



## Kapalı Ortamlarda Kişi Tespitinde Makine Öğrenmesi Algoritmalarının Karşılaştırmalı Başarım Analizi

\*<sup>1</sup>Pelin Yıldırım Taşer, <sup>2</sup>Vahid Khalilpour Akram

<sup>1</sup> İzmir Bakırçay Üniversitesi, Mühendislik ve Mimarlık Fakültesi, İzmir, Türkiye,

pelin.taser@bakircay.edu.tr, 

<sup>2</sup> Ege Üniversitesi, Uluslararası Bilgisayar Enstitüsü, İzmir, Türkiye,

vahid.akram@ege.edu.tr, 

Araştırma Makalesi

Geliş Tarihi: 07.07.2020

Kabul Tarihi: 14.10.2020

### Öz

Günümüzde, iç mekan konumlandırma ve kişi takip sistemlerinin uygulama alanları her geçen gün artış göstermektedir. Özellikle, hasta, personel, cihaz ve müşteri takip sistemleri ile akıllı binalar ve kalabalık tahminleme gibi birçok alanda, kişilerin konumlarının veya mekan içerisinde bulunma durumlarının doğru tespiti büyük önem taşımaktadır. İç mekan konumlandırma sistemlerinde genellikle, hedef mobil varlığın üzerine periyodik olarak radyo sinyali gönderen küçük bir cihaz yerleştirilir ve bu cihazdan elde edilen sinyaller ile varlığın konumu belirlenir. Fakat bazı ortamlarda, üzerinde herhangi bir sinyal göndericisi taşımayan varlıkların konumlarının tespit edilmesine ihtiyaç duyulmaktadır. Dolayısıyla, mobil cihaz kullanılmayan radyo tabanlı takip sistemlerinde, radyo sinyallerinde meydana gelen dalgalanmalar analiz edilerek, ortamdaki hareketlilik tahmin edilmeye çalışılır. Bu sistemlerde, ortamın çeşitli noktalarına periyodik olarak radyo sinyalleri gönderen ve diğer cihazların gönderdiği sinyalleri alabilen cihazlar yerleştirilir. Ortamda bulunan herhangi bir nesnenin hareket etmesi durumunda, sinyal gücündeki dalgalanmalar analiz edilerek, ortamdaki hareketlilik ve yoğunluk tahmin edilebilir. Ancak bazı durumlarda, radyo sinyallerinde hareketten kaynaklanmayan, geçici ama nispeten şiddetli dalgalanmalar yaşanabilmektedir. Bu tür dalgalanmalar, yanlış tespitlere sebep olarak, sistemin hassasiyetini ve doğruluğunu düşürmektedir. Makine öğrenmesi tekniklerinin, veriler arasındaki gizli örüntü ve karmaşık ilişkileri ortaya çıkarmadaki başarıları sayesinde, makine öğrenmesi tekniklerine dayalı kişi tespit sistemleri, geleneksel yöntemlere göre doğruluğu daha yüksek tahminleme becerisi sunmaktadır. Dolayısıyla, bu çalışmada, kapalı ortamda kişi tespiti için makine öğrenmesi algoritmalarından yararlanılmıştır. Deneysel çalışmalar kapsamında, 10 farklı geleneksel (Naive Bayes, Çok Katmanlı Algılayıcı (MLP), Destek Vektör Makineleri (SVM) ve K-en Yakın Komşuluk (K-NN)) ve karar ağacı tabanlı (C4.5, Random Forest, Random Tree, REPTree, Decision Stump ve HoeffdingTree) sınıflandırma algoritmaları, kapalı ortamdaki 3 farklı telsiz duyurga düğümünden elde edilen ve 23585 kayıttan oluşan veri seti üzerinde ayrı ayrı uygulanmış, doğruluk oranı ve model oluşturma süresi performanslarına göre karşılaştırılmıştır. Elde edilen sonuçlar incelendiğinde, çalışmada uygulanan tüm algoritmaların kapalı alandaki kişi tespitinde %80'in üzerinde başarı performansı sunduğu ve en başarılı algoritmanın %99.68 doğruluk oranı ile Random Forest olduğu gözlemlenmiştir. Ayrıca, geleneksel ve karar ağacı tabanlı sınıflandırma algoritmaları sağlamış oldukları ortalama doğruluk oranlarına göre kıyaslandığında ise karar ağacı tabanlı algoritmaların %95.78 ile daha yüksek tahminleme becerisi sunduğu görülmektedir.

**Anahtar Kelimeler:** İç Mekan Kişi Tespiti, Makine Öğrenmesi, Alınan Sinyal Gücü Göstergesi, Sınıflandırma, Telsiz Duyurga Ağları.

## Comparative Performance Analysis of Machine Learning Algorithms for Human Detection in Indoor Environments

\*<sup>1</sup>Pelin Yıldırım Taşer, <sup>2</sup>Vahid Khalilpour Akram

<sup>1</sup> İzmir Bakırçay University, Faculty of Engineering and Architecture, Izmir, Turkey, pelin.taser@bakircay.edu.tr

<sup>2</sup> Ege University, International Computer Institute, Izmir, Turkey, vahid.akram@ege.edu.tr

### Abstract

Nowadays, the application areas of indoor positioning and person tracking systems are increasing day by day. Especially, in many fields such as patient, personnel, device and customer tracking systems and smart buildings and crowd forecasting, it is

\*<sup>1</sup>Sorumlu Yazar: İzmir Bakırçay Üniversitesi, Mühendislik ve Mimarlık Fakültesi, İzmir, Türkiye, pelin.taser@bakircay.edu.tr, 05544758645

Doi: 10.21541/apjes.765795

crucial to determine the location of humans or their presence in the environment. Generally, in the indoor tracking systems, a small device, that periodically sends a radio signal, is attached to the mobile target asset and the location of the asset is determined with the signals obtained from this device. However, in some environments, it is necessary to determine the location of assets that do not carry any signal transmitters on them. Hence, in the systems without mobile devices, the mobility in the environment is estimated by analyzing fluctuations in the radio signals. In these systems, there are some static devices at various points in the environment that send radio signals periodically and can receive signals sent by other devices. If an object moves in the environment, the mobility and density of mobile objects in the environment can be detected by analyzing the fluctuations in signal strength. However, in some cases, temporary but relatively violent fluctuations, which are not related to any movement, may occur in radio signals. These fluctuations reduce the sensitivity and accuracy of the system by causing false detections. Thanks to the success of machine learning techniques in revealing hidden patterns and complex relations among data, human detection systems based on machine learning techniques offer higher accuracy estimation ability than traditional methods. Therefore, in this study, machine learning algorithms are utilized for human detection in indoor environment. Within the experimental studies, 10 different traditional (Naive Bayes, Multilayer Perceptron (MLP), Support Vector Machines (SVM), and K-nearest Neighbors (K-NN)) and decision tree-based (C4.5, Random Forest, Random Tree, REPTree, Decision Stump, and HoeffdingTree) classification algorithms were applied separately on a data set consisting of 23585 records, obtained from 3 different wireless sensor nodes in the indoor environment, and were compared according to accuracy rate and model build time performances. When the obtained results are examined, it is observed that all algorithms applied in the study offer over 80% success performance in the detection of human in indoor environments and the most successful algorithm is Random Forest with an accuracy rate of 99.68%. In addition, when the traditional and decision tree-based classification algorithms compared according to the average accuracy rates they provided, it is seen that decision tree-based algorithms present higher estimation ability with 95.78%.

**Keywords:** Indoor Human Detection, Machine Learning, Received Signal Strength Indicator, Classification, Wireless Sensor Networks.

## 1. GİRİŞ

Günümüzde dış mekân konumlandırma ve takip sistemlerinde meydana gelen gelişmeler, bu sistemlerin hassasiyet ve doğruluk oranlarını oldukça arttırmıştır. Ancak, dış mekân konumlandırma sistemlerinin hemen hepsinin GPS sistemine dayalı olması ve GPS sisteminin kapalı ortamlarda çalışmamasından dolayı, bu sistemlerin kapalı ortamlardaki başarılı konumlandırma oranları oldukça sınırlıdır. Diğer taraftan, iç mekân konumlandırma ve takip sistemlerinin uygulama alanları her geçen gün artmaktadır. Kapalı ortamlarda çeşitli varlıkların anlık konum bilgileri, arama, izleme, kontrol, otomasyon ve güvenlik gibi çeşitli işlemlerin verimliliğini arttırabilmektedir. Örneğin, hastanelerde el hijyen izleme sistemlerinde, sağlık çalışanlarının anlık konum bilgilerinin tutulması ile hijyen durumları kontrol edilebilir. Ayrıca, alışveriş merkezlerinde, müşterilerin konumları ve takip ettikleri güzergah bilgileri üzerinden, reklam ve pazarlama stratejileri belirlenerek, satış verimliliği arttırılabilir. Bir diğer uygulanabilir örnek olarak güvenlik sistemlerinde ise, izinsiz giriş, yasak bölge ihlali, ve kalabalık kontrolü gibi çeşitli işlemlerde, iç mekan konumlandırma sistemleri önemli derecede yarar sağlayabilir. Bunların yanı sıra, akıllı binalar, depolar, otoparklar, şantiyeler, fabrikalar ve oteller gibi bir çok benzer uygulama alanlarında da, iç mekan kişi ve hareket tespiti önemli avantajlar sağlamaktadır [1]. Bu motivasyonlar dikkate alındığında, önerilen çalışmada, makine öğrenmesi teknikleri kullanılarak, kapalı ortamda cihazsız kişi tespitinin gerçekleştirilmesi amaçlanmıştır. Ayrıca, geleneksel sınıflandırma algoritmalarının yanı sıra, karar ağacı tabanlı sınıflandırma algoritmalarının uygulanması ile kişi tespit sisteminin doğruluğunun artırılması hedeflenmiştir.

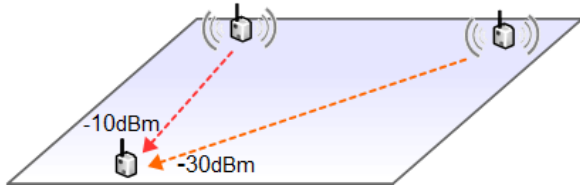
Kapalı alanlarda kişilerin takibi ve konumlandırılması için Telsiz Duyarga Ağları (TDA), RFID, Bluetooth ve WiFi gibi çeşitli teknolojiler kullanılmaktadır [1,2]. Genel olarak bu teknolojilerin çoğu, hedef varlık tarafından gönderilen radyo sinyallerinin gücünü ölçerek varlığın konumunu belirlemektedir. Dolayısıyla bu yöntemlerde, periyodik olarak hedef varlıktan bir radyo sinyali gönderilmektedir. Bu işlemi gerçekleştirmek için, takip edilmesi gereken varlıkların üzerine bir radyo sinyal göndericisi yerleştirilmektedir. Bu sinyal göndericisi bir TDA düğümü, bir RF etiketi veya bir Bluetooth sinyal göndericisi olabilir. Örneğin, günümüzde kullanılan birçok konum takip sisteminde kişilerin cep telefonlarından gönderilen Bluetooth sinyallerinden yararlanılmaktadır [3,4]. Kişilerin üzerinde bulunan bir radyo göndericisine ihtiyaç duyan bu sistemlerin en büyük dezavantajı, cihaz taşımayan kişilerin takip edilememesidir.

Birçok durumda, ortamda üzerinde herhangi bir sinyal göndericisi taşımayan kişilerin veya hareket eden varlıkların konumlarının tespit edilmesine ihtiyaç duyulmaktadır. Örneğin hasta, müşteri, personel ve cihaz takip sistemlerinde [4,5], güvenlik ihlali tespit sistemlerinde [6], kalabalık tahminleme uygulamalarında [7,8], arama kurtarma sistemlerinde [9-11] ve enerji tasarruflu akıllı binaların tasarımında [12], ortamdaki kişi veya hareketliliğin tespiti gerekmektedir.

Radyo tabanlı cihazsız hareket takip sistemleri, genel olarak ortamda yayılan sinyallerin sarsılmasını baz almaktadır. Ortamda meydana gelen herhangi bir hareketlilik, radyo sinyallerinin yayılmasını etkileyip, alıcı cihaz tarafından alınan sinyalin gücünü değiştirebilir. Genel olarak, bir alıcı

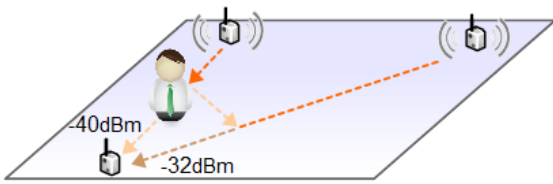
cihazın aldığı sinyalin gücüne RSSI (Received Signal Strength Indicator) denir.

Ortamda bulunan herhangi bir yeni varlık, sinyallerin önünde engel teşkil ederek, sinyallerin gücünü azaltır veya sinyallerin yansımalarına sebep olur. Dolayısıyla, ortamda meydana gelen bir hareketlilik, yayılmakta olan tüm radyo sinyallerini etkileyebilir. Örneğin, Şekil 1’ de gösterilen odada, iki radyo sinyal göndericisi ve bir alıcı cihaz bulunmaktadır. Bu şekilde, göndericiler ve alıcının arasında herhangi bir varlık bulunmamaktadır. Bu sebepten dolayı, alınan radyo sinyallerinin gücü, ağırlıklı olarak, alıcıyla göndericinin arasındaki mesafeden etkilenmektedir. Bir radyo sinyali, göndericiden uzaklaşırken gücünü kaybeder. Dolayısıyla, yakın bir göndericiden alınan sinyalin RSSI değeri, daha uzakta bulunan göndericinin sinyalinden daha yüksek olabilir.



Şekil 1. Herhangi bir varlık bulunmayan kapalı ortamda RSSI değerleri.

Ortamda bir varlığın bulunması veya bu varlığın hareket etmesi, tüm sinyallerin RSSI değerlerini etkileyebilir. Örneğin; Şekil 2’de ortamda bir kişinin bulunması, bazı sinyallerin RSSI değerlerini doğrudan azaltabilir (gönderici ve alıcının arasında bir engel oluşturarak) veya yansımalarından dolayı diğer sinyalleri etkileyebilir. Kişinin odada yer değiştirmesi, alıcı tarafından alınan RSSI değerlerini doğrudan (bir engel oluşturarak) veya dolaylı yoldan (sinyallerin yansımalarından dolayı) değiştirebilir.



Şekil 2. Bir kişi bulunan kapalı ortamda RSSI değerleri.

Fakat bazı durumlarda ortamda herhangi bir varlık veya hareketlilik olmasa bile, radyo sinyallerinin yansımaları, ortamın sıcaklığı veya göndericinin enerji seviyesinin değişmesi gibi çeşitli sebeplerden dolayı, alınan RSSI değerlerinde değişiklik olabilir. Çoğu durumda bu değişiklik sınırlıdır, fakat sarsıntılarının büyüklüğü istatistik, eşik değer veya filtreleme tabanlı geleneksel kişi tespit sistemlerini yanıltarak, yanlış tespitlere sebep olabilmektedir. Makine öğrenmesi tekniklerinin, veriler arasındaki gizli örüntü ve karmaşık ilişkileri ortaya çıkarmadaki başarıları sayesinde, ortamda bulunan herhangi bir varlık veya hareketlilik dışı sebeplerden oluşan RSSI değeri değişimleri kolayca ayırt edilebilmektedir. Böylece, makine öğrenmesi tekniklerine

dayalı kişi tespit sistemleri, geleneksel yöntemlere nazaran doğruluğu daha yüksek tahminleme becerisi sunmaktadır.

Kapalı ortamda cihazsız kişi veya hareket tespit sistemlerini, görüntü ve radyo tabanlı olmak üzere iki ana gruba ayırmak mümkündür. Görüntü tabanlı sistemler, kameradan elde edilen görüntüleri kullanarak, arka plan resmini çeşitli yöntemlerle algılar ve fotoğrafta bulunan diğer varlıkların hareketini tespit eder [7]. Görüntü tabanlı sistem içerisinde gerekli olan çok sayıda hesaplama işlemlerinin, hızlı bir şekilde gerçekleştirilmesi gerekmektedir. Dolayısıyla bu sistemler, güçlü işlemci, yüksek bellek ve geniş iletişim bandına ihtiyaç duymaktadır. Ayrıca, yeterince ışığın bulunmadığı veya arka planın değişebildiği ortamlar da, bu sistemlerin başarı oranını oldukça düşürecektir.

Literatürde, kapalı ortamlarda RSSI değerleri üzerinden hareket veya kişi tespiti gerçekleştiren birçok çalışma bulunmaktadır. Örneğin; Parkinson hastalarının hareketini ve evin hangi bölgelerinde bulduklarını tespit eden RSSI tabanlı bir yöntem geliştirilmiştir [5]. Önerilen yöntemde RSSI değerleri belli bir seviyenin altına indiğinde, ilgili bölgede sinyalleri engelleyen bir varlığın olduğu tespit edilir. TDA’larda RSSI değerlerinin dalgalanması üzerinden ortamdaki hareketlilik veya kişi tespiti için eşik değerle karşılaştırma [9,10], uyarlamalı filtreleme [11] ve zaman diliminde değişim oranlarının hesaplanması [12] gibi çeşitli yöntemler kullanılmaktadır. Ancak bu yöntemler, anlık veriler üzerinden karar verme işlemi gerçekleştirdikleri için, oluşan geçici ve sebepsiz dalgalanmalar sistemi yanıltabilecektir.

Kapalı alandaki kalabalığın tespiti için geliştirilen çalışmaların birinde [8], RSSI ve LQI (Link Quality Indicator) değerlerine dayalı pasif bir tahminleme sistemi gerçekleştirilmiştir. Bu çalışmada, ortamda kalabalığın artması sebebiyle düğümlerin ortalama RSSI ve LQI değerlerinin düştüğü varsayılarak, alınan ortalama RSSI ve LQI değerleri üzerinden kişi sayısı tahmin edilmektedir.

Kapalı ortamlarda TDA tabanlı izinsiz kişi girişi tespitinde, RSSI değerlerinin dalgalanmasına dayalı çeşitli çalışmalar bulunmaktadır [13-20]. Bu dalgalanmalar, dağıtık hesaplama, kanal durum denetleme ve derin göç öğrenim gibi yöntemlerle tespit edilmektedir. Fakat hareketten kaynaklanmayan dalgalanmalardan dolayı, tüm bu yöntemlerin başarısı sınırlı kalmaktadır.

Literatürde, makine öğrenmesi tekniklerinden faydalanan kalabalık tahmin sistemi çalışmaları mevcuttur [21-23]. Örneğin; Yoshida ve Taniguchi [21], odada bulunan kişi sayısını, TDA düğümleri yerine WiFi sinyalleri üzerinden, doğrusal regresyon ve destek vektör regresyon yöntemleri kullanarak tahmin etmektedir. Benzer bir diğer çalışmada da [22], TDA düğümlerinden gelen RSSI değerleri üzerinde K-means kümeleme algoritması kullanılarak, kalabalık yoğunluğu ortaya çıkarılmaktadır. Bunların yanı sıra, hedef varlık tarafından gönderilen radyo sinyalleri aracılığıyla, varlığın konumunu belirleyen makine öğrenmesi tabanlı

çalışmalar da bulunmaktadır [24,25]. Ancak bu yöntemlerde radyo sinyali göndermeyen varlıklar takip edilemez.

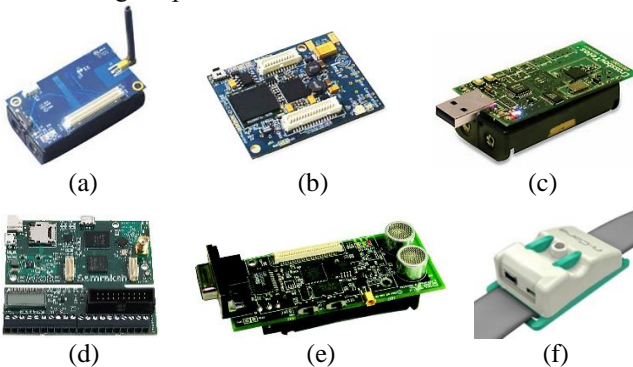
Son yıllarda WiFi sinyalleri ve makine öğrenmesi tekniklerini kullanan cihazsız takip sistemleri yaygın bir şekilde kullanılmaya başlanmıştır [26,27]. Örneğin; Huang ve Lin [26], derin evrişimli sinir ağı kullanarak WiFi sinyalleri üzerinden kişi tespiti sistemi geliştirmişlerdir. Mevcut çalışmalardan farklı olarak bu çalışmada, WiFi sinyalleri yerine TDA düğümleri tarafından gönderilen veriler üzerinde 10 farklı geleneksel (Naive Bayes, MLP, SVM ve K-NN) ve karar ağacı tabanlı (C4.5, Random Forest, Random Tree, REPTree, Decision Stump ve HoeffdingTree) sınıflandırma algoritmaları uygulanarak, kapalı ortamda cihazsız kişi tespiti gerçekleştirilmiştir.

## 2. PROBLEM TANIMI VE ÖN BİLGİ

### 2.1. Telsiz Duyarga Ağları

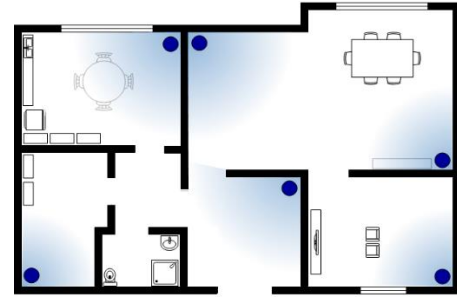
TDA'larda düğümler, algıladıkları, ışık, sıcaklık, nem, rüzgâr ve kimyasal gazlar gibi çeşitli bilgileri, radyo mesajları aracılığıyla bir baz istasyonuna gönderirler. Baz istasyonu, gelen verileri bilgisayara aktararak, bu verilerin işlenmesini ve üzerinde gereken işlemlerin yapılmasını sağlar. Günümüzde, TDA'lara özel çeşitli sensör ve donanımlar üretilmektedir. TDA düğümleri, genel amaçlı veya özel bir uygulamaya yönelik tasarlanabilir. Örneğin, Iris (Şekil 3.a), Imote2 (Şekil 3.b), Telosb (Şekil 3.c) ve Now (Şekil 3.d) gibi düğümlerin üzerinde çeşitli sensörler veya geliştirme slotlarının bulunması, bu düğümlerin farklı uygulamalarda kullanılabileceğini göstermektedir. Cricket (Şekil 3.e) veya Sirius Quantum (Şekil 3.f) gibi düğümler ise özel amaçlar için tasarlanmışlardır.

TDA düğümleri tarafından gönderilen radyo mesajlarında, sinyalin gücü ortamda ilerledikçe veya engellerden geçerken azalır. Ortamda insan vücudu gibi hareketli engellerin bulunması, yayılan sinyallerin gücünü etkiler ve böylece ortamdaki hareketli varlık tespit edilebilir. Dolayısıyla, kapalı bir ortamın farklı noktalarına TDA düğümleri yerleştirilip, bu düğümlerden periyodik olarak radyo sinyalleri gönderilirse, alınan sinyallerin gücünde meydana gelen dalgalanmalardan dolayı ortamda bir kişinin bulunup bulunmadığı tespit edilebilir.



Şekil 3. a) Iris, b) Imote2, c) Telosb, d) Now, e) Cricket, f) Sirius Quantum düğümleri.

Örneğin, Şekil 4'te gösterilen bir kat planında, her odada en az bir TDA düğümü konumlandırılmıştır. Böylece eğer bir odanın düğümlerinden gelen sinyallerde normalin üzerinde bir dalgalanma olursa, o odada en az bir hareketli varlığın bulunduğu varsayılabilir. Düğümlerin radyo menzili ve odaların büyüklüğü göz önüne alınarak, her odaya yeterli sayıda düğüm yerleştirilmelidir. Az sayıda TDA düğümünün kullanılması, bazı bölgelere sinyallerin yeterli ölçüde ulaşmamasına sebep olarak, tespitin doğruluk oranını düşürebilir.



Şekil 4. TDA düğümlerinin örnek ortamda konumlandırılması.

Genel olarak, TDA'larda her düğüme benzersiz bir kimlik numarası verilir. Düğümler radyo mesajı gönderirken, kimlik numaralarını da mesaja eklerler. Böylece alıcı düğüm, mesajın kimden geldiğini tespit edebilir. Bir düğüm radyo mesajı alırken, mesajın içindeki bilgilerin yanı sıra, mesajı taşıyan sinyalin gücünü de ölçebilir. Gönderici ile alıcının konumları sabit ve ortamda hareketli bir varlık yoksa, alınan mesajların gücündeki değişiklik sınırlı olur. Ancak ortamda hareketli bir varlık bulunuyorsa, varlığın, göndericinin ve alıcının konumlarına göre alınan mesajların gücünde şiddetli dalgalanmalar oluşabilir. Bu nedenle ortamda daha fazla düğümün bulunması, hareket eden varlıkların daha doğru tespit edilmesini sağlar.

Baz istasyonunda, alınan mesajların kimlik numarasından, mesajın gönderildiği düğümün konumu belirlenir. Daha sonrasında ise, mesajın RSSI değeri incelenerek, gönderici düğümün yakınında yeni bir varlığın olup olmadığı belirlenir. Böylece, odadaki bir kişinin varlığı tespit edilebilir. Bu çalışmada Iris düğümleri kullanılarak, ortamın ışık seviyesi ve gönderilen mesajların RSSI değerleri ölçülmüştür. Bu veriler ile makine öğrenmesi algoritmaları eğitilerek, ortamda bir kişinin olup olmadığı tespit edilmiştir.

### 2.2. Makine Öğrenmesi

Makine öğrenmesi, sistemlere herhangi bir insan müdahalesine gerek olmadan öğrenme becerisi kazandırmayı konu edinen, yapay zekânın bir alt dalıdır [28]. Bu disiplindeki amaç, matematiksel ve istatistiksel yöntemler kullanılarak büyük miktardaki ham verinin işlenmesi ve bu veriler içerisinden tahminler yapılmasını sağlayan modellerin oluşturulmasıdır. Günümüzde makine öğrenmesi teknikleri iki temel başlık altında toplanmaktadır: denetimli öğrenme ve denetimsiz öğrenme. Denetimli

öğrenme sınıflandırma ve regresyon metotlarını içerirken, denetimsiz öğrenme tekniğinde ise kümeleme ve birliktelik kuralı analizi yöntemleri uygulanmaktadır. Bu çalışma kapsamında gerçekleştirilen kapalı ortamlarda kişi tespitinde sınıflandırma tekniği uygulanmıştır.

Denetimli öğrenme yönteminde sistem, daha önceden etiketlenmiş (sınıf etiketi bilinen) gözlem verileri ile eğitilir. Sonrasında ise eğitilen bu sistem, mevcut gözlem verilerine bakarak, sınıf etiketi bilinmeyen örneklemin etiket değerini tahmin etmeye çalışır. Sınıflandırma, çıktı değeri (sınıf etiketi) bilinmeyen bir örneklemin belirli sınıflardan birine dâhil etme işlemi gerçekleştiren, makine öğrenmesinde en yaygın uygulanan denetimli öğrenme yöntemlerinden birisidir. Literatürde yaygın olarak kullanılan sınıflandırma algoritmaları, Naive Bayes, Yapay Sinir Ağları, K-en Yakın Komşuluk, Destek Vektör Makineleri ve Karar Ağacıdır. Bu çalışmada, yaygın kullanılan geleneksel sınıflandırma algoritmalarının yanı sıra, karar ağacı tabanlı sınıflandırma algoritmaları da deneysel çalışmalar kapsamında uygulanmıştır.

### Geleneksel Sınıflandırma Algoritmaları

**Naive Bayes:** Temeli Bayes teoremine dayanan, sınıflandırılacak örneklemin her bir özneliğindeki değerlerin eğitim verilerindeki olasılıklarını hesaplayıp, sınıf etiketini olasılığı en yüksek değer olarak seçen istatistiksel bir sınıflandırma algoritmasıdır [29]. Algoritmanın en büyük avantajı, dengesiz ve az sayıdaki veri setleri üzerinde de yüksek doğruluk oranına sahip sınıflandırma performansı sergilemesidir.

**Yapay Sinir Ağları (YSA):** İnsan beyninin biyolojik sinir sisteminden esinlenilerek oluşturulmuş, örneklemin girdi değerlerini işleyerek sınıf etiketini tahminlemeye çalışan bir sınıflandırıcıdır [30]. YSA'da biyolojik sinir sistemindeki nöron, dentrit, çekirdek, akson ve sinaps yapıları, sırasıyla girdi, toplama fonksiyonu, aktivasyon fonksiyonu, çıktı ve ağırlık olarak modellenmektedir. Literatürde birçok YSA algoritmaları bulunmaktadır: tek katmanlı algılayıcılar (TKA), basit algılayıcılar, Adaline, çok katmanlı algılayıcılar, ileri beslemeli ağlar ve geri beslemeli ağlar. Bu çalışma kapsamında çok katmanlı algılayıcı (MLP) modeli uygulanmıştır.

MLP en popüler ileri beslemeli YSA modellerinden birisidir. MLP algoritması, örneklemin özniteliklerini, birbirine bağlı birden fazla katmandan oluşan ağırlıklı bir ağ yapısına girdi olarak gönderir ve sonucunda da çıktı değerini örneklemin sınıf etiketi olarak kabul eder.

**K-en Yakın Komşuluk (K-NN):** K-NN algoritması, örneklemin sınıf etiketi belirlenirken ilk olarak örnekleme en benzer  $k$  (kullanıcı tanımlı sabit değer) eğitim verisini belirler [31]. Bu verilerin belirlenmesinde, literatürdeki mevcut uzaklık ölçütleri kullanılmaktadır. Örneğin; öznitelik türü sayısal olan veriler için Euclidean, Manhattan ve Minkowski ölçütleri kullanılırken, kategorik değerlerde ise

Hamming ölçütü uygulanmaktadır. K en yakın kayıdın belirlenmesinin ardından, bu kayıtların sınıf etiketlerinin çoğunluğu ile örneklem verisi sınıflandırılır. K-NN algoritması hem kolay uygulanabilir hem de kolay anlaşılabilir olmasından dolayı sıklıkla tercih edilen makine öğrenmesi algoritmalarından birisidir.

**Destek Vektör Makineleri (SVM):** SVM istatistiksel öğrenme teorisi, karar düzlemleri ve risk minimizasyonu kavramlarını temel alan ikili bir sınıflandırma algoritmasıdır [32]. SVM algoritmasında, eğitim verisindeki her bir kayıt  $n$  (öznitelik sayısı) boyutlu uzayda birer nokta olarak ele alınır ve bu uzay üzerindeki noktaları iki farklı sınıfa ayırmak için bir hiper düzlem oluşturulur. Algoritmanın amacı oluşturulan hiper düzlemin en büyük marjine sahip olmasını sağlamaktır. Böylece sistemin aşırı ezberlemesi ya da düşük sınıflandırma performansı göstermesinin önüne geçilebilmektedir.

**Karar ağacı:** Karar ağacı, eğitim veri setlerinden ağaç formunda bir model oluşturarak, örneklemelerin sınıf etiketlerini belirleme işlemi gerçekleştiren bir sınıflandırıcıdır. Karar ağacı herhangi bir girdi almayan kök düğüm ve girdi alan iç düğümlerden meydana gelir. Bir düğümün çıktısı bir diğer düğüme girdi olarak aktarılır. Girdiler ve çıktılar ağaç kavramında dal olarak adlandırılır. Ağaç üzerindeki bir düğümün çıktısı başka bir düğüme girdi olarak aktarılmıyorsa, o düğüme yaprak düğüm adı verilmektedir. Ağaç yapısındaki düğüm, dal ve yaprak kavramları veri setinde sırasıyla öznitelik, öznitelik değerleri ve sınıf etiketlerini sembolize etmektedir.

Karar ağaçları, oluşturulması ve yorumlanması kolay algoritmalar olmalarından dolayı birçok sınıflandırma çalışmasında sıklıkla kullanılmaktadırlar [33]. Dışarıdan herhangi bir parametre almıyor oluşu ve hem kategorik hem de sayısal değerlerin tahminlenmesinde kullanılabilir olması ise, karar ağaçlarının bir diğer avantajlarıdır. Karar ağaçlarını diğer sınıflandırıcılardan ayıran en önemli özellik, bir örneklemin sınıf etiketi tahmin edilirken hangi özneliğin daha belirleyici olduğunun da ortaya çıkmasıdır.

Gözetimli öğrenme alanında karar ağacı tabanlı birçok sınıflandırma algoritması bulunmaktadır: C4.5, Random Forest, Random Tree, REPTree, Decision Stump ve HoeffdingTree. Bu algoritmalar, makalede gerçekleştirilen deneysel çalışma kapsamında uygulanmış ve diğer geleneksel sınıflandırma algoritmaları ile kıyaslanmıştır.

### Karar Ağacı Tabanlı Sınıflandırma Algoritmaları

**C4.5:** C4.5, Ross Quinlan tarafından geliştirilen ve en yaygın uygulanan karar ağacı algoritmalarından birisidir [34]. Bu algortmada eğitim verisi ile ağaç modeli oluşturulurken, veri seti içerisindeki en ayırt edici öznitelik belirlenir ve kök düğüme yerleştirilir. Bu özneliğin belirlenmesinde ise information gain (bilgi kazanımı) ölçütü kullanılır.  $S$  eğitim veri seti,  $S_v$  örnek uzayın alt kümesi,  $A$  öznitelik ve  $v$



özniteliğin olası değerleri olmak üzere, information gain (IG) formülü aşağıdaki şekilde tanımlanmıştır:

$$IG(S, A) = Entropy(S) - \sum_{v \in A} \frac{|S_v|}{|S|} Entropy(S_v) \quad (1)$$

Entropy değeri ise, veri setindeki herhangi bir özniteliğin homojenliğini belirlemek için kullanılır.  $c$  sınıf sayısı,  $p_i$   $i$ . sınıfa ait olasılık olmak üzere, Entropy formülü aşağıdaki şekilde hesaplanmaktadır:

$$Entropy(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (2)$$

Kök düğümden itibaren ağacın dallandığı karar noktalarının (düğümlerin) belirlenmesinde de yine yüksek information gain değerine sahip öznitelik üzerinden ilerlenir. Algoritmanın en büyük avantajı, eksik öznitelik değerlerine sahip kayıtları içeren veri setleri üzerinde de çalışabilir olmasıdır.

**Random Forest:** Random Forest, Leo Breiman and Adele Cutler tarafından geliştirilen bir karar ağacı tabanlı sınıflandırma algoritmasıdır [35]. Random Forest algoritmasında, eğitim veri setindeki kayıtların rastgele seçilmesinden elde edilen birden fazla sayıda karar ağacı modeli oluşturulur. Sınıf etiketi tahmin edilecek örneklem ormandaki her bir karar ağacı modeline girdi olarak verilir ve modellerin üretmiş oldukları çıktı değerlerinin çoğunluğuna bakılarak, örnekleme sınıf etiketi atanır. Random Forest algoritması ile tek bir sınıflandırma modeli yerine birden fazla sınıflandırma modeli kullanarak sınıflandırma performansı yüksek sonuçlar elde edilebilmektedir.

**Random Tree:** Random Tree algoritması, bagging (torbalama) yöntemi kullanarak eğitim veri setindeki rastgele veriler ile birden fazla ağaç modeli oluşturur [36]. Eşit şansa sahip tüm ağaçlar içerisinde rastgele bir ağaç seçilir.

**REPTree:** REPTree algoritmasında, regresyon ağacı veya karar ağaçları oluşturulurken information gain ölçütü kullanılır [37]. Oluşturulan bu ağaçlar arasında en iyi performansa sahip olan seçilir. Seçilen ağaç hata budaması işlemine tabi tutulur.

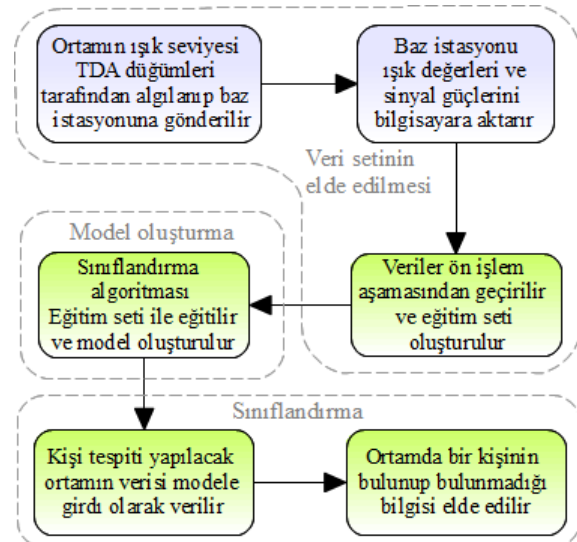
**Decision Stump:** Decision Stump, sadece kök düğüm ve yapraklardan oluşan tek seviyeli bir karar ağacı tabanlı sınıflandırıcıdır [38]. Örneklemlerin sınıflandırılması işlemi sadece tek bir öznitelige bakılarak değerlendirilir. Algoritma yaygın olarak Boosting topluluk öğrenmesi metodunun temel sınıflandırıcısı olarak kullanılır.

**HoeffdingTree:** Domingos and Hulten tarafından 2000 yılında geliştirilen Hoeffding Tree algoritması, veri setindeki her bir kayıdı sadece bir kez işlemesinden dolayı, özellikle büyük veri setlerinde hızlı çalışan bir karar ağacı algoritmasıdır [39]. Algoritma, geleneksel topluluk öğrenmesi yöntemlerine benzer şekilde karar ağaçları

oluşturur. Ağaç üzerindeki karar noktaları Hoeffding sınırı isimli matematiksel değer ile belirlenir.

### 3. DENEYSEL ÇALIŞMA

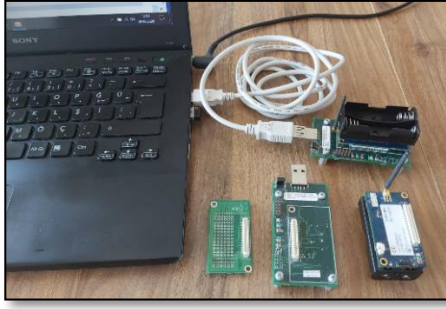
Bu makalede sunulan çalışma kapsamında, 10 farklı sınıflandırma algoritması (Naive Bayes, MLP, SVM, K-NN, C4.5, Random Forest, Random Tree, REPTree, Decision Stump ve HoeffdingTree) kapalı alanda kişi tespiti amacıyla uygulanmış ve göstermiş oldukları doğruluk oranı ve model oluşturma süresi performanslarına göre karşılaştırılmıştır. Uygulanan sınıflandırma modellerinin doğruluğu 10-katlı çapraz geçerlilik tekniği kullanılarak elde edilmiştir. Deneysel çalışmalar kapsamında geliştirilen sınıflandırma uygulamasında, Weka açık kaynak kodlu veri madenciliği kütüphanesi kullanılmıştır [40]. Deneyler, Intel Core i7-7500U 2.90 GHz CPU 8 GB RAM sistem özelliklerine sahip bir kişisel bilgisayar üzerinde gerçekleştirilmiştir. Çalışma kapsamında geliştirilen sistemin genel iş akış diyagramı Şekil 5' te sunulmuştur.



Şekil 5. Geliştirilen sistemin genel iş akış diyagramı.

#### 3.1. Test Yatağı

Deneysel çalışma kapsamında, kapalı alanda kişi tespiti için ortamın ışık seviyesi ve gönderilen paketlerin RSSI değerleri, Iris düğümleri [41] ile ölçülmüştür. Iris düğümleri 2.4 GHz IEEE 802.15.4 radyo göndericisi, 128 kB flaş bellek, 8 kB RAM bellek ve bir Atmel ATmega1281 işlemcisine sahiptir. TinyOS [42] işletim sistemini destekleyen Iris düğümleri üzerinde, herhangi bir sensör bulunmamaktadır. Ancak düğümün üzerindeki 51 pinlik geliştirme slot aracılığıyla farklı sensörlere bağlanabilir. Örneğin; MDA100, ışık ve sıcaklık sensörü bu slota takılabilir. Iris düğümleri, MIB520 geçit donanımı ile bilgisayarın USB girişine bağlanıp, bir baz istasyonu olarak radyodan gelen paketleri bilgisayara aktarabilirler. Şekil 6'da, çalışmada kullanılan Iris, MIB520, MDA100 donanımları ve Iris düğümünün MIB520 ile birlikte bilgisayara bağlı olduğu (baz istasyonu) görülmektedir.



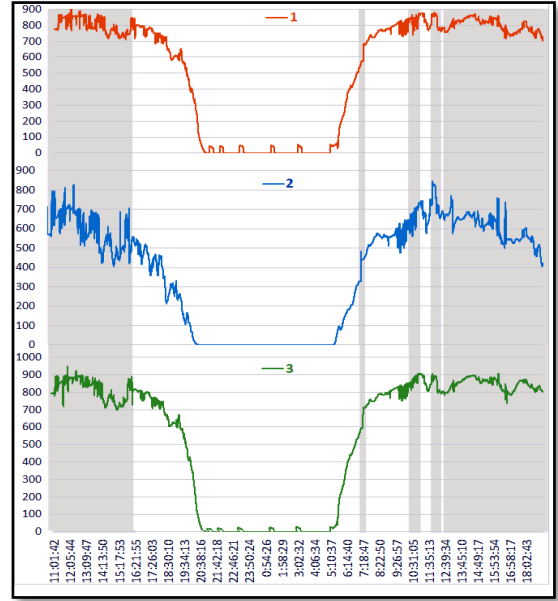
Şekil 6. Deneylerde kullanılan donanımlar

Çalışmada gerçekleştirilen deneylerde, 3 Iris düğümü gönderici olarak ve bir baz istasyonu (Iris ile MIB520 birlikte) alıcı olarak kullanılmıştır. Iris düğümleri odanın farklı köşelerine yerleştirilirken, baz istasyonu da aynı odada bir bilgisayara bağlanmıştır. Baz istasyonu dışında her bir Iris düğümüne bir MDA100 sensörü takılarak, odanın ışık seviyesi her 5 saniyede bir ölçülüp, bir radyo mesajı aracılığıyla baz istasyonuna gönderilmiştir. Ortamın radyo gürültüsünün etkisini en aza indirmek için, IRIS düğümlerin kabul ettiği radyo sinyallerinin asgari gücü -100 dBm'den -40 dBm'e yükseltilmiştir. Ayrıca radyo sinyallerinin çakışma olasılığını düşürmek için de MAC (Media Access Control) protokolü olarak CSMA (Carrier Sense Multiple Access) kullanılmıştır. Baz istasyonu, gelen mesajların RSSI değerlerini bulup, gönderen düğümün kimlik numarası ve ışık verisiyle birlikte USB girişi üzerinden, bilgisayarda çalışan Java uygulamasına aktarmıştır. TinyOS kütüphanesi ile geliştirilen Java uygulaması, gelen verileri bilgisayarın tarih ve saat bilgileri ile beraber bir log dosyasına göndermiştir. Böylece her 5 saniyede bir, ortamın ışık seviyesi ve radyo sinyal dalgalanmaları, 3 düğüm tarafından log dosyasına kaydedilmiştir.

### 3.2. Veri Seti

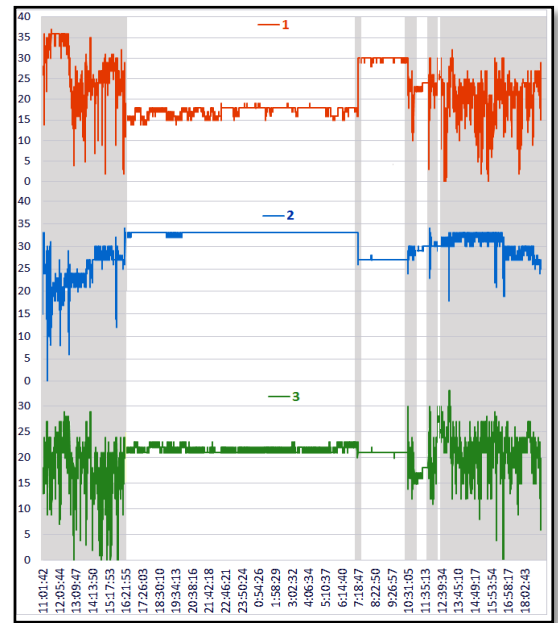
Deneyel çalışmada, bir oda içerisinde konumlandırılan 3 TDA düğümü 32 saat boyunca her 5 saniyede bir ortamın ışık yoğunluğunu ölçüp, kendi kimlik numaralarıyla beraber baz istasyonuna göndermişlerdir. Baz istasyonunun, gelen verileri, radyo paketlerinin RSSI değerleri ile birlikte bilgisayara aktarması ile veri seti oluşturulmuştur. Deneyin yapıldığı odada bulunan kişinin odaya giriş ve çıkış saatleri sisteme girilerek, odanın boş ve dolu olduğu zamanlar kaydedilmiştir.

Ortamın ışık yoğunluğu, MDA100 sensörleri tarafından 0 (tamamen karanlık) ile 1000 (maksimum ışık yoğunluğu) arasında bir tam sayı değeri olarak ölçülmüştür. 32 saat boyunca, 3 sensör tarafından toplanan ışık verileri Şekil 7'de gösterilmiştir. Bu şekildeki kırmızı, mavi ve yeşil renkler, 3 farklı sensörden gelen verileri göstermektedir. Şekil 7'deki gri renk ile gösterilen zaman aralıklarında odada en az bir kişi bulunurken, beyaz renk ile gösterilen zaman aralıklarında ise odada kimse bulunmamıştır.



Şekil 7. Deneyel çalışmada toplanan ışık yoğunluğu verileri.

Deneyel çalışmada, 3 düğüm tarafından toplanan RSSI değerleri Şekil 8'de gösterilmiştir. Şekilden anlaşıldığı üzere, odada bir kişinin bulunduğu zamanlarda (gri renk ile gösterilen bölgelerde) RSSI değerlerinin dalgalanma oranı, odanın boş olduğu zamanlara göre (beyaz renk ile gösterilen bölgelerde) daha fazladır. Ancak odada kimsenin olmadığı bazı zamanlarda da, RSSI değerlerinde geçici, ama nispeten şiddetli dalgalanmalar yaşanmıştır.



Şekil 8. Deneyel çalışmada toplanan RSSI değerleri.

Deneyel çalışma kapsamında kullanılan veri seti, 3 telsiz duyurga düğümünün 32 saat boyunca her 5 saniyede bir ortamın ışık seviyesini ve diğer düğümlerden gönderilen radyo sinyallerini ölçmesi ile 23585 kayıttan oluşmaktadır. Her bir kayıt, tarih, saat, düğüm kimlik numarası, her

düğümün algıladığı ışık seviyesi ve gönderdiği sinyal gücünü tutmaktadır. Çalışmada kullanılan sınıflandırma algoritmalarının uygulanabilmesi için, veri seti bir ön işlem aşamasından geçirilmiştir. Bu ön işlem aşamasında, bilgisayarın tarih, saat bilgileri ile düğümlerin kimlik numaraları çıkarılmıştır. Ön işlemden geçirilmiş veri setine ait öznelik, öznelik türü, özneliklerin asgari, azami, ortalama ve standart sapma değerleri Tablo 1’de sunulmuştur.

**Tablo 1.** Veri seti özellikleri

Öznelik	Öznelik Türü	Asgari Değer	Azami Değer	Ortalama Değer	Standart Sapma
Light_1	Sayısal	0	915	530.89	351.44
RSSI_1	Sayısal	0	37	21.10	5.74
Light_2	Sayısal	0	846	375.12	276.63
RSSI_2	Sayısal	0	34	30.26	3.64
Light_3	Sayısal	0	949	549.25	365.68
RSSI_3	Sayısal	0	33	20.80	3.15
Person	Kategorik	0: Kişi yok 1: Kişi var			

#### 4. DEĞERLENDİRME

Bu çalışmada, 10 farklı sınıflandırma algoritması (Naive Bayes, MLP, SVM, K-NN, C4.5, Random Forest, Random Tree, REPTree, Decision Stump ve HoeffdingTree) kapalı alanda kişi tespiti amacıyla, 3 farklı TDA düğümünden elde edilen veriler kullanılarak eğitilmiştir. K-NN algoritmasında  $k$  değeri veri seti kayıt sayısının karekökü alınarak hesaplanmıştır ve 154 olarak seçilmiştir ( $\sqrt{23585} \approx 154$ ) [43, 44]. Kullanılan bu algoritmalar doğruluk oranı, f-ölçüm değeri ve model oluşturma süreleri performanslarına göre karşılaştırılmıştır. Algoritmaların göstermiş oldukları doğruluk oranları 10-katlı çapraz geçerlilik tekniği kullanılarak elde edilmiştir.

DeneySEL çalışmalar kapsamında, geleneksel ve karar ağacı tabanlı sınıflandırma algoritmalarının sağlanmış oldukları ortalama doğruluk oranı ve model oluşturma süresi performansları da ayrıca hesaplanmıştır. Elde edilen tüm deneysel karşılaştırma sonuçları grafik ve tablolar halinde sunulmuştur.

Doğruluk oranı, algoritmaların sınıflandırma başarımını göstermede en yaygın kullanılan yöntemlerden birisidir ve algoritmanın test verisindeki doğru tahminlerinin toplam veri sayısına oranını ile elde edilir.

Bir diğer sık tercih edilen sınıflandırma performans ölçütü ise f-ölçüm değeridir. Algoritmanın f-ölçüm değeri hesaplanırken, kesinlik ve hassasiyet değerlerinin harmonik ortalaması alınır (Denklem 3).

$$F - \text{ölçüm} = \frac{2 * \text{Kesinlik} * \text{Hassasiyet}}{\text{Kesinlik} + \text{Hassasiyet}} \quad (3)$$

$DP$  ve  $DN$  doğru sınıflandırılan pozitif ve negatif değerler,  $YP$  ve  $YN$  yanlış sınıflandırılan pozitif ve negatif değerler olmak üzere, kesinlik ve hassasiyet değerleri aşağıdaki formüller ile elde edilir:

$$\text{Kesinlik} = \frac{DP}{DP + YP} \quad (4)$$

$$\text{Hassasiyet} = \frac{DP}{DP + YN} \quad (5)$$

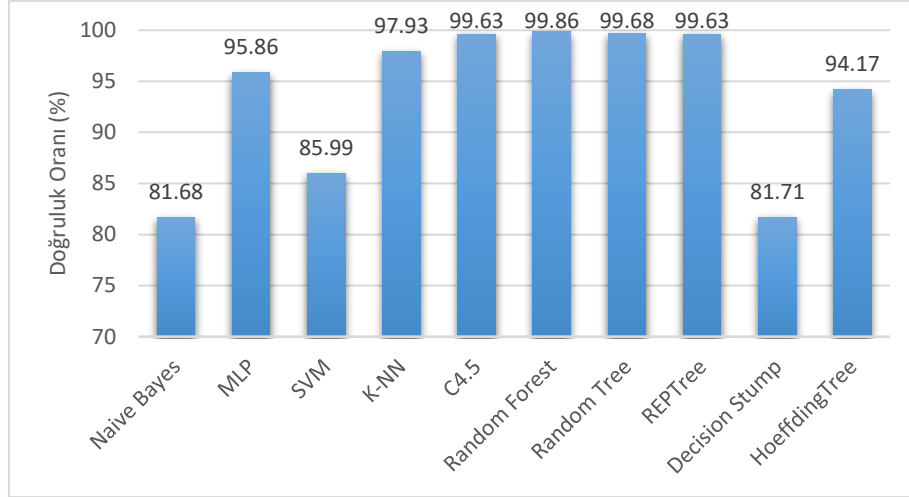
Çalışmada kullanılan sınıflandırma algoritmalarının 3 farklı TDA düğümünden elde edilen veriler üzerinde sağlanmış oldukları ağırlıklı kesinlik, hassasiyet ve f-ölçüm değerleri Tablo 2’de verilmiştir. Tablo’dan elde edilen sonuçlar göstermektedir ki, en yüksek f-ölçüm değerine sahip olan algoritma Random Forest algoritmasıdır. Random Forest algoritmasını takip eden Random Tree, C4.5 ve REPTree algoritmalarının da sırasıyla 0.997, 0.996 ve 0.996 f-ölçüm değerleri sunduğu görülmektedir. Dolayısıyla, en başarılı sınıflandırma becerisi sunan algoritmanın Random Forest olduğunu ve onu takip eden Random Tree, C4.5 ve REPTree algoritmalarının da yüksek sınıflandırma performansı gösterdiklerini söylemek mümkündür.

**Tablo 2.** Sınıflandırma algoritmalarının ağırlıklı kesinlik, hassasiyet ve f-ölçüm değerleri.

Algoritma	Kesinlik	Hassasiyet	F-ölçüm
Naive Bayes	0.866	0.817	0.817
MLP	0.959	0.959	0.959
SVM	0.862	0.860	0.860
K-NN	0.980	0.979	0.979
C4.5	0.996	0.996	0.996
Random Forest	0.999	0.999	<b>0.999</b>
Random Tree	0.997	0.997	0.997
REPTree	0.996	0.996	0.996
Decision Stump	0.866	0.817	0.818
HoeffdingTree	0.942	0.942	0.942

Benzer şekilde, sınıflandırma algoritmalarının kapalı alanda kişi tespiti amacıyla 3 farklı TDA düğümünden elde edilen veriler üzerinde sağlanmış oldukları doğruluk oranlarının yüzdesi Şekil 9’da sunulmuştur. Grafikten elde edilen sonuçlara göre, uygulanan tüm algoritmaların %80’in üzerinde sınıflandırma başarısı sunduğu görülmektedir. Bu sonuçlar, kapalı ortamda kişilerin tespit edilmesi probleminin makine öğrenmesi teknikleri ile başarılı bir şekilde çözülebildiğini göstermektedir. Ayrıca, algoritmaların elde etmiş oldukları doğruluk oranları birbirleri ile kıyaslandığında, yine en yüksek doğruluk oranına sahip olan algoritmanın %99.68 ile Random Forest olduğu anlaşılmaktadır. Bunun yanı sıra, Random Tree, C4.5 ve REPTree algoritmaları da %99’un üzerinde doğruluk oranı sunmaktadır.





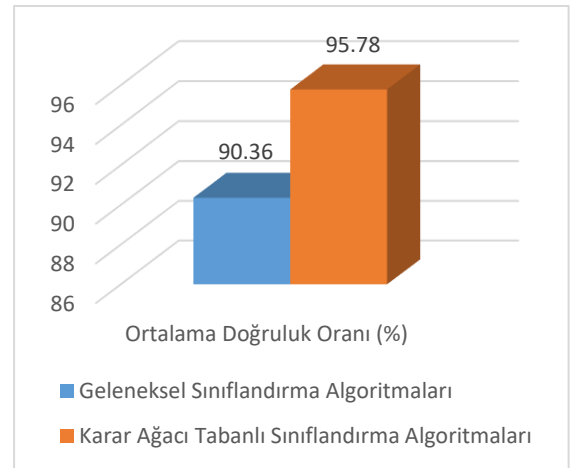
Şekil 9. Sınıflandırma algoritmalarının doğruluk oranlarının karşılaştırması.

Çalışmada tercih edilecek algoritmanın belirlenmesinde doğruluk oranının yanı sıra, modelin oluşturulması için gerek duyulan sürenin de incelenmesi gerekir. Tablo 3'te algoritmaların sınıflandırma modellerinin oluşturulma süreleri saniye cinsinden listelenmiştir. Sonuçlara göre, model oluşumu için en az süreye ihtiyaç duyan algoritmanın K-NN algoritması olduğu ve bu algoritmayı Naive Bayes, Decision Stump ve Random Tree algoritmanın takip ettiği anlaşılmaktadır.

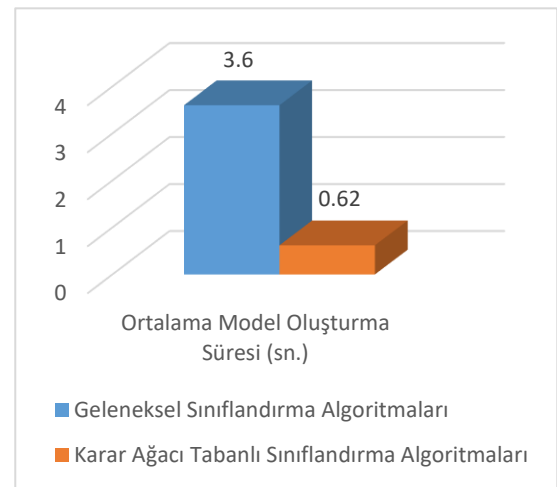
Tablo 3. Sınıflandırma algoritmalarının model oluşturma süreleri.

Algoritma	Model Oluşturma Süresi (sn.)
Naive Bayes	0.04
MLP	13.66
SVM	0.68
K-NN	<b>0.01</b>
C4.5	0.56
Random Forest	2.45
Random Tree	0.08
REPTree	0.25
Decision Stump	0.05
HoeffdingTree	0.34

Bireysel algoritma kıyaslamalarının yanı sıra, deneysel çalışmalar kapsamında geleneksel ve karar ağacı tabanlı sınıflandırma algoritmalarının sağladıkları ortalama doğruluk oranı ve model oluşturma süresi performansları da hesaplanmıştır. Elde edilen sonuçlar Şekil 10 ve Şekil 11'de sunulmuştur. Şekil 10'daki ortalama doğruluk oranı değerlerine göre, %95.78 doğruluk oranı ile karar ağacı tabanlı sınıflandırma algoritmalarının (C4.5, Random Forest, Random Tree, REPTree, Decision Stump ve HoeffdingTree), %90.36 doğruluk oranı sunan geleneksel sınıflandırma algoritmalarına (Naive Bayes, MLP, SVM ve K-NN) göre daha başarılı sınıflandırma performansı sunduğu anlaşılmaktadır. Ayrıca, sağladığı yüksek doğruluk oranına ek olarak, karar ağacı tabanlı sınıflandırma algoritmalarının ortalama 0.62 sn ile geleneksel algoritmalara göre daha az model oluşturma süresine ihtiyaç duyduğu Şekil 11'deki grafikten anlaşılmaktadır.



Şekil 10. Geleneksel ve karar ağacı tabanlı sınıflandırma algoritmalarının ortalama doğruluk oranlarının karşılaştırması.



Şekil 11. Geleneksel ve karar ağacı tabanlı sınıflandırma algoritmalarının ortalama model oluşturma sürelerinin karşılaştırması.

## 5. TARTIŞMA

Çalışma kapsamında, kapalı alanda kişi tespiti amacıyla geliştirilen sistemde, TDA düğümlerinden elde edilen ışık verileri ve radyo sinyallerinden örnek bir veri seti oluşturulmuştur. Bu veri seti üzerinde geleneksel ve karar ağacı tabanlı sınıflandırma algoritmaları uygulanarak, yüksek doğrulukla kapalı ortamda kişi tespiti gerçekleştirilmiştir. Literatürde, makine öğrenmesi tekniklerinden faydalanan kalabalık tahminleme sistemleri bulunmaktadır [21-25]. Ancak bu sistemlerde hedef varlıklar, üzerlerinde radyo sinyali gönderen bir cihaz taşımak zorundadır. Dolayısıyla böyle bir cihazı taşımayan kişiler bu sistemler tarafından tespit edilemez. Son yıllarda, cihaz taşımayan kişilerin tespitine yönelik, WiFi sinyalleri ve makine öğrenmesi tekniklerini kullanan kalabalık ve takip sistemleri önerilmiştir [26,27]. Ancak, sınırlı sayıda kaynaktan gönderilen WiFi radyo sinyallerinin kullanılması ve bazı verilerin göz ardı edilmesinden dolayı bu çalışmaların doğru tespit oranları en iyi durumda %95'e ulaşabilmiştir.

Literatürdeki benzer çalışmalardan farklı olarak, önerilen bu çalışmada, sadece radyo sinyal gücü ve istatistiksel hesaplamalar kullanmak yerine, çeşitli TDA düğümleri tarafından gönderilen ışık seviyeleri ve alınan radyo sinyallerinin gücünden oluşan bir veri seti oluşturulmuştur. Elde edilen veri seti, çalışmada uygulanan sınıflandırma algoritmalarının eğitiminde kullanılarak, yüksek tahmin yeteneğine sahip modeller elde edilmiştir. Deneysel çalışmalar göstermiştir ki, Random Tree, C4.5 ve REPTree sınıflandırma algoritmaları ile doğru tespit oranı %99.6'nın üzerine ulaşabilmektedir.

## 6. SONUÇ

Kapalı alanlarda, müşteri, personel, hasta ve çeşitli cihazların takip edilmesi ihtiyacı, yüksek hassasiyete sahip iç mekân konumlandırma ve insan tespit sistemlerinin her geçen gün daha fazla rağbet görmesine sebep olmuştur. Kapalı ortamlarda GPS sisteminin kullanılmamasından dolayı, bu teknolojiye dayalı sistemler, iç mekân konumlandırmasında kullanılamaz. Kapalı bir ortamda bulunan kişilerin hareket etmesi, o ortamda yayılmakta olan sinyallerin gücünü değiştirir. Birçok çalışmada, herhangi bir radyo göndericisi taşımayan hedef varlıkların takibi için, sabit noktalardan yayılan radyo sinyallerinin gücünde meydana gelen dalgalanmalardan yararlanılmaktadır. Bu tür çalışmaların en temel problemi, ortamdaki hareketlilikten kaynaklanmayan sinyal dalgalanmalarıdır. Geleneksel istatistiksel, filtreleme ve eşik değer kontrolü gibi yöntemler, hareketlilik dışı durumlardan meydana gelen sinyal dalgalanmaları sebebiyle hatalı konum tespitleri gerçekleştirebilmektedirler. Bu problemin önüne geçebilmek amacıyla, çalışma kapsamında, kapalı ortamda kişi tespiti için 10 farklı geleneksel (Naive Bayes, MLP, SVM ve K-NN) ve karar ağacı tabanlı (C4.5, Random Forest, Random Tree, REPTree, Decision Stump ve HoeffdingTree) makine öğrenmesi algoritmaları

kullanılmıştır. Bu algoritmalar, deneysel çalışma kapsamında, kapalı ortamdaki 3 farklı TDA düğümünden elde edilen veriler üzerinde ayrı ayrı uygulanmış, doğruluk oranı ve model oluşturma süresi performanslarına göre karşılaştırılmıştır. Algoritmaların doğruluk oranı performansları göz önüne alındığında, tüm algoritmaların kapalı alandaki kişi tespitinde %80'in üzerinde başarı gösterdiği ve en yüksek doğruluk oranı sunan algoritmanın ise %99.68 ile Random Forest olduğu görülmüştür. Ayrıca, geleneksel ve karar ağacı tabanlı sınıflandırma algoritmaları kendi aralarında sağlamış oldukları ortalama doğruluk oranlarına göre kıyaslanmış ve karar ağacı tabanlı algoritmalar %95.78 ile daha yüksek tahminleme becerisi sunmuştur.

Gelecek çalışma olarak, topluluk tabanlı makine öğrenmesi yöntemleri kullanarak ortamda bulunan kişi sayısı ve kişilerin konumunun tahmin edilmesi amaçlanmaktadır. Ayrıca, bu çalışmada kullanılan RSSI sinyal verileri yerine Bluetooth sinyalleri kullanılarak, kapalı alandaki kişi tespitinin makine öğrenmesi algoritmaları ile yüksek doğrulukla tahminlenmesi hedeflenmektedir.

## KAYNAKÇA

- [1].F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and Technologies," *IEEE Commun.*, vol. 21, no 3, pp. 2568-2599, 2019.
- [2].O. Dağdeviren and V.K. Akram, "TinyOS Tabanlı Telsiz Duyarga Ağları için Bir Konumlandırma ve k-Bağlılık Denetleme Sistemi," *Bilişim Teknolojileri Dergisi*, vol. 10, no 2, pp. 139-152, 2017.
- [3].P. Kriz, F. Maly, and T. Kozel, "Improving Indoor Localization Using Bluetooth Low Energy Beacons," *Mob. Inf. Syst.*, vol. 2016, pp. 1-11, 2016.
- [4].N. Karimpour, B. Karaduman, A. Ural, M. Challenger, and O. Dagdeviren, "IoT based Hand Hygiene Compliance Monitoring," in 2019 International Symposium on Networks, Computers and Communications (ISNCC), Istanbul, Turkey, 2019, pp. 1-6.
- [5].S. Chakraborty, S. K. Ghosh, A. Jamthe, and D. P. Agrawal, "Detecting Mobility for Monitoring Patients with Parkinson's Disease at Home using RSSI in a Wireless Sensor Network," *Procedia Comput. Sci.*, vol. 19, pp. 956-961, 2013.
- [6].O. Kaltiokallio and M. Bocca, "Real-Time Intrusion Detection and Tracking in Indoor Environment through Distributed RSSI Processing," in 2011 IEEE 17th International Conference on Embedded and Real-Time Computing Systems and Applications, Toyama, 2011, pp. 61-70.
- [7].T. Teixeira and A. Savvides, "Lightweight People Counting and Localizing for Easily Deployable Indoors WSNs," *IEEE J Sel. Top. Signal Process.*, vol. 2, no. 4, pp. 493-502, 2008.
- [8].M. Nakatsuka, H. Iwatani, and J. Katto, "A study on passive crowd density estimation using wireless sensors," in 4th International Conference on Mobile Computing and

- Ubiquitous Networking, Miraikan, Tokyo Japan, 2008, pp. 1-6.
- [9].S. Shukri, L. M. Kamarudin, G.C. Cheik, R. Gunasagaran, A. Zakaria, K. Kamarudin, and S.N. Azemi, "Analysis of RSSI-based DFL for human detection in indoor environment using IRIS mote," in 3rd International Conference on Electronic Design (ICED), Phuket, Thailand, 2016, pp. 216-221.
- [10].A. Booranawong, N. Jindapetch, and H. Saito, "A System for Detection and Tracking of Human Movements Using RSSI Signals," *IEEE Sens. J.*, vol. 18, no. 6, pp. 2531–2544, 2018.
- [11].A. Booranawong, N. Jindapetch, and H. Saito, "Adaptive Filtering Methods for RSSI Signals in a Device-Free Human Detection and Tracking System," *IEEE Syst. J.*, vol. 13, no. 3, pp. 2998–3009, 2019.
- [12].B. Mrazovac, M. Bjelica, D. Kukolj, B. Todorovic, and D. Samardzija, "A human detection method for residential smart energy systems based on Zigbee RSSI changes," *IEEE Trans. Consum. Electron.*, vol. 58, no. 3, pp. 819–824, 2012.
- [13].Y. Jin, Z. Tian, M. Zhou, , Z. Li, and Z. Zhang, "A whole-home level intrusion detection system using WiFi-enabled IoT," in 14th International Wireless Communications & Mobile Computing Conference (IWCMC), Limassol, Cyprus, 2018, pp. 494-499.
- [14].O. Kaltiokallio, M.Bocca, and L. M. Eriksson, "Distributed RSSI processing for intrusion detection in indoor environments," in Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, Stockholm, Sweden, 2010, pp. 404-405.
- [15].T. Wang, D. Yang, S. Zhang, Y. Wu, and S. Xu, "Wi-Alarm: Low-Cost Passive Intrusion Detection Using WiFi," *Sensors*, vol. 19, no. 10, pp. 2335, 2019.
- [16]. J.W. Davis, V. Sharma, A. Tyagi, and M. Keck, *Human Detection and Tracking*, Boston,MA: Springer, 2009.
- [17].H. Chen, Q. Zhang, Y. Liu, Y. Yang, and Z. Guo, "A Tentative Study on Zigbee-Based Indoor Human Intrusion Detection," *Advances in Intelligent Systems and Computing. Recent Developments in Intelligent Computing, Communication and Devices*, pp. 501–506, 2018.
- [18].B. Chatfield and R. J. Haddad, "RSSI-based spoofing detection in smart grid IEEE 802.11 home area networks," 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington D.C., USA, 2017, pp. 1-5.
- [19].Y. Bao, L. Dong, Y. Zheng, and Y. Liu, "WiSafe: a real-time system for intrusion detection based on wifi signals," in Proceedings of the ACM Turing Celebration Conference-China, Chengdu, China, 2019, pp. 1-5.
- [20].D. Li, L. Deng, M. Lee, and H. Wang, "IoT data feature extraction and intrusion detection system for smart cities based on deep migration learning," *Int. J. Inf. Manag.*, vol. 49, pp. 533–545, 2019.
- [21].T.Yoshida and Y.Taniguchi, "Estimating the number of people using existing wifi access point in indoor environment," in Proceedings of the 6th European Conference of Computer Science (ECCS'15), Rome, Italy, 2015, pp. 46-53.
- [22].Y. Yuan, C. Qiu, W. Xi, and J. Zhao, "Crowd density estimation using wireless sensor networks," in Seventh international conference on mobile Ad-hoc and sensor networks, Beijing, China, 2011, pp. 138-145.
- [23].N. Matsumoto, J. Kawasaki, M. Suzuki, S. Saruwatari, and T. Watanabe, "Crowdedness Estimation Using RSSI on Already-deployed Wireless Sensor Networks," in 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring), Kuala Lumpur, Malaysia, 2019, pp. 1-7.
- [24].D. Ghosh, P. Roy, C. Chowdhury, and S. Bandyopadhyay. "An ensemble of condition based classifiers for indoor localization," in IEEE International conference on advanced networks and telecommunications systems (ANTS), Bangalore, India, 2016, pp. 1-6.
- [25].D. Alshamaa, A. Chkeir, and F. Mourad-Chehade, "Target localization using machine learning and belief functions: Application for elderly people in indoor environments," in 2019 IEEE Symposium on Computers and Communications (ISCC), Barcelona, Spain, 2019, pp. 1-6.
- [26].H. Huang and S. Lin, "WiDet: Wi-Fi based device-free passive person detection with deep convolutional neural networks," *Comput. Commun.*, vol. 150, pp. 357-366, (2020).
- [27].M. A. Al-qaness, "Device-free human micro-activity recognition method using WiFi signals", *Geo. Spat. Inf. Sci.*, vol. 22, no. 2, pp.128-137, 2019.
- [28].E. Alpaydin, *Introduction to machine learning*. Cambridge, MA: The MIT Press, 2020.
- [29].P. Yildirim and D. Birant, "Naive Bayes classifier for continuous variables using novel method (NBC4D) and distributions," in 2014 IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA) Proceedings, Alberobello, Italy, 2014, pp. 110-115.
- [30].P. Yıldırım, D. Birant and T. Alpyıldız, "Improving prediction performance using ensemble neural networks in textile sector," in 2017 International Conference on Computer Science and Engineering (UBMK), Antalya, Turkey, 2017, pp. 639-644.
- [31].Z. Zhang, "Introduction to machine learning: k-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, 2016.
- [32].S. Huang, N. Cai, P. P. Pacheco, S. Narandes, Y. Wang, and W. Xu, "Applications of Support Vector Machine (SVM) Learning in Cancer Genomics," *Cancer Genom. Proteom.*, vol. 15, no. 1, pp. 41-51, 2017.
- [33].Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Arch. Psychiatry*, vol. 27, no. 2, pp. 130-135, 2015.
- [34].J. R. Quinlan, *C4.5 - programs for machine learning*. San Mateo, CA: Kaufmann, 1992.
- [35]. X. Gao, J. Wen, and C. Zhang, "An Improved Random Forest Algorithm for Predicting Employee Turnover," *Math. Probl. Eng.*, vol. 2019, pp. 1–12, 2019.
- [36].A. Onan, "Şirket İflaslarının Tahminlenmesinde Karar Ağacı Algoritmalarının Karşılaştırmalı Başarım Analizi," *Bilişim Teknolojileri Dergisi*, vol. 8, no. 1, 2015.
- [37]. K. H. Raviya and B. Gajjar, "Performance Evaluation of Different Data Mining Classification Algorithm Using WEKA," *Paripex Indian J.*, vol. 2, no. 1, pp. 19–21, 2012.

- [38].S. Chen, B. Shen, X. Wang, and S.-J. Yoo, "A Strong Machine Learning Classifier and Decision Stumps Based Hybrid AdaBoost Classification Algorithm for Cognitive Radios," *Sensors*, vol. 19, no. 23, p. 5077, 2019.
- [39].A. Kumar, P. Kaur, and P. Sharma, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *Int. J. Sci. Res.*, vol. 5, no. 4, pp. 2094–2097, 2016.
- [40].Weka 3 - Data Mining with Open Source Machine Learning Software in Java, URL: <https://www.cs.waikato.ac.nz/ml/weka/> (Erişim zamanı; Temmuz, 4, 2020).
- [41].Wireless Measurement System - Memsic Inc., URL: [http://www.memsic.com/userfiles/files/Datasheets/WSN/IRIS\\_Datasheet.pdf](http://www.memsic.com/userfiles/files/Datasheets/WSN/IRIS_Datasheet.pdf) (Erişim zamanı; Temmuz, 4, 2020).
- [42].P. Levis, S. Madden, J. Polastre, R. Szewczyk, K. Whitehouse, A. Woo, and Culler, D. "TinyOS: An operating system for sensor networks," In: Weber W., Rabaey J.M., Aarts E. (eds) *Ambient Intelligence*. Springer, Berlin, Heidelberg.
- [43]. P. Nadkarni, (2016) *Core Technologies: Data Mining and "Big Data"*. In: *Clinical Research Computing*. Academic Press, Iowa City, IA, United States.
- [44]. B. Lantz, *Machine learning with R*, Birmingham: Packt Publishing, 2015.