anemi hastalığını tahmin etmek için farklı bir yöntem olabileceğini göstermektedir. Ancak, gelecekteki araştırmalarda veri ön işleme, özellik seçimi gibi yöntemlerin kullanılmasının topluluk öğrenme modellerinin performansını artırmada etkili olabileceği öngörülmektedir.

Anahtar Kelimeler – Makine Öğrenmesi, Topluluk Öğrenmesi, Anemi Teşhisi, Sınıflandırma

I. GİRİŞ

Anemi hastalığı, kırmızı kan hücrelerinin sayısının azalması veya hemoglobin proteininin normal seviyenin altına düşmesi durumunda ortaya çıkan bir hastalıktır. Aneminin üç önemli nedeni vardır: kan kaybı, yetersiz beslenme ve vücudun kırmızı kan hücrelerini üretme yeteneğindeki bozukluklar. Anemi belirtileri arasında halsizlik, çabuk yorulma, uyku isteği ve soluk cilt yer alır.

Özellikle okul çağı altındaki çocukların anemi olma olasılığı daha yüksektir. Bu yaş grubundaki çocukların %42'sinin anemi hastalığına sahip olduğu tahmin edilmektedir. Aneminin birkaç çeşidi vardır, ancak özellikle beş yaşın altındaki çocuklar

arasında demir eksikliği anemisi en yaygın anemi türlerinden biridir [1], [2].

Hastalık tahmininde ve tedavisinde, makine öğrenmesi yöntemleri kullanılabilir. Bu yöntemlerin avantajları arasında şunlar yer alır:

- Verilerin daha hızlı ve daha doğru bir şekilde analiz edilmesi;
- Verilerin daha iyi anlaşılması ve daha iyi bir şekilde yorumlanması;
- Hastalıkların daha erken teşhis edilmesi ve tedavi edilmesi;
- Hastalıkların daha iyi bir şekilde takip edilmesi ve yönetilmesi;

Makine öğrenmesi yöntemleri, hastalık tahmini yapmak için kullanılabilir birçok farklı algoritma içerir. Bu algoritmaların her biri farklı özelliklere sahip olup, hangisinin kullanılacağına karar vermek için veri setinin boyutu, verilerin özellikleri ve hastalığın türü gibi faktörler göz önünde bulundurulmalıdır. Bu algoritmalar arasında en yaygın olanları şunlardır:

- Rastgele ormanlar,
- Destek vektör makineleri,
- Karar ağaçları,
- Lojistik regresyon,
- Yapay Sinir Ağları.

Çalışmanın ikinci bölümünde konuyla ilgili çalışmalar yer almaktadır. Bir sonraki bölümde hastalığın tahmini için kullanılan makine öğrenme ve topluluk öğrenme yöntemlerinden ve kullanılan veri setinden bahs edilmiştir. Dördüncü bölümde kullanılan yöntemlerin sonuçları karşılaştırılmıştır. Son bölüm ise sonuçların tartışmasını içermektedir.

II. İLGİLİ ÇALIŞMALAR

Sağlık alanı, günümüzde yapay zeka ve onun alt dalları olan veri madenciliği ve makine öğrenmesi yöntemleriyle en fazla araştırma ve uygulama yapılan alanlarından biridir [3]. Yapay zeka teknikleri son zamanlarda sağlık sektöründe hastalıkların teşhisi konusunda büyük başarı elde etmiştir [4]. Bu bölümde, hastalık teşhisinde son üç yılda yapay zeka tekniklerinin kullanımına yönelik literatür taraması yapılmaktadır.

Saygın ve Baykara “Karaciğer Yetmezliği Teşhisinde Özellik Seçimi Kullanarak Makine Öğrenmesi Yöntemlerinin Başarılarının Ölçülmesi” başlıklı yaptıkları çalışmada, sarmal (wrapper) özellik seçimi yöntemlerinden İleri Yönde Seçim yöntemi (Sequential Forward Selection) kullanılarak karaciğer hastalıklarının erken teşhisinde kullanılacak en önemli özellikler belirlenmiş. Makine öğrenmesi teknikleri kullanılarak, önemli olarak belirlenen özellikler üzerinde yapılan çalışmalar, karaciğer yetmezliği hastalığının teşhis doğruluğunu ölçmek için gerçekleştirilmiş. Hafif Gradyan Güçlendirme Makinesi Sınıflandırıcısı (HGGM), Karar Ağacı (KA), Çok Katmanlı Algılayıcılar (ÇKA), Destek Vektör Makineleri (DVM) ve Lojistik Regresyon (LR) yöntemleri, karaciğer yetmezliği hastalığının

tahmininde kullanılmıştır. Kullanılan yöntemlerin sırasıyla elde ettikleri teşhis doğruluğu şu şekildedir: LR; %77.80, DVM; %77.87, ÇKA; %81.13, KA; %81.13 ve HGGM; %82.12 [5].

Bilgin 2021 yılında yayınlanan “Makine Öğrenmesi Algoritmaları Kullanarak Erken Dönemde Diyabet Hastalığı Riskinin Araştırılması” adlı çalışmasında, 520 kişinin verilerinden oluşan bir veritabanı kullanılmış ve diyabete yakalanma olasılığının tahmini için makine öğrenmesi yöntemleri kullanılmış. Çalışmada, makine öğrenmesi yöntemlerinden K-En Yakın Komşu (KEYK), KA, ÇKA, Doğrusal Diskriminant Analizi (DDA), Topluluk Öğrenme Algoritmaları (TÖA) ve Destek Vektör Makinaları (DVM) yöntemleri kullanılmış. Denenmiş yöntemlerden KEYK algoritması en yüksek doğruluğu %99,81 elde etmiştir [6].

Coşar ve Deniz 2021 yılında yayınlanan “Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi” adlı çalışmalarında, farklı makine öğrenmesi teknikleri kullanarak kalp rahatsızlıklarının tespiti için bir örnek veri kümesi üzerinde çalışmalar yapmışlardır. Üç farklı makine öğrenmesi yöntemi kullanılarak örnek bir model oluşturulmuştur ve kalp rahatsızlığı olan bireyler tespit edilmiştir. Veri seti incelenmiş ve kalp rahatsızlığına dair belirtilerin hangi verilerde yer alabileceği belirlenmiştir. Karşılaştırılan sonuçlara göre, Rastgele Orman(RO) algoritması %88'lik bir doğruluk oranı elde etmiştir. LR %85 doğruluk oranıyla ikinci sırada yer alırken, KEYK algoritması %70 doğruluk oranıyla üçüncü sırada yer almış [7].

III. MATERYAL VE YÖNTEM

A. Veri Seti

Bu çalışmada kullanılan veri seti 600 örneği içermektedir [1-2]. Bu örneklerden 429'u anemi hastası iken, 171'i anemi hastası değildir. Verilerin %80'i eğitim seti olarak, %20'i ise test seti olarak kullanılmıştır. İlgili veri setinde her bir örneğe ait 31 özellik bulunmaktadır. Veri setinde kullanılan tüm özellikler Tablo 1'de açıklanmıştır.

Tablo 1. Veri seti özelliklerinin açıklaması.

No	Özellik adı	Türü	Açıklama
1	Id	Sayısal	Birincil anahtar 600 satırdan oluşur
2	Çocuğun yaşı	Sayısal	6 aydan 6 yaşa kadar
3	Cinsiyet	Sayısal	0 = Kadın, 1 = Erkek
4	Annenin yaşı	Sayısal	0 = Yaş < 30, 1 = Yaş ≥ 30
5	Annenin eğitim düzeyi	Sayısal	0 = Okur yazarlık yok, 1 = İlkokul, 2 = Ortaokul, 3 = Üniversite ve üzeri
6	Annenin mesleki düzeyi	Sayısal	0 = İşsiz, 1 = Çalışan
7	Babanın eğitim düzeyi	Sayısal	0 = Okur yazarlık yok, 1 = İlkokul, 2 = Ortaokul, 3 = Üniversite ve üzeri
8	Babanın mesleki düzeyi	Sayısal	0 = İşsiz, 1 = Çalışan
9	Konut	Sayısal	0 = kırsal, 1 = kentsel
10	Sosyo-ekonomik durum	Sayısal	0 = Yoksul, 1 = Orta düzey, 2 = İyi
11	Hemogloblin seviyesi	Sayısal	
12	Yaşına göre kısa boy	Sayısal	0 = Anormal, 1 = Normal
13	Son 15 gün içinde ateş	Sayısal	0 = Hayır, 1 = Evet
14	Anemiye ilişkin önceki tıbbi geçmiş.	Sayısal	0 = Hayır, 1 = Evet
15	Son 15 gün içinde ishal	Sayısal	0 = Hayır, 1 = Evet
16	Emzirme türü	Sayısal	0 = Anormal, 1 = Normal, 3 = Maksimum
17	Süt tozu tüketim	Sayısal	0 = Hayır, 1 = Evet
18	Şekerli içecek tüketim	Sayısal	0 = Hayır, 1 = Evet
19	Yoğurt tüketim	Sayısal	0 = Hayır, 1 = Evet
20	Katı/yarı katı gıda tüketim	Sayısal	0 = Hayır, 1 = Evet

21	Emzirme süresi	Sayısal	0 = Anormal, 1 = Normal
22	Et tüketimi	Sayısal	0 = Hayır, 1 = Evet
23	Koyu yeşil yapraklı sebzelerin tüketimi.	Sayısal	0 = Hayır, 1 = Evet
24	Demir kaynağı gıdaların tüketimi.	Sayısal	0 = Hayır, 1 = Evet
25	Karaciğer tüketimi.	Sayısal	0 = Hayır, 1 = Evet
26	Tamamlayıcı beslenmenin optimal zamanını bilmek.	Sayısal	0 = Bilinmiyor, 1 = Biliniyor, 2 = Emin Değilim
27	İlk tamamlayıcı gıdayı bilmek.	Sayısal	0 = Bilinmiyor, 1 = Biliniyor, 2 = Emin Değilim
28	Takviye demirin optimal gıdasını bilmek.	Sayısal	0 = Bilinmiyor, 1 = Biliniyor, 2 = Emin Değilim
29	Anemi ile ilişkili besinleri bilmek.	Sayısal	0 = Bilinmiyor, 1 = Biliniyor, 2 = Emin Değilim
30	Emzirmenin optimal zamanını bilmek.	Sayısal	0 = Bilinmiyor, 1 = Biliniyor, 2 = Emin Değilim
31	Anemi durumu	Sayısal	0 = Anemi Hastası Değil (171 örnek), 1 = Anemi hastası (429 örnek)

B. Veri ön işleme

Yapay zeka algoritmalarının doğru çalışmasını olumsuz etkileyen ve düşük performans seviyesine neden olan birçok yaygın sorun vardır. Bu sorunlar arasında kayıp gözlem, sınıf dengesizliği, ilgisiz değişken ve aykırı gözlem gibi sorunlar yer almaktadır [8]. Veri ön işleme, makine öğrenime teknikleri için önemli bir aşamadır. Bu algoritmaların temel hedefi, veriden bilgi elde etmektir ve bunu hesaplama yöntemlerini kullanarak yapmaktalar [9]. Makine öğrenmesinde

veri önışleme, verilerin doğru bir şekilde analiz edilebilmesi için hazırlanmasıdır. Veri setinde tutarsız, gürültülü, eksik ve ya gereksiz veriler olabilir. Bu durum, algoritmaların performansını olumsuz yönde etkiler ve yanlış öğrenilen bilgi yanlış sonuçlar verebilir. Veri önışleme, verilerin temizlenmesi, dönüştürülmesi ve özellik çıkarımı gibi işlemleri içerir. Bu işlemler, makine öğrenmesi algoritmalarının daha doğru ve güvenilir sonuçlar üretmesine yardımcı olur.

Veri temizleme işlemi, verilerdeki eksik veya yanlış bilgilerin düzeltilmesini içerir. Bu işlem, veri setlerindeki hataları ortadan kaldırarak, verilerin doğru bir şekilde analiz edilebilmesi ve daha doğru sonuçlar elde etmek için yapılan bir veri ön işleme adımındır. Veri temizleme işlemi, özellikle büyük veri setlerinde önemli bir rol oynar çünkü büyük veri setleri, birçok hata ve bozuk veri içerebilir. Veri temizleme, veri setindeki hatalı veya bozuk verileri belirlemek için bir dizi teknik kullanır. Bunlar arasında manuel veri incelemesi, istatistiksel yöntemler, makine öğrenimi ve yapay zeka teknikleri yer alır. Veri temizleme işlemi sırasında yapılacak işlemler şunlardır :

- Eksik verilerin tamamlanması;
- Yanlış verilerin düzeltilmesi;
- Aykırı verilerin tespit edilmesi ve düzeltilmesi;
- Verilerin birleştirilmesi.

Veri dönüştürme, bir veri kümesini bir formattan başka bir formata dönüştürme sürecidir. Bu işlem, verilerin bir sistemden diğerine taşınması, işlenmesi veya saklanması sırasında gereklidir. Verileri hedef sistem tarafından kabul edilen bir formata dönüştürerek farklı sistemler arasında uyumluluğu sağlar, veri kaynakları arasındaki farklılıkların giderilmesine yardımcı olur, veri bütünlüğünü korur ve verilerin okunabilirliğini artırarak insanların anlaması için daha kolay hale getirebilir. Birçok farklı işlemde kullanılabilir. Örneğin veri entegrasyonu, veri madenciliği, veri analizi, raporlama ve iş zekası gibi alanlarda sıkça kullanılmaktadır. Veri dönüştürme işlemi, manuel olarak yapılabileceği gibi otomatik olarak da gerçekleştirilebilir. Otomatik veri dönüştürme işlemleri, daha hızlı ve daha doğru sonuçlar sağlar ve insan hatalarını azaltır.

Veri normalleştirme, bir veritabanındaki verilerin düzenlenmesidir. Veritabanı tasarımı sırasında aynı verinin farklı yerlerde tekrarlanma ihtimali mevcuttur. Bu durumda, veri eklemek, silmek veya güncellemek gibi işlemlerde problemler oluşabilir. Normalleştirme işlemi, verilerin tekrarlanmasını önleyerek veritabanının daha az yer kaplamasını sağlar. Veritabanı normalleştirme, veritabanını yapılandırma ve organize etme sürecidir, veri tekrarını azaltmak için kullanılır.

Makine öğrenimi problemlerinde, öznelik seçimi kabul edilebilir bir sınıflandırma doğruluğunu elde etmek için önemli bir adımdır. Bu adım aynı zamanda büyük ölçekli bir veri kümesinde yer alan gereksiz, alakasız ve gürültülü özneliklerin elenmesini de içermektedir. Öznelik seçimi, sınıflandırıcının doğruluğunu arttırmak ve ilgili hesaplama maliyetlerini en aza indirmek için kullanılmaktadır.

C. Sınıflandırma Yöntemleri

Sınıflandırma yöntemi, veri madenciliğinin öğrenme algoritmalarına dayanan temel yöntemlerinden biridir [10]. Makine öğrenmede iki tür öğrenme vardır: denetimli ve denetimsiz öğrenme. Denetimli öğrenmede, algoritmalar girdileri (özellikler) ve çıktıları (hedefler) kullanarak bir modeli eğitirler ve yeni veri girdileri için makul tahminler oluşturmayı amaçlarlar. Denetimsiz öğrenmede ise algoritmalar sadece girdileri kullanarak bir modeli eğitirler ve çıktıları belirtmezler. En önemli ve yaygın olarak kullanılan denetimli makine öğrenimi tekniklerinden biri sınıflandırmadır. Çıktıların sayısına bağlı olarak iki tür sınıflandırma vardır: ikili ve çoklu sınıflandırma. Bu çalışmada, çocuklarda anemi durumunu (anemik veya anemik olmayan) tahmin ettiğimiz için ikili sınıflandırma kullanılmıştır.

Bu çalışmada Anemi hastalığını tahmin etmek için KA, DVM, RO, LR, KEYK, LR, ÇKA gibi çeşitli makine öğrenmesi algoritmaları ve torbalama (bagging), artırma (boosting), istifleme (stacking) olmak üzere üç farklı topluluk öğrenme teknikleri kullanılmıştır.

Lojistik Regresyon (LR): LR, sınıflandırma problemlerinde kullanılan bir öğrenme yöntemidir. Bu yöntem, çıkış değerlerinin sınırlı olmasını

sağlamak için tahmini etiket değerlerini lojistik bir fonksiyondan geçirir. LR, günümüzde sıkça kullanılan ve bu sorunu çözmek için önerilen bir öğrenme yöntemidir.

LR'nin çalışma prensibi şu şekildedir: Veri kümesindeki her bir örnek, bir veya daha fazla öznitelikle temsil edilir. Bu özniteliklerin ağırlıklarını ve bir eşik değerini hesaplar. Bu ağırlıklar ve eşik değeri, veri kümesindeki her bir örneğin sınıflandırılmasında kullanılır. Amaç, doğru sınıflandırma yapmak için en uygun ağırlıkları ve eşik değerini bulmaktır. Sınıflandırma işlemi sırasında, algoritma bağımsız ve bağımlı değişkenler arasındaki ilişkiyi analiz eder [11].

Rastgele Orman (RO): RO, makine öğrenmesinde sınıflandırma ve regresyon problemlerinde sıkça kullanılan bir yöntemdir. Bu yöntem, birden fazla karar ağacını bir araya getirerek bir tahmin oluşturur. Her bir RO, rastgele seçilen alt küme verileri üzerinde eğitilir. Bu sayede, her ağaç kendine özgü öznitelikler üzerinde eğitilir ve farklı sonuçlar elde eder. RO, her bir ağacın sonucunu toplayarak ortalamasını alarak daha güvenilir bir tahmin elde eder. Yöntem, tek bir ağaç kullanıldığında beliren aşırı öğrenme sorununu halletmek için geliştirilmiştir [12]. Bu yöntem, yüksek boyutlu verilerde iyi performans gösterir, verilerdeki gürültüyü azaltabilir, yüksek doğrulukla çalışır ve çeşitli uygulama alanlarında kullanılabilir.

K-En Yakın Komşu (KEYK): KEYK, makine öğrenmesinde yaygın olarak kullanılan sınıflandırma yöntemlerinden biridir. Bu yöntemde, yeni veri örneği, en yakın komşularından etkilenir ve KEYK yöntemi öğrenme verilerindeki örneklerin uzaklığına dayalı olarak tahmin veya sınıflandırma yapar. Bu yöntem, basit ama etkili bir yöntemdir. Ancak, yüksek boyutlu verilerde verimlilik sorunları ortaya çıkabilir. KEYK yönteminin avantajları arasında eğitim verilerindeki gürültüye dirençli olması, yeni veri örnekleri eklenirken yeniden eğitim gerektirmemesi ve kolay anlaşılır olması yer alır. Dezavantajları ise, veri setinin büyüklüğüne bağlı olarak işlem yükü ve maliyetin artmasıdır [13]. Bu yöntem, sınıflandırma ve örüntü tanıma gibi birçok uygulama alanında kullanılır ve birçok makine öğrenmesi algoritmasına temel oluşturur.

Karar Ağacı (KA): KA yöntemi, makine öğrenmesinde kullanılan bir yöntemdir ve veri setindeki özellikleri bir ağaç yapısı şeklinde sıralar. Her bir düğüm, bir özellik seçer ve buna göre veri setini bölerek işleme devam eder. Bu işlem, bölümlerin sonunda elde edilen alt kümeler için de tekrarlanır ve sonunda sınıflandırma ya da tahmin yapılır. KA yöntemi, yüksek açıklanabilirliği ve yorumlanabilirliği sayesinde tercih edilir. Ayrıca, veri ön işlemi gerektirmez ve yüksek boyutlu verilerde kullanılabilir. Ancak, aşırı uyum sorunu oluşabilir ve bazı durumlarda diğer yöntemlere kıyasla daha az doğru sonuçlar verebilir.

Destek Vektör Makineleri (DVM): DVM, büyük veri setlerinin olduğu sınıflandırma problemlerinde sıklıkla kullanılan bir ayırım metodudur [14]. Sınıflandırma yapmak için farklı sınıflara ait verileri birbirinden ayıran bir hiper düzlem oluşturur. Bu hiper düzlem, maksimum marjinal ayırım yapacak şekilde optimize edilir. DVM, yüksek boyutlu verilerde de etkili bir şekilde çalışabilir ve aşırı uyuma karşı dirençlidir. Ayrıca, DVM, doğrusal veya doğrusal olmayan ayırım için farklı çekirdek fonksiyonları kullanılabilir. Bu yöntem, verilerdeki gürültüye ve eksik verilere karşı da dirençli olabilir. Bununla birlikte, DVM, büyük veri setlerinde eğitim süresi uzun olabilir ve ölçeklendirme sorunlarına dikkat edilmesi gerekir.

Torbalama: Topluluk öğrenme yöntemi olan torbalama, makine öğrenmesinde sınıflandırma ve regresyon problemlerini çözmek için kullanılan etkili bir tekniktir. Bootstrap Aggregating kısaltması Bagging olarak ifade edilir ve Bootstrap örnekleme yöntemini esas almaktadır [15]. Bu yöntem, birden fazla zayıf öğrencinin bir araya getirilerek güçlü bir öğrenci oluşturulmasını sağlar. Torbalama yöntemi, her zayıf öğrencinin kendi küçük veri setinde eğitilmesini ve sonunda tüm zayıf öğrencilerin tahminlerinin toplanarak bir tahmin vermesini içerir. Bu teknik, yüksek varyanslı modellerde kullanılmak üzere özellikle uygundur ve aşırı uyum problemini azaltmak için etkilidir. Ayrıca torbalama yöntemi, gürültüye dayanıklıdır ve yüksek boyutlu verilerde genellikle tercih edilir. Ancak, hesaplama maliyeti yüksek olabilir ve çok sayıda zayıf öğrenci kullanıldığında yetersiz bellek problemleri ortaya çıkabilir.

Artırma: Makine öğrenmesinde sıkça kullanılan bir teknik olan Artırma (Boosting) topluluk öğrenmesi, birden fazla öğrenme algoritmasının bir araya getirilmesiyle daha doğru sonuçlar elde etmeyi amaçlar. Artırma, birden fazla modelin sıralı bir şekilde eğitilmesiyle her bir modelin önceki modelin hatalarından öğrenerek daha doğru sonuçlar üretmeyi hedefleyen bir yöntemdir. Bu yöntemde, algoritmalar eğitim verilerine göre kendi modellerini oluştururlar ve bu modeller birleştirilerek sonuçlar üretilir. Böylece, her bir algoritmanın güçlü yönlerinden faydalanılır ve zayıf yönleri telafi edilir. Artırma, modelin yanlılığını azaltarak yetersiz uyum problemini önlemeye yardımcı olur.

İstifleme: İstifleme (Stacking) topluluk öğrenmesi, yapay zeka ve makine öğrenmesi alanlarında sıklıkla kullanılan bir tekniktir. Bu yöntem, karmaşık sorunların çözümü için idealdir ve farklı model tiplerinin bir araya getirilerek özelleştirilmiş çözümler oluşturulmasına olanak sağlar. İstifleme, çeşitli model türlerinin tahminlerini meta-öğrenici için girdi olarak kullanarak daha yüksek performanslı bir tahmin üretir [16]. Topluluk içindeki her model birbirini tamamlar ve sonraki modelin çıktısı, önceki modelin girdisi olarak kullanılır. Böylece, modeller birbirleriyle etkileşim içinde çalışarak daha doğru sonuçlar üretebilirler. İstifleme topluluk öğrenmesi, büyük veri kümeleriyle çalışırken özellikle yararlıdır ve gürültüyü azaltmak ve doğru sonuçları elde etmek için kullanılabilir. Bu teknik, aşırı uyum ve yetersiz uyum problemlerini önlemeye yardımcı olarak modelin yanlılığını ve varyansını azaltır.

D. Performans ölçümü

Sınıflandırma tekniklerinin performansı, doğruluk, duyarlılık ve kesinlik gibi üç kalite ölçütü kullanılarak değerlendirildi. Anemi hastası olan örnekler "1" olarak pozitif kabul edilirken, anemi hastası olmayan örnekler "0" olarak negatif kabul edilmiştir.

- Gerçek Pozitifler (GP): Anemik çocuklar, anemik olarak tahmin edilen çocuklar.
- Gerçek Negatifler (GN): Anemik olmayan çocuklar, anemik olmayan olarak tahmin edilen çocuklar.

- Yanlış Pozitifler (YP): Anemik olmayan çocuklar, anemik olarak tahmin edilen çocuklar.
- Yanlış Negatifler (YN): Anemik çocuklar, anemik olmayan olarak tahmin edilen çocuklar.

$$\text{Doğruluk} = (\text{GN} + \text{GP}) / (\text{GN} + \text{GP} + \text{YN} + \text{YP})$$

$$\text{Kesinlik} = \text{GP} / (\text{GP} + \text{YP})$$

$$\text{Duyarlılık} = \text{GP} / (\text{GP} + \text{YN}) \quad [17]$$

IV. DENEYSEL SONUÇLAR

Veri seti üzerinde farklı topluluk tekniklerinin performansları değerlendirilmiştir. Tablo 2'deki sonuçlara göre topluluk öğrenme teknikleri bireysel sınıflandırıcılara göre daha az doğrulukla tahminde bulunmuştur. Ayrıca, topluluk öğrenme teknikleri arasında Artırma yönteminin en yüksek doğruluğa (%91) eriştiği görülmüştür.

Sonuçlar incelendiğinde bireysel sınıflandırıcıların daha doğru bir tahmin yaptığı açıkça görülmektedir. Bunun başlıca nedenlerinden birisi, veri setinin eğitim kısmında öğrenmeyi gerçekleştirirken RO gibi algoritmaların önemli özellikleri tespit etmesi gösterilebilir.

Tablo 2. Sınıflandırma yöntemlerinin sonuçlarının karşılaştırılması.

Sınıflandırma Modeli	Duyarlılık	Kesinlik	Doğruluk
Lojistik Regresyon	0.90	0.90	0.90
Rastgele Orman	0.94	0.95	0.94
K-En Yakın Komşu	0.75	0.73	0.75
Karar Ağacı	0.98	0.98	0.98
Destek Vektör Makineleri	0.84	0.87	0.84
Torbalama	0.86	0.88	0.86
İstifleme	0.83	0.85	0.83
Artırma	0.91	0.92	0.91

V. SONUÇ

Bu çalışmada, beş farklı makine öğrenmesi tekniği (Karar Ağacı, Destek Vektör Makineleri, Rastgele Orman, Lojistik Regresyon, En Yakın Komşu) ve üç farklı topluluk öğrenme yöntemi (torbalama, artırma, istifleme) çocuklarda anemi hastalığını tahmin etmek için analiz edilmiştir. Makine öğrenme yöntemlerinden Karar Ağacı

yöntemi 98% doğrulukla en yüksek sonucu, K- En Yakın Komşu yöntemiyle 75% doğrulukla en düşük sonucu elde etmiştir. Diğer taraftan, topluluk öğrenme yöntemlerinden Artırma yöntemi 91% doğrulukla en yüksek sonucu, İstifleme yöntemiyle 83% doğrulukla en düşük sonuca ulaşmıştır. Gelecekte veri ön işleme adımlarında farklı teknikler kullanılarak topluluk öğrenme yöntemlerinin başarısı artırılabilir.

KAYNAKLAR

- [1] Q. Saihood and E. Sonuç, "The Efficiency of Classification Techniques in Predicting Anemia Among Children: A Comparative Study," in *Communications in Computer and Information Science*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 167–181. doi: 10.1007/978-3-030-97255-4_12.
- [2] Q. T. Luay SAIHOOD, "EXPLORATION OF MACHINE LEARNING TECHNIQUES IN PREDICTING THE CHILDHOOD ANEMIA," 2021.
- [3] İ. E. EMRE, C. TAŞ, and Ç. EROL, "Psikiyatride Makine Öğrenmesi Yöntemlerinin Kullanımı," *Psikiyatride Guncel Yaklaşımlar - Current Approaches in Psychiatry*, vol. 13, no. 2, pp. 332–353, Jun. 2021, doi: 10.18863/pgy.779987.
- [4] M. TEKE, "Prediction of Liver Diseases with Machine Learning Method," *Sürdürülebilir Mühendislik Uygulamaları ve Teknolojik Gelişmeler Dergisi*, Jun. 2022, doi: 10.51764/smutgd.1106793.
- [5] E. SAYGIN and M. BAYKARA, "Karaciğer Yetmezliği Teşhisinde Özellik Seçimi Kullanarak Makine Öğrenmesi Yöntemlerinin Başarılarının Ölçülmesi," *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, Mar. 2021, doi: 10.35234/fumbd.832264.
- [6] G. BİLGİN, "Investigation of The Risk of Diabetes in Early Period using Machine Learning Algorithms," *Journal of Intelligent Systems: Theory and Applications*, vol. 4, no. 1, pp. 55–64, Mar. 2021, doi: 10.38016/jista.877292.
- [7] M. COŞAR and E. DENİZ, "Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi," *European Journal of Science and Technology*, Oct. 2021, doi: 10.31590/ejosat.1012986.
- [8] A. S. Mert Demirarslan, "Rutin kan testleriyle COVID-19 tanı tahmininde makine öğrenmesi yöntemleriyle bir mobil uygulama geliştirilmesi," 2021.
- [9] K. KARADAĞ, "KAN VERMEYE ELVERİŞLİ DONÖRLERİN MAKİNE ÖĞRENME YÖNTEMLERİ İLE TESPİTİ," *Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi*, Oct. 2021, doi: 10.54365/adyumbd.993772.
- [10] B. ÖZLÜER BAŞER, M. YANGIN, and E. S. SARIDAŞ, "Makine Öğrenmesi Teknikleriyle Diyabet Hastalığının Sınıflandırılması," *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, Feb. 2021, doi: 10.19113/sdufenbed.842460.
- [11] G. KABA and S. BAĞDATLI KALKAN, "KARDİYOYOVASKÜLER HASTALIK TAHMİNİNDE MAKİNE ÖĞRENMESİ SINIFLANDIRMA ALGORİTMALARININ KARŞILAŞTIRILMASI," *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, Sep. 2022, doi: 10.55071/ticaretfd.1145660.
- [12] K. AKYOL and A. KARACI, "Diyabet Hastalığının Erken Aşamada Tahmin Edilmesi İçin Makine Öğrenme Algoritmalarının Performanslarının Karşılaştırılması," *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, pp. 123–134, Dec. 2021, doi: 10.29130/dubited.1014508.
- [13] G. AKGÜL, A. A. ÇELİK, Z. ERGÜL AYDIN, and Z. KAMIŞLI ÖZTÜRK, "Hipotiroidi Hastalığı Teşhisinde Sınıflandırma Algoritmalarının Kullanımı," *Bilişim Teknolojileri Dergisi*, vol. 13, no. 3, pp. 255–268, Aug. 2020, doi: 10.17671/gazibtd.710728.
- [14] G. BİLGİN and A. ÇİFCİ, "Eritematöz Skuamöz Hastalıkların Teşhisinde Makine Öğrenme Algoritmalarının Etkisi," *Journal of Intelligent Systems: Theory and Applications*, vol. 4, no. 2, pp. 195–202, Sep. 2021, doi: 10.38016/jista.901670.
- [15] Adem DOĞANER and C. Çolak, "TOPLULUK ÖĞRENME YÖNTEMLERİ İLE RENAL HÜCRELİ KARSİNOM'UN TAHMİN EDİLMESİ," 2020.
- [16] Mehmet KIVRAK and C. Çolak, "SINIFLANDIRMA PROBLEMLERİNDE TOPLULUK ÖĞRENME YÖNTEMLERİNİN İNCELENMESİ VE KÜÇÜK HÜCRELİ DIŞI AKCİĞER KANSERİ VERİLERİ ÜZERİNE BİR UYGULAMASI," 2021.
- [17] Ö. EKREM, O. K. M. SALMAN, B. AKSOY, and S. A. İNAN, "YAPAY ZEKÂ YÖNTEMLERİ KULLANILARAK KALP HASTALIĞININ TESPİTİ," *Mühendislik Bilimleri ve Tasarım Dergisi*, vol. 8, no. 5, pp. 241–254, Dec. 2020, doi: 10.21923/jesd.824703.