

Kablosuz Sinyal Gücünü Kullanarak İç Mekan Kullanıcı Lokalizasyonu için Karar Ağacı Algoritmalarının Karşılaştırılması

A Comparison of Decision Tree Algorithms for Indoor User Localization Using Wireless Signal Strength

Ebru Efeoğlu¹ 



ÖZ

İç mekanda kullanıcı ve cihazları yerleştirmek geniş bir uygulama alanına sahiptir. Akıllı ev sistemleri, sınırlı bölgelerdeki suçluları bulma, bir erişim noktasındaki kullanıcı sayısını belirlemek için kullanılabilir. Bu çalışmanın amacı kablosuz sinyal gücüne dayalı olarak iç mekanda kullanıcıların konumunu belirlemektir. Bunun yanı sıra tasarlanacak izleme cihazlarında kullanılacak en iyi karar ağacı sınıflandırma algoritmasını saptamaktır. Bu amaçla çalışmada 12 farklı algoritma kullanılmış ve performans analizi yapılarak algoritmaların performansları karşılaştırılmıştır. Performans analiz yöntemi olarak 10 kat çapraz doğrulama kullanılmıştır. Performans değerlendirmesi yapılırken algoritmaların hem çaprazdoğrulama yapılmadan önceki sınıflandırma performansı hemde çapraz doğrulama sonrası yapılan sınıflandırma performansları karşılaştırılmıştır. Çalışmada Dengeli bir veri seti kullanıldığı için Performans analizinde dengeli veri setlerinin sınıflandırılmasında kullanılan performans metrikleri tercih edilmiştir. Performans analizinde doğruluk, karışıklık matrisi, kesinlik, duyarlılık, F-skoru, Kappa istatistiği, Kök ortalama hata değeri ve ROC değeri kullanılmıştır. Analiz sonucunda Analizden sonra en iyi performansı Random Forest Rasgele orman algoritmasının elde ettiği gözlemlenmiştir. Algoritmanın çapraz doğrulama öncesi ve sonrasında hesaplanan tüm metric değerleri diğer algoritmalarından daha yüksektir.

Anahtar Kelimeler: İç mekan, Karar ağaçları, Kablosuz sinyal gücü, Konumlandırma, Performans analizi

¹(Dr. Öğr. Üyesi) Kütahya Dumlupınar Üniversitesi, Yazılım Mühendisliği Bölümü, Kütahya, Türkiye

ORCID: E.E. 0000-0001-5444-6647

Corresponding author:

Ebru EFOĞLU
Kütahya Dumlupınar Üniversitesi, Yazılım
Mühendisliği Bölümü, Kütahya, Türkiye
E-mail address: ebru.efoглу@dpu.edu.tr

Submitted: 20.02.2022

Revision Requested: 19.06.2022

Last Revision Received: 21.06.2022

Accepted: 19.07.2022

Published Online: 19.09.2022

Citation: Efeoğlu, E. (2022). Kablosuz sinyal gücünü kullanarak iç mekan kullanıcı lokalizasyonu için karar ağacı algoritmalarının karşılaştırılması. *Acta Infologica*, 6(2), 163-173. <https://doi.org/10.26650/acin.1076352>

ABSTRACT

Localizing users and devices indoors has a wide range of applications. Smart home systems can be used to locate criminals in restricted areas and determine the number of users at an access point. The aim of this study is to determine the location of users indoors using wireless signal strength as well as the best decision tree classification algorithm that can be used in monitoring devices that will be designed. For this purpose, the study uses 12 different algorithms and compares their performances by conducting a performance analysis. The study uses 10-fold cross validation as the performance analysis method. While evaluating the performance, the algorithms' classification performance were compared before and after the cross-validation. Due to the study using a balanced dataset, the performance metrics used for classifying balanced datasets have been preferred in the performance analysis. As a result of the analysis, the random forest algorithm was observed to have achieved the best performance. All metric values calculated before and after the cross-validation of the random forest algorithm were higher than those for the other algorithms.

Keywords: Indoor, Decision Trees, Wi-Fi, Localization, Performance analysis

1. GİRİŞ

Konumlandırma teknolojileri iç ve dış mekan konumlandırma sistemleri olarak iki gruba ayrılabilir. Dış mekan konumlandırmada (Dardari, Closas, ve Djurić, 2015), yaygın olarak kullanılan navigasyon sistemleri, metre düzeyinde doğrulukla konum hizmetleri sağlayabilir. İç mekanlarda engellerden kaynaklanan uydu sinyal kayıpları, çok yollu etki ve tutarsız zaman gecikmesi sorunları nedeniyle uydu tabanlı konumlandırma sistemleri doğru sonuçlar vermez. Ayrıca bir iç mekan konum hizmetinin gereksinimlerini karşılayamaz (J. Shi, 2013). Uydu tabanlı konumlandırma sistemlerinin iç mekan konum belirlemede karşılaştığı teknik sorunlar (Van Diggelen ve Abraham, 2001) tarafından ele alınmıştır.

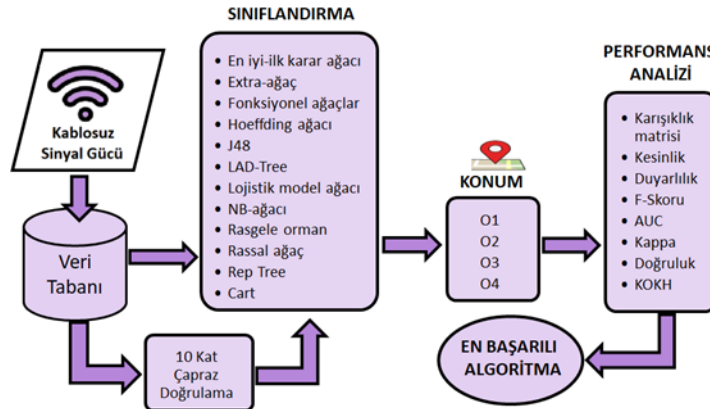
İç mekan senaryoları insan yaşamının büyük bir kısmı ile ilgili olması nedeniyle insanların veya nesnelerin iç mekanlardaki konumlarını sağlamak için sürekli ve gerçek zamanlı olarak çalışan sistemler olarak düşünülebilir (Gu, Lo, ve Niemegeers, 2009). Bu sistemler, yaşlı ve engelli kişilerin günlük fiziksel aktivitelerini tanımda ve yaşamsal belirtilerinin (Barsocchi et al., 2015) gerçek zamanlı olarak izlenmesinde (Wang, Yang, ve Dong, 2017 ; Huang, Wang, Zhang, Hu, ve Jin, 2019) ve acil durumlarda kullanılabilir. Bunun yanı sıra anaokulu güvenliğinde (Lin, Lee, Syu, ve Chen, 2010), görme engelli kişiler (Abu Doush, Alshatnawi, Al-Tamimi, Alhasan, ve Hamasha, 2017; Sáenz ve Sánchez, 2009; Zhang ve Ye, 2020) ve turistler için navigasyon sağlamada iç mekan konumlandırma hizmetlerine ihtiyaç duyulur. İnsanlar dışında hastanelerdeki pahalı ekipmanları takip etmek, hırsızlığı önlemek ve ameliyatlar sırasında robotik asistanlar için hassas konumlandırma yapmada İç mekan konumlandırma sistemleri talep edilmektedir (Correa, Llado, Morell, ve Vicario, 2016; Gu et al., 2009). İç mekan konumlandırma üzerine çeşitli çalışmalar yapılmıştır (Akleyek, Kiliç, Söylemez, Aruk, & Çavuş, 2020). Bu amaç için birçok teknolojiden yararlanılmıştır. Bu teknolojiler (Arslantaş & Ökdem, 2019) tarafından anlatılmıştır. Artırılmış gerçeklik teknolojisi ile iç mekân navigasyonu uygulanmıştır (Çalık & Gülgen, 2021). Akıllı ev teknolojisi için kablosuz akıllı Kit. (Kuncan & Ömer, 2019) ve bir akıllı telefon uygulaması önerilmiştir (Homayounvala, Nabati, Shahbazian, Ghorashi, & Moghtadaiee, 2019). Zigbee modülü kullanılarak ses kontrollü bir ev otomasyonunun geliştirilmiştir (Çubukçu, Kuncan, Kaplan, & Ertunc, 2015). Bluetooth sinyal gücüne dayalı (Rida, Liu, Jadi, Algawhari, ve Askourih, 2015; Subhan, Hasbullah, ve Ashraf, 2013) konum belirleme çalışmaları yapılmıştır. Akıllı mobil cihazlar için bluetooth ve WiFi iletişimine dayalı hibrit bir iç mekan konum mekanizması (Su, Liao, Lin, ve Lin, 2015) tasarlanmıştır. Radyo Frekans sinyali gücü ölçümlerinden iç mekan konum tahmini (Seco, Jiménez, ve Zampella, 2013) yapılmış ve konum belirleme için yeni bir yöntem önerilmiştir (Chen ve Huang, 2009). Ayrıca maliyeti düşürme çalışmalarında bulunulmuş ve düşük maliyetli iç mekan konumlandırma sistemi tasarlanmıştır (Randell ve Muller, 2001). Çoklu optik alıcılar kullanarak iç mekan konum takibi (Yasir, Ho, ve Vellambi, 2015) yapılmıştır. Bu teknolojiler ile birlikte konum belirlemek için makine öğrenmesi yöntemleri sıklıkla kullanılmaktadır. İç mekan konumlandırma için makine öğrenimi için anket çalışması yapılmıştır (Roy & Chowdhury, 2021). İç mekanda gerçek zamanlı kişi tespitinde makine öğrenmesi algoritmalarının karşılaştırmalı Başarım analizi yapılmıştır (Taşer & Akram, 2021). Dinamik yapay sinir ağı ile iç mekân konum kestirimi yapılmıştır (Mert, Ferdi, & Hakan, 2020) . Birçok makine öğrenmesi teknikleri vardır. Bina içi lokalizasyon uygulamalarında hangi makine öğrenmesi yönteminin kullanıldığı çok önemlidir (Bozkurt, Elibol, Gunal, ve Yayan, 2015). Çalışmada karar ağaçlarını tercih edilmesinin nedeni, Karar Ağaçlarının karar verirken insan düşünme yeteneğini taklit etmesi nedeniyle anlaşılması kolay ve başarı oranının yüksek olmasıdır.

2. MATERIAL VE YÖNTEM

2.1. Veri Seti

Çalışmada kullanılan veriler UCI (Frank, 2010) kütüphanesinden elde edilmiştir. Veri setinde 4 farklı odada bulunan kişilerin akıllı telefonları ile 7 farklı kablosuz sinyal gücü ölçümünden oluşan 2000 adet ölçüm bulunmaktadır. Bu veriler ilk olarak (Rohra, Perumal, Narayanan, Thakur, ve Bhatt, 2017)'de kullanılmıştır. Bu çalışmada (Rohra et al., 2017) den farklı olarak, 12 farklı karar ağacı algoritması kullanılmıştır. Bu veriler kullanılarak Karar ağaçları algoritmalarının dışındaki sınıflandırma algoritmalarının karşılaştırması yapılmıştır (Sabancı, Yigit, Ustun, Toktas, ve Aslan, 2018). Fakat yapılan çalışmada algoritmaların performans analizlerinin yapılmasında hata matrisi gösterilmiş ve doğruluk ölçütü göz önüne alınmıştır. Oysaki algoritmaların performanslarının karşılaştırılmasında sadece doğruluk ölçütünün kullanımı yeterli değildir. Çünkü

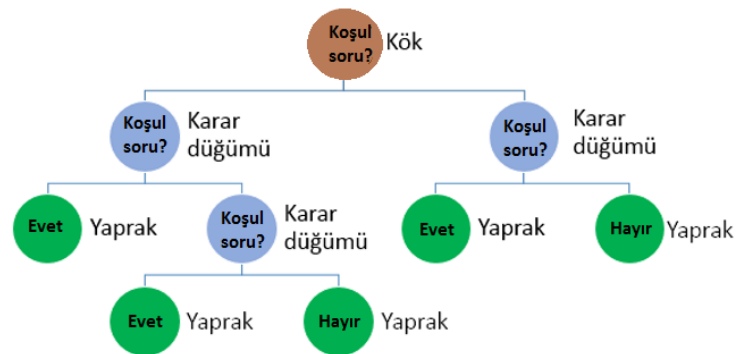
sınıflandırma algoritmasının tüm tahminleri sayıca fazla olan sınıftan yapması durumunda, algoritma neredeyse hiçbir şey öğrenmediği halde doğruluk oranı yüksek çıkabilir (Joshi, 2016). Ayrıca sınıflandırma algoritmaları arasında problemin çözümünde en etkili algoritmayı seçebilmek için algoritmanın veri tabanında bulunmayan bir veriyi sınıflandırabilme yeteneğinin ölçülmesi önemlidir. Bu nedenle, bu çalışmada algoritmaların veri tabanında bulunmayan bir veriyi sınıflandırabilme yeteneğinin ölçülmesi için 10 kat çapraz doğrulama işlemi gerçekleştirilmiştir. Algoritmaların performans analizi yapılırken hem çapraz doğrulama öncesi performansları hem de çapraz doğrulamadan sonraki performansları karşılaştırılmış ve en iyi sonuç veren algoritma belirlenmiştir. Çalışmayı anlatan akış diyagramı Şekil 1'de verilmiştir.



Şekil 1. Çalışmanın akış diyagramı

2.2. Karar Ağaçları

Karar ağaçlarında model, kökleri yukarıda, ters çevrilmiş bir ağaca benzetilebilir. Bir karar ağacı, karar vermede kullanılan karar düğümleri, karar çıktıları olarak sayılan yaprak düğümleri ve dallardan oluşur. Bir yaprak düğümünden sonra ağaç daha fazla ayrılmaz. Bir karar ağacına basitçe bir soru sorularak bazı özellikler test edilir. Sorunun cevabına göre (Evet/Hayır) karar düğümü alt düğümlere bölünür. Bir karar ağacının genel yapısını açıklayan diyagram Şekil 2'de verilmiştir. Karar ağacındaki ilk düğüme kök düğüm adı verilir ve dallanma oradan başlar.



Şekil 2. Karar Ağacı yapısı

Karar ağaçlarının çalışma prensibi kısaca şu şekilde özetlenebilir. Öncelikle, kök özneliğinin değerlerini kayıt özneliği ile karşılaştırır sonrasında karşılaştırmaya dayalı olarak dallanır ve bir sonraki düğüme geçer. Yine öznelik değerini alt düğümlerle karşılaştırır ve bir sonraki düğüme geçer. Bu işlem yaprak düğüme gelinceye kadar devam eder. Bir karar ağacını uygulamada kök düğüm ve alt düğümler için en iyi özneliğin nasıl seçileceği oldukça önemlidir. Öz nitelik seçimi için bilgi kazancı ve Gini endeksi olmak üzere iki popüler yöntem bulunmaktadır. Bilgi Kazancı, bir özelliğin bir sınıf ile ilgili ne ölçüde bilgi sağladığını hesaplar. Bilgi kazancının değeri kullanılarak düğümler bölünür ve karar ağacı oluşturulur. Bir

karar ağacı algoritması bilgi kazancının maksimum olmasını ister. Bölünme sırasında yüksek bilgi kazancına sahip bir düğüm/öznitelik önce bölünür. Gini İndeksi, karar ağacının oluşturulmasında bir saflık ölçütü olarak kullanılır. Nitelik seçiminde düşük gini indeksine sahip olan nitelik tercih edilir. Karar ağaçlarında çok büyük ağaçlar fazla uyum riski oluşturabilmesine karşın küçük ağaçlar da veri kümesine ait önemli özelliklerini kaçırabilir. Bu durumun üstesinden gelip optimal karar ağacı elde etmek için gereksiz düğümlerin ağaçtan atılması işlemi olarak açıklanan budama yöntemi kullanılır. Birçok karar ağacı algoritması vardır. Lojistik Model Ağacında ayrılma (LMT) J48 algoritmasına benzemektedir. Karar ağacı indüksiyonu ile lojistik regresyon modeli birleştirilmiştir. Budama yapılarak ağaç sadeleştirilir (Landwehr, Hall, ve Frank, 2005). Ekstra Ağaçlar (Extra tree) algoritmasında, eğitim veri kümesinden çok sayıda budanmamış karar ağacı oluşturularak çalışır. Tahminler, regresyon durumunda karar ağaçlarının tahmininin ortalaması alınarak veya sınıflandırma durumunda çoğunluk oylaması kullanılarak yapılır (Geurts, Ernst, ve Wehenkel, 2006). Hoefding ağacındaki (Hoefding tree) düğümlerin parçalanmasına hoefding sınırı ile bilinen bir istatistiksel değer kullanılarak karar verilir (Hulten, Spencer, ve Domingos, 2001). İlk olarak (J. R. Quinlan, 1987) tarafından önerilen Rep tree algoritması, hatayı en aza indirmek için entropi ile bilgi kazanımını kullanır (Srinivasan ve Mekala, 2014). CART Algoritmasında, her karar düğümü farklı ayırma kriteri kullanılarak ağaç iki dala ayrılır (Breiman, Friedman, Olshen, ve Stone, 1984). En iyi-ilk karar ağacı (BF tree) Gini İndeksini (H. Shi, 2007) ve C4.5 olarak ta bilinen J48 algoritması özellik seçiminde bilgi kazancını kullanır (R. Quinlan, 1993). Fonksiyonel ağaçlar (FT tree) algoritmasında veri bir örnek uzayında hiper-dikdörtgenlere bölünür. Karar uzayı test edilen özelliğe dik diğer özelliklere ise paralel olmalıdır (Gama, 2004). NB tree algoritması, Bayes kuralının karar ağaçlarına uygulanması ile oluşturulmuştur (Kohavi, 1996). Rasgele orman nitelik seçiminde düğümleri dallara ayırmadan düğümlerden rasgele nitelikler alır ve bu nitelikler arasından en iyisini seçer sonrasında bu niteliklere göre düğümleri dallara ayırır. Ağaç budama işlemi yapılmaz (Breima, 2010). Rasgele ağaç algoritması ise her düğümden belli sayıda özellik alınarak ağaç oluşturulur (Breima, 2010).

2.3. Performans Analizi

Sınıflandırma algoritmalarının performans değerlendirilmesi için K-kat çapraz doğrulama tekniği tercih edilmiştir. K-katlama çapraz doğrulama tekniği, olası aşırı uyumu önlemek ve modelin daha önce görmediği bir veri kümesi üzerinde nasıl performans gösterdiğini anlamak için veri kümesini eğitim ve test kümelerine böler. Çünkü aşırı uyumda model eğitim setinde başarılı olurken, hiç görmediği veri setleri üzerinde başarısız tahminler yapar. K-katlama çapraz doğrulama tekniği, eğitim veri setini rastgele k parçaya böler. Eğitim için k-1, test seti için 1 kısım kullanılır ve bu k defa tekrarlanır. Her turda elde edilen değerler toplanır ve modelin performansı değerlendirilir. K sayısı bu çalışmada olduğu gibi genellikle 10'dur. Bir sınıflandırma algoritmasının sınıflandırma sürecinin sonunda ne kadar iyi performans gösterdiğini değerlendirmek için çeşitli ölçütler bulunmaktadır. Bu ölçütlerden en sık kullanılanı, gerçek ve tahmin edilen sınıflar hakkında bilgiler içeren karışıklık matrisidir. Bu matriste Doğru Pozitif, Doğru Negatif, Yanlış Pozitif ve Yanlış Negatif olmak üzere 4 farklı durum ortaya çıkmaktadır. Yanlış Pozitif ve Yanlış Negatif değerler, tahmin edilen sınıf ile gerçek sınıf aynı olmadığında ortaya çıkar. Bunlar, algoritmanın yanlış tahmin ettiği sınıfların sayısını gösterir. Ayrıca bu değerler kullanılarak performans değerlendirilmesi için bazı metrikler hesaplanabilir. Örneğin, tahmin edilen gözlem sayısını toplam gözlem sayısına bölerek Doğruluk değeri, Doğru Pozitif gözlem sayısını toplam pozitif gözlem sayısına bölerek Kesinlik değeri, doğru gözlem sayısını toplam gözlem sayısına bölerek duyarlılık, kesinlik ve duyarlılık değerlerinin harmonik ortalaması alınarak F-Skoru hesaplanabilir. Duyarlılık, pozitif olarak tahmin edilmesi gerekenlerin ne kadarını pozitif olarak tahmin edildiğini gösteren bir performans ölçütüdür. Kesinlik pozitif olarak tahmin edilen değerlerin gerçekten kaç tanesinin pozitif olduğunu göstermektedir. Ayrıca bir olasılık eğrisi olan ROC eğrilerinin incelenmesi performans değerlendirmesinde önemli bir yere sahiptir. ROC eğrisinde X eksenini Yanlış Pozitif Oranını, Y eksenini de Doğru Pozitif Oranını göstermektedir. Bu eğrilerin altında kalan alan değerine bakılarak algoritmanın sınıflandırma performansı hakkında yorum yapılabilir. Bu metriklerin dışında, performans değerlendirme çalışmalarında yaygın olarak Kappa istatistiksel değeri kullanılmaktadır. Kappa değeri, yapılan sınıflandırma ile gerçek sınıflandırma arasındaki uyumu ölçer ve -1 ile 1 arasında değişir. -1 değeri tam bir uyumsuzluğu, 1 değeri ise mükemmel uyumu gösterir. Kök otalama kare hata değeri ise sınıflandırmanın ne kadar hata ile yapıldığını gösteren bir ölçüttür. Sınıflandırmada hata oranının düşük olması sınıflandırmanın başarılı olduğunu ifade eder.

3. BULGULAR VE TARTIŞMA

Farklı odalarda bulunan kullanıcıların akıllı telefonları ile 7 farklı Wi-Fi kaynağından gelen sinyal güçleri ölçülmüştür. Toplam 2000 adet ölçüm alınmıştır. Bu ölçümler ve 12 adet farklı karar ağacı algoritması kullanılarak kullanıcının hangi odada olduğunu tespiti yapılmıştır. Sınıflandırma uygulamasında kullanıcılar 4 farklı odada bulunduğu için veriler, Oda1(O1), Oda2(O2), Oda3(O3), Oda4(O4) olmak üzere 4 farklı sınıfa ayrılmıştır. Algoritmaların sınıflandırma performansının belirlenmesi için çapraz doğrulama öncesi ve sonrasında performans analizi yapılmıştır. Algoritmaların çapraz doğrulama uygulanmadan yapılan sınıflandırmadan elde edilen karışıklık matrisi Şekil 3’ de verilmiştir.

		En iyi-ilk karar ağacı							Extra-ağaç							Fonksiyonel ağaçlar							
		Tahmin edilen Sınıf							Tahmin edilen Sınıf							Tahmin edilen Sınıf							
		O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4
Gerçek	O1	499	0	0	1	Gerçek	O1	500	0	0	0	Gerçek	O1	500	0	0	0	Gerçek	O1	500	0	0	0
	O2	0	490	10	0		O2	0	500	0	0		O2	0	483	17	0		O3	0	3	495	2
	O3	2	4	494	0		O3	0	0	500	0		O3	0	0	1	499		O4	0	0	1	499
	O4	2	0	0	498		O4	0	0	0	500		O4	0	0	0	1		499				
		Hoeffding -ağacı							J48							LAD Tree							
		Tahmin edilen Sınıf							Tahmin edilen Sınıf							Tahmin edilen Sınıf							
		O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4
Gerçek	O1	0	0	133	367	Gerçek	O1	499	0	1	0	Gerçek	O1	497	0	3	0	Gerçek	O1	497	0	3	0
	O2	0	471	29	0		O2	0	496	4	0		O2	0	482	18	0		O3	1	14	481	4
	O3	0	1	497	2		O3	0	4	496	0		O3	1	14	481	4		O4	3	0	4	493
	O4	0	0	2	498		O4	3	0	1	496		O4	3	0	4	493						
		Lojistik model ağacı							NB-Tree							Rasgele orman							
		Tahmin edilen Sınıf							Tahmin edilen Sınıf							Tahmin edilen Sınıf							
		O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4
Gerçek	O1	499	0	0	1	Gerçek	O1	499	0	0	1	Gerçek	O1	500	0	0	0	Gerçek	O1	500	0	0	0
	O2	0	485	15	0		O2	0	478	22	0		O2	0	500	0	0		O3	0	0	500	0
	O3	0	5	495	0		O3	2	0	496	2		O3	0	0	500	0		O4	0	0	0	500
	O4	0	0	0	500		O4	2	0	1	497		O4	0	0	0	500						
		Rassal ağaç							Rep tree							Cart							
		Tahmin edilen Sınıf							Tahmin edilen Sınıf							Tahmin edilen Sınıf							
		O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4
Gerçek	O1	500	0	0	0	Gerçek	O1	499	0	1	0	Gerçek	O1	499	0	1	0	Gerçek	O1	499	0	1	0
	O2	0	500	0	0		O2	0	478	22	0		O2	0	478	22	0		O3	1	6	489	4
	O3	0	0	500	0		O3	1	4	491	4		O3	1	6	489	4		O4	4	0	3	493
	O4	0	0	0	500		O4	1	0	3	496		O4	4	0	3	493						

Şekil 3. Çapraz doğrulama uygulanmadan yapılan sınıflandırmadan elde edilen karışıklık matrisi.

Burada yeşil renkle gösterilen köşegen değerleri Doğru Pozitif ve Doğru Negatif değerlerini ifade eder. Bu değerler algoritmanın doğru tahmin ettiği veri sayısıdır. Diğer değerler ise yanlış tahmin edilen verilerin sayısını gösterir. Örneğin En iyi karar algoritmasının toplam doğru tahmin ettiği veri sayısı 1981’dir. Algoritma O1’de bulunan 1 kişinin O4’ te olduğunu, O2’ de olan 10 kişiyi O3’ te olduğunu, O3’ te olan 2 kişinin O1’ de, 4 kişiyi de O2’de ve O4’te olan 2 kişiyi de O1’de olduğu tahmininde bulunmuştur. Algoritmanın konumunu yanlış tahmin ettiği kişi sayısı toplam 19’dur. Karışıklık matrisinden yararlanılarak Kesinlik, Duyarlılık, F1-skoru ve Kappa değeri hesaplanmıştır. Bu değerlerden ROC eğrileri çizilmiş ve eğri altında kalan alanlar hesaplanarak hesaplanan diğer performans ölçütleri ile birlikte Tablo 1’de verilmiştir.

Tablo 1. 10 kat çapraz doğrulama öncesi performans metrikleri

Algoritma	Sınıf	Kesinlik	Duyarlılık	F1 skoru	AUC	Kappa
En iyi-ilk karar ağacı	O 1	0.99	0.99	0.99	0.99	0.98
	O 2	0.99	0.98	0.98	0.99	
	O 3	0.98	0.98	0.98	0.99	
	O 4	0.99	0.99	0.99	0.99	
Extra-ağaç	O 1	1	1	1	1	1
	O 2	1	1	1	1	
	O 3	1	1	1	1	
	O 4	1	1	1	1	
Fonksiyonel ağaçlar	O 1	1	1	1	1	0.98
	O 2	0.99	0.96	0.98	0.99	
	O 3	0.96	0.99	0.97	0.99	
	O 4	0.99	0.99	0.99	1	
Hoeffding-ağacı	O 1	---	0.0	---	0.5	0.64
	O 2	0.99	0.94	0.96	0.99	
	O 3	0.75	0.99	0.85	0.97	
	O 4	0.57	0.99	0.72	0.93	
J48	O 1	0.99	0.99	0.99	0.99	0.99
	O 2	0.99	0.99	0.99	0.99	
	O 3	0.99	0.99	0.99	0.99	
	O 4	1	0.99	0.99	0.99	
LAD-Tree	O 1	0.99	0.99	0.99	1	0.96
	O 2	0.97	0.96	0.96	0.99	
	O 3	0.95	0.96	0.95	0.99	
	O 4	0.99	0.98	0.98	1	
Lojistik model ağacı	O 1	1	0.99	0.99	1	0.98
	O 2	0.99	0.97	0.98	0.99	
	O 3	0.97	0.99	0.98	0.99	
	O 4	0.99	1	0.99	1	
NB-ağacı	O 1	0.99	0.99	0.99	1	0.98
	O 2	1	0.95	0.97	0.99	
	O 3	0.95	0.99	0.97	0.99	
	O 4	0.99	0.99	0.99	1	
Rasgele orman	O 1	1	1	1	1	1
	O 2	1	1	1	1	
	O 3	1	1	1	1	
	O 4	1	1	1	1	
Rassal ağaç	O 1	1	1	1	1	1
	O 2	1	1	1	1	
	O 3	1	1	1	1	
	O 4	1	1	1	1	
Rep Tree	O 1	0.99	0.99	0.99	0.99	0.97
	O 2	0.99	0.95	0.97	0.99	
	O 3	0.95	0.98	0.96	0.98	
	O 4	0.99	0.99	0.99	0.99	
Cart	O 1	0.99	0.99	0.99	0.99	0.97
	O 2	0.98	0.95	0.97	0.99	
	O 3	0.95	0.97	0.96	0.98	
	O 4	0.99	0.98	0.98	0.99	

Bu ölçütlere göre sınıflandırmanın başarılı olması değerlerin 1'e yaklaşması anlamına gelmektedir. Değer 1 olursa mükemmel bir sınıflandırma yapıldığı ve algoritmanın çok başarılı olduğu kabulü yapılır.

Algoritmaların veri tabanında bulunmayan bir veriyi sınıflandırma başarısını test etmek için çapraz doğrulama yöntemi kullanılmıştır. 10 Kat çapraz doğrulama sonrası elde edilen karışıklık matrisi Şekil 4'de diğer performans ölçütleri ise Tablo 2'de verilmiştir.

		En iyi-ilk karar ağacı						Extra-ağaç						Fonksiyonel ağaçlar			
		Tahmin edilen Sınıf						Tahmin edilen Sınıf						Tahmin edilen Sınıf			
		O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4
Gerçek	O1	498	0	1	1	Gerçek	O1	485	2	9	4	Gerçek	O1	496	0	1	3
	O2	0	472	28	0		O2	1	470	29	0		O2	0	481	19	0
	O3	2	17	479	2		O3	9	31	451	9		O3	5	7	484	4
	O4	5	0	3	492		O4	8	0	7	485		O4	2	0	2	496
		Hoeffding -ağacı						J48						LAD Tree			
		Tahmin edilen Sınıf						Tahmin edilen Sınıf						Tahmin edilen Sınıf			
		O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4
Gerçek Sınıf	O1	499	0	1	0	Gerçek Sınıf	O1	496	0	4	0	Gerçek Sınıf	O1	495	0	5	0
	O2	0	478	22	0		O2	0	481	19	0		O2	0	475	25	0
	O3	3	3	492	2		O3	2	17	477	4		O3	3	19	474	4
	O4	2	0	1	497		O4	5	0	6	489		O4	4	0	4	492
		Lojistik model ağacı						NB-Tree						Rasgele orman			
		Tahmin edilen Sınıf						Tahmin edilen Sınıf						Tahmin edilen Sınıf			
		O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4
Gerçek Sınıf	O1	499	0	1	0	Gerçek Sınıf	O1	494	0	2	4	Gerçek Sınıf	O1	499	0	1	0
	O2	0	483	17	0		O2	0	472	28	0		O2	0	481	19	0
	O3	3	7	488	2		O3	1	7	487	5		O3	0	5	494	1
	O4	2	0	2	496		O4	3	0	2	495		O4	3	0	1	496
		Rassal ağaç						Rep tree						Cart			
		Tahmin edilen Sınıf						Tahmin edilen Sınıf						Tahmin edilen Sınıf			
		O1	O2	O3	O4			O1	O2	O3	O4			O1	O2	O3	O4
Gerçek Sınıf	O1	491	0	5	4	Gerçek Sınıf	O1	499	0	1	0	Gerçek Sınıf	O1	499	0	1	0
	O2	0	474	26	0		O2	0	471	29	0		O2	0	476	24	0
	O3	2	25	471	2		O3	5	14	477	4		O3	1	12	483	4
	O4	4	0	8	488		O4	5	0	3	492		O4	5	0	2	493

Şekil 4.Çapraz doğrulama uygulamasından sonra elde edilen karışıklık matrisi.

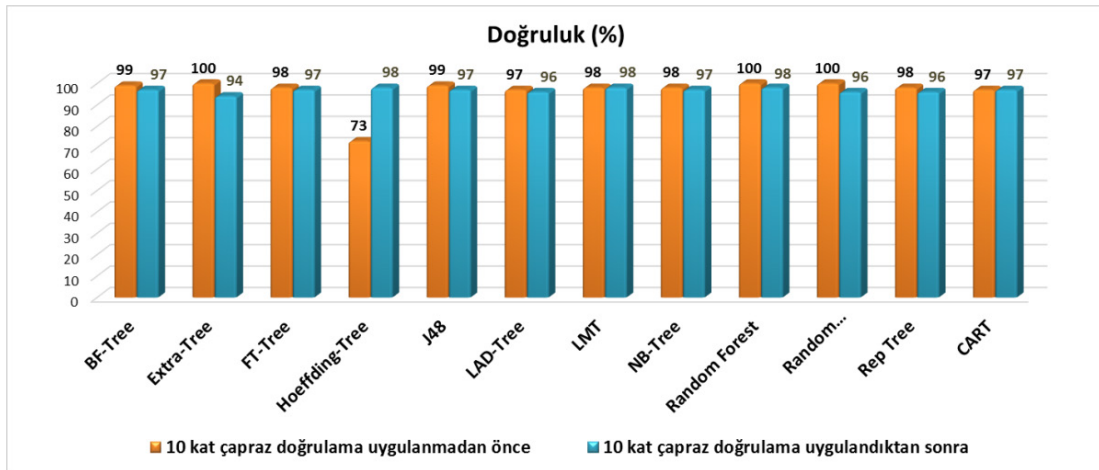
Çapraz doğrulamadan sonra Extra tree algoritması 2000 veriden 1891 tanesini doğru, 109 tanesini yanlış sınıflandırmış, Rasgele orman 1970 doğru, 30 yanlış Rassal ağaç 1924 doğru 76 yanlış sınıflandırmıştır. Hoeffding tree algoritmasının doğru sınıflandırdığı veri sayısı 196 yanlış sınıflandırdığı veri sayısı ise 34 olmuştur.

Tablo 2. 10 kat çapraz doğrulama sonrası performans metrikleri

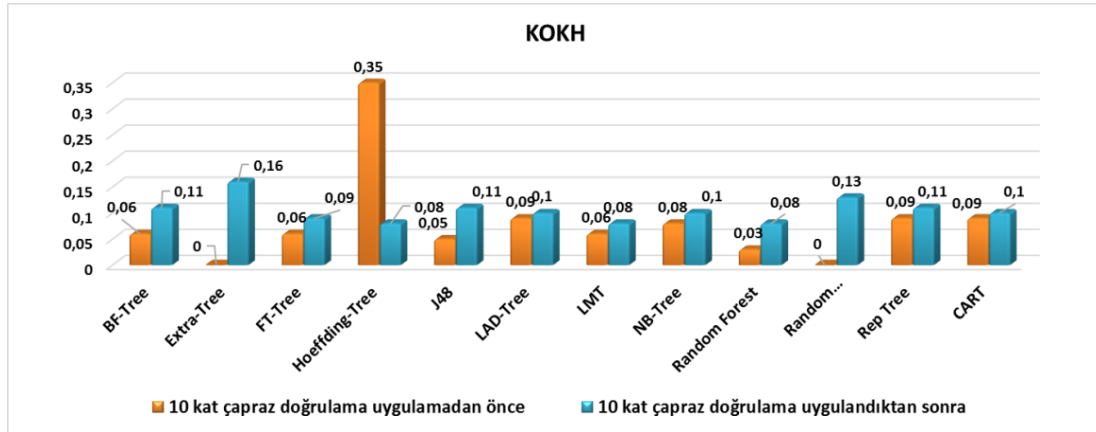
Algoritma	Sınıf	Kesinlik	Duyarlılık	F1 skoru	AUC	Kappa
En iyi-ilk karar ağacı	O 1	0.98	0.99	0.99	0.99	0.96
	O 2	0.96	0.94	0.95	0.98	
	O 3	0.93	0.95	0.94	0.97	
	O 4	0.99	0.98	0.98	0.99	
Extra-ağaç	O 1	0.96	0.97	0.96	0.97	0.92
	O 2	0.93	0.94	0.93	0.95	
	O 3	0.90	0.90	0.90	0.93	
	O 4	0.97	0.97	0.97	0.98	
Fonksiyonel ağaçlar	O 1	0.98	0.99	0.98	0.99	0.97
	O 2	0.98	0.96	0.97	0.99	
	O 3	0.95	0.96	0.96	0.98	
	O 4	0.98	0.99	0.98	0.99	
Hoeffding-ağacı	O 1	0.99	0.99	0.99	1	0.97
	O 2	0.99	0.95	0.97	0.99	
	O 3	0.95	0.98	0.96	0.99	
	O 4	0.99	0.99	0.99	1	

J48	O1	0.98	0.99	0.98	0.99	0.96
	O2	0.96	0.96	0.96	0.98	
	O3	0.94	0.95	0.94	0.97	
	O4	0.99	0.97	0.98	0.98	
LAD-Tree	O1	0.98	0.99	0.98	0.99	0.95
	O2	0.96	0.95	0.95	0.99	
	O3	0.93	0.94	0.94	0.99	
	O4	0.99	0.98	0.98	0.99	
Lojistik model ağacı	O1	0.99	0.99	0.99	1	0.97
	O2	0.98	0.96	0.97	0.99	
	O3	0.96	0.97	0.96	0.99	
	O4	0.99	0.99	0.99	1	
NB-ağacı	O1	0.99	0.98	0.99	1	0.96
	O2	0.98	0.94	0.96	0.99	
	O3	0.93	0.97	0.95	0.99	
	O4	0.98	0.99	0.98	0.99	
Rasgele orman	O1	0.99	0.99	0.99	1	0.98
	O2	0.99	0.96	0.97	0.99	
	O3	0.95	0.98	0.97	0.99	
	O4	0.99	0.99	0.99	1	
Rassal ağaç	O1	0.98	0.98	0.98	0.98	0.94
	O2	0.95	0.94	0.94	0.96	
	O3	0.92	0.94	0.93	0.95	
	O4	0.98	0.97	0.98	0.98	
Rep Tree	O1	0.98	0.99	0.98	0.99	0.95
	O2	0.97	0.94	0.95	0.99	
	O3	0.93	0.95	0.94	0.98	
	O4	0.99	0.98	0.98	0.99	
Cart	O1	0.98	0.99	0.99	0.99	0.96
	O2	0.97	0.95	0.96	0.98	
	O3	0.94	0.96	0.95	0.97	
	O4	0.99	0.98	0.98	0.99	

En yüksek Kappa değeri Rasgele orman algoritmasına aittir. Rasgele orman algoritmasında diğer odaların performans ölçütlerine oranla en düşük F1 skoru ve AUC değeri Oda3 için elde edilmiştir. Algoritmaların eğitim ve çapraz doğrulamadaki doğruluk ve hata oranları sırasıyla Şekil 5 ve Şekil 6’te verilmiştir.



Şekil 5. Algoritmaların doğruluk oranı.



Şekil 6. Algoritmaların Kök ortalama kare hata deęeri.

4. SONUÇLAR

Kablosuz sinyal gücü kullanılarak 4 farklı odada bulunan akıllı telefon kullanıcılarının konumunun belirlenmesinde 12 farklı karar ağacı algoritması kullanılmıştır. Çalışmada fazla sayıda algoritma kullanılması bu problemin çözümünde en başarılı algoritmanın bulunabilmesi için önemlidir. Ayrıca algoritmaların performans analizinde sadece bir kaç metrik deęeri ile yetinilmeyip bir çok metrik deęeri hesaplanmış ve en iyi performansı belirleyebilmek için yorumlanmıştır. Performans deęerlendirmesi için de sadece veri tabanında bulunan örneklerin sınıflandırılmasındaki performanslar deęil aynı zamanda algoritmanın daha önce hiç görmedięi bir örneęi sınıflandırma performansları incelenmiştir. Tüm bu durumlar çalışmayı daha önceki çalışmalardan üstün kılmaktadır. Çalışmada en popüler performans analiz yöntemi olan 10 kat çapraz doęrulama yöntemi ile performans analizi yapılmıştır. Analizinde doęruluk, karışıklık matrisi, kesinlik, duyarlılık, F-skoru, Kappa istatistięi, Kök ortalama hata deęeri ve ROC deęeri kullanılmıştır. Algoritmaların hangi odadaki kişilerin tespitinde daha başarılı olduęunu anlamak amacıyla bu deęerler her sınıf için ayrı ayrı hesaplanmıştır. Performans metriklerinin her sınıf için ayrı hesaplanması da yine çalışmayı dięer çalışmalardan farklılaştırmıştır. Performans analizi sonucunda algoritmaların birbirlerine yakın sonuçlar verdięi ve çapraz doęrulama sonrası yapılan sınıflandırmada tüm algoritmaların performanslarında doęal bir düşüş gözlenmiştir. Çapraz doęrulama öncesi hesaplanan performans metriklerine bakıldığında Extra ağaç, Rasgele orman ve Rassal ağaç algoritmalarının metrik deęerlerinin 1 tam puan aldıęı görülmektedir. Bu durum veri tabanında bulunan verilerle yapılan sınıflandırmalarda bu algoritmaların performansının dięer algoritmalarından daha iyi olduęunu gösterir. En düşük metrik deęerleri ise Hoeffding ağacı algoritmasına aittir. Algoritma Odal de bulunan hiç kimseyi doęru sınıflandıramamıştır. Bu nedenle Precision deęeri Odal için hesaplanamamıştır. Odal dışındaki odalar için 1466 tanesi doęru 534 yanlış sınıflandırma yapmıştır. Kappa deęeri de dięer algoritmalarla göre oldukça düşüktür. Bu nedenle çapraz doęrulama öncesinde yapılan sınıflandırmada en kötü performansı Hoeffding ağacı algoritmasının gösterdięi anlaşılmaktadır.

Çapraz doęrulama sonrasında Extra ağaç ve Rassal ağaç çapraz doęrulama öncesi gösterdikleri performansı gösteremedikleri için bu algoritmaların veri tabanında bulunmayan örnekleri sınıflandırmada iyi performans gösteremedikleri sonucuna varılabilir. Çapraz doęrulama sonrasında genel olarak en yüksek performans metrikleri rasgele orman algoritmasıdır. Bu metrik deęerleri sınıf bazında incelendiğinde Oda2 ve Oda3 sınıfına ait metrik deęerlerinin dięer sınıflardan daha düşük olduęu görülmektedir. Bu durum bu algoritmanın en çok Oda2 ve Oda3' te bulunan kişilerin tespitinde zorlandıęı ve bu odalarda bulunan bazı kişilerin yerini yanlış tespit ettięi anlamına gelmektedir. Oda1 ve Oda4 için kesinlik ve duyarlılık deęerleri 0,99 iken Oda2 ve Oda3 için bu deęerler düşüş göstermiştir. Oda2 de bulunan 19 kişiyi tespit edemedięinden duyarlılık deęeri 0,96 olmuştur. Konumunu Oda3 olduęunu tahmin ettięi 21 kişinin gerçekte farklı odalarda olması nedeniyle de kesinlik deęeri 0,95 olmuştur. Yapılan performans deęerlendirmesi sonucunda en iyi performansı Rasgele orman algoritması elde etmiştir. En kötü performans ise Hoeffding algoritmasının olmuştur. Çalışma genelinde hem çapraz doęrulama öncesi hemde çapraz doęrulama sonrası hesaplanan en yüksek metrik deęerleri Rasgele orman algoritmasıdır. Bu nedenle en başarılı algoritmanın bu olduęu sonucuna varılabilir. Algoritmanın çapraz doęrulama öncesi doęruluk oranı %100 çapraz

doğrulamadaki doğruluk oranı %98'dir. Hata oranları ise çapraz doğrulama öncesi 0.03, çapraz doğrulamada 0.08'dir. Çapraz doğrulama öncesi ve sonrası hata oranı en yüksek, doğruluk oranı ve diğer metrik değerleri en düşük olan algoritma Hoeffding ağacı algoritması olmuştur.

Sonuç olarak Kablosuz Sinyal Gücünü Kullanarak İç Mekan Kullanıcı Lokalizasyonu için Rasgele orman algoritmasının kullanımı önerilmektedir. Gelecek çalışmalarda veri sayısı ve oda sayıları artırılarak kişilerin konumlarının bulunması amaçlanmaktadır.

Hakem Değerlendirmesi: Dış bağımsız.

Çıkar Çatışması: Yazar çıkar çatışması beyan etmemiştir.

Finansal Destek: Yazar bu çalışma için finansal destek almadığını beyan etmiştir.

Peer-review: Externally peer-reviewed.

Conflict of Interest: The author has no conflict of interest to declare.

Grant Support: The author declared that this study has received no financial support.

Kaynaklar/References

- Abu Doush, I., Alshatnawi, S., Al-Tamimi, A.-K., Alhasan, B., ve Hamasha, S. (2017). ISAB: integrated indoor navigation system for the blind. *Interacting with Computers*, 29(2), 181-202.
- Akleylek, S., Kiliç, E., Söylemez, B., Aruk, T. E., & Çavuş, A. (2020). Kapalı mekan konumlandırma üzerine bir çalışma. *Mühendislik Bilimleri ve Tasarım Dergisi*, 8(5), 90-105.
- Arslantaş, H., & Ökdem, S. (2019). *İç Mekân Konumlandırma Yöntemleri*. Paper presented at the 4th International Symposium on Innovative Approaches in Engineering and Natural Sciences.
- Barsocchi, P., Cimino, M. G., Ferro, E., Lazzeri, A., Palumbo, F., ve Vaglini, G. (2015). Monitoring elderly behavior via indoor position-based stigmergy. *Pervasive and Mobile Computing*, 23, 26-42.
- Bozkurt, S., Elilob, G., Gunal, S., ve Yayan, U. (2015). *A comparative study on machine learning algorithms for indoor positioning*. Paper presented at the 2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA).
- Breima, L. (2010). Random Forests. Machine Learning.
- Breiman, L., Friedman, J., Olshen, R., ve Stone, C. (1984). Classification and regression trees—crc press. Boca Raton, Florida.
- Chen, R.-C., ve Huang, S.-L. (2009). *A new method for indoor location base on radio frequency identification*. Paper presented at the WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering.
- Correa, A., Llado, M. B., Morell, A., ve Vicario, J. L. (2016). Indoor pedestrian tracking by on-body multiple receivers. *IEEE Sensors Journal*, 16(8), 2545-2553.
- Çalık, S. H., & Gülgen, F. (2021). Artırılmış gerçeklik teknolojisi ile iç mekân navigasyonu. *Türkiye Coğrafi Bilgi Sistemleri Dergisi*, 3(1), 48-52.
- Çubukçu, A., Kuncan, M., Kaplan, K., & Ertunc, H. M. (2015). *Development of a voice-controlled home automation using Zigbee module*. Paper presented at the 2015 23rd Signal Processing and Communications Applications Conference (SIU).
- Dardari, D., Closas, P., ve Djurić, P. M. (2015). Indoor tracking: Theory, methods, and technologies. *IEEE Transactions on Vehicular Technology*, 64(4), 1263-1278.
- Frank, A. (2010). UCI machine learning repository. <http://archive.ics.uci.edu/ml>.
- Gama, J. (2004). Functional trees. *Machine learning*, 55(3), 219-250.
- Geurts, P., Ernst, D., ve Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1), 3-42.
- Gu, Y., Lo, A., ve Niemegeers, I. (2009). A survey of indoor positioning systems for wireless personal networks. *IEEE Communications surveys ve tutorials*, 11(1), 13-32.
- Homayounvala, E., Nabati, M., Shahbazian, R., Ghorashi, S. A., & Moghtadaiee, V. (2019). *A novel smartphone application for indoor positioning of users based on machine learning*. Paper presented at the Adjunct proceedings of the 2019 ACM international joint conference on pervasive and ubiquitous computing and proceedings of the 2019 ACM international symposium on wearable computers.
- Huang, X., Wang, F., Zhang, J., Hu, Z., ve Jin, J. (2019). A posture recognition method based on indoor positioning technology. *Sensors*, 19(6), 1464.
- Hulten, G., Spencer, L., ve Domingos, P. (2001). *Mining time-changing data streams*. Paper presented at the Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining.
- Joshi, R. (2016). Accuracy, precision, recall ve f1 score: Interpretation of performance measures. Retrieved April, 1(2018), 2016.
- Kohavi, R. (1996). *Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid*. Paper presented at the Kdd.
- Kuncan, M., & Ömer, Ç. (2019). Akıllı Ev Teknolojisi için Kablosuz Akıllı Kit. *Avrupa Bilim ve Teknoloji Dergisi*(17), 271-282.
- Landwehr, N., Hall, M., ve Frank, E. (2005). Logistic model trees. *Machine learning*, 59(1-2), 161-205.
- Lin, C.-J., Lee, T.-L., Syu, S.-L., ve Chen, B.-W. (2010). *Application of intelligent agent and RFID technology fo indoor position: Safety of kindergarten*

- as example. Paper presented at the 2010 International Conference on Machine Learning and Cybernetics.
- Mert, T., Ferdi, K., & Hakan, K. (2020). Dinamik Yapay Sinir Ağı ile İç Mekân Konum Kestirimi. *El-Cezeri Journal of Science and Engineering*, 7(2), 858-870.
- Quinlan, J. R. (1987). Simplifying decision trees. *International journal of man-machine studies*, 27(3), 221-234.
- Quinlan, R. (1993). 4.5: Programs for machine learning morgan kaufmann publishers inc. *San Francisco, USA*.
- Randell, C., ve Muller, H. (2001). *Low cost indoor positioning system*. Paper presented at the International Conference on Ubiquitous Computing.
- Rida, M. E., Liu, F., Jadi, Y., Algawhari, A. A. A., ve Askourih, A. (2015). *Indoor location position based on bluetooth signal strength*. Paper presented at the 2015 2nd International Conference on Information Science and Control Engineering.
- Rohra, J. G., Perumal, B., Narayanan, S. J., Thakur, P., ve Bhatt, R. B. (2017). *User localization in an indoor environment using fuzzy hybrid of particle swarm optimization ve gravitational search algorithm with neural networks*. Paper presented at the Proceedings of Sixth International Conference on Soft Computing for Problem Solving.
- Roy, P., & Chowdhury, C. (2021). A survey of machine learning techniques for indoor localization and navigation systems. *Journal of Intelligent & Robotic Systems*, 101(3), 1-34.
- Sabancı, K., Yigit, E., Ustun, D., Toktas, A., ve Aslan, M. F. (2018). *Wifi based indoor localization: application and comparison of machine learning algorithms*. Paper presented at the 2018 XXIIIrd International Seminar/Workshop on Direct and Inverse Problems of Electromagnetic and Acoustic Wave Theory (DIPED).
- Sáenz, M., ve Sánchez, J. (2009). *Indoor position and orientation for the blind*. Paper presented at the International Conference on Universal Access in Human-Computer Interaction.
- Seco, F., Jiménez, A. R., ve Zampella, F. (2013). *Joint estimation of indoor position and orientation from RF signal strength measurements*. Paper presented at the International Conference on Indoor Positioning and Indoor Navigation.
- Shi, H. (2007). *Best-first decision tree learning*. The University of Waikato,
- Shi, J. (2013). The Challenges of Indoor Positioning. *National University of Singapore: Singapore*.
- Srinivasan, D. B., ve Mekala, P. (2014). Mining social networking data for classification using REPTree. *International Journal of Advance Research in Computer Science and Management Studies*, 2(10).
- Su, H.-K., Liao, Z.-X., Lin, C.-H., ve Lin, T.-M. (2015). *A hybrid indoor-position mechanism based on bluetooth and WiFi communications for smart mobile devices*. Paper presented at the 2015 International Symposium on Bioelectronics and Bioinformatics (ISBB).
- Subhan, F., Hasbullah, H., ve Ashraf, K. (2013). Kalman filter-based hybrid indoor position estimation technique in bluetooth networks. *International Journal of Navigation and Observation*, 2013.
- Taşer, P. Y., & Akram, V. (2021). Kapalı ortamlarda gerçek zamanlı kişi tespitinde makine öğrenmesi algoritmalarının karşılaştırmalı başarımların analizi. *Academic Platform Journal of Engineering and Science*, 9(1), 182-193.
- Van Diggelen, F., ve Abraham, C. (2001). Indoor GPS technology. *CTIA Wireless-Agenda, Dallas*, 89.
- Wang, Z., Yang, Z., ve Dong, T. (2017). A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time. *Sensors*, 17(2), 341.
- Yasir, M., Ho, S.-W., ve Vellambi, B. N. (2015). Indoor position tracking using multiple optical receivers. *Journal of Lightwave Technology*, 34(4), 1166-1176.
- Zhang, H., ve Ye, C. (2020). *A visual positioning system for indoor blind navigation*. Paper presented at the 2020 IEEE International Conference on Robotics and Automation (ICRA).

