

## Zaman Serisi Sınıflandırması ile Ayakkabı Üzerinden Aktivite Tanımlama

### ÖZET

Veri madenciliği son yıllarda bilişim alanındaki gelişmeleri günlük hayatımızın içerisine fazlasıyla girerek çeşitli alanlarda büyük faydalar sağlamaktadır. Bu çalışma içerisinde de ayakkabıya yerleştirilecek bir cihazdan gelen veriler zaman serisi sınıflandırma yöntemleri kullanılarak kullanıcıların yürüme, koşma, bisiklete binme, merdivenden inme ve merdiven çıkma aktivitelerini tanımlanması amaçlanmıştır. Çalışma için gerekli olan verileri toplayabilmek için küçük ve maliyeti düşük bir cihaz oluşturulmuştur. Cihaz ayakkabının üzerine yerleştirilerek hareketleri gerçekleştirirken Bluetooth ile 3 eksen ivmeölçer ve 3 eksen jiroskop olmak üzere 6 eksenli veri, zaman etiketli bir şekilde bilgisayara aktarılmıştır. Aktarılan veriler python yazılım dilinde oluşturulan kodlar ile işlenmiş ve csv formatında kaydedilmiştir. Elde edilen veriler model eğitime uygun olmadığından ve bazı hatalar içerdiğinden dolayı veri madenciliği ön işlemlerinden geçirilmiş ve tek dosyada aktivite etiketleriyle birlikte birleştirilmiştir. Birleştirilen veriler ile farklı sınıflandırma algoritmalarında modeller eğitilmiş ve sınıflandırma performansları incelenmiştir. Yapılan incelemeler ve kıyaslamalar sonucu genel olarak algoritmaların sınıflandırma başarısı %93'ün üzerinde olduğu görülmüş ve çalışmanın sonucunda ayakkabı üzerinden aktivite tanımlaması yapılabileceği ortaya konulmuştur. Giyilebilir cihazlar üzerine geliştirilen sağlık ve spor alanında uygulamalar çok fazla talep edildiği görülmektedir. Bu çalışma kapsamında ele alınan akıllı ayakkabı ile de bu alanda güzel bir alternatif oluşturularak sağlık ve spor alanındaki sorunlara çözüm getirebildiği öngörülmüştür.

Anahtar Kelimeler: Veri Madenciliği, Zaman Serisi Sınıflandırma, Akıllı Ayakkabı, İnsan Aktivitesi Tanıma, Giyilebilir Cihaz

## GİRİŞ

Teknolojinin ve bilimin gelişmesiyle birçok alanda yeni ürünler ve yöntemler ortaya çıkmaya devam etmektedir. Özellikle son zamanlarda nesnelerin interneti ve veri madenciliği alanları birçok girişimcinin ve bilim insanının ilgisini çekmektedir. Nesnelerin internetinin kullanıcılar tarafından en çok talep gören uygulamalarından biri de giyilebilir cihazlardır. Özellikle de giyilebilir cihazların sağlık alanında kalp atış hızı, adım sayısı, uyku ve aktivite takibi gibi uygulamaları popülerliğini arttırmaktadır. Veri madenciliğinde ön zorlu problemlerden biri olan zaman serisi sınıflandırması son zamanlarda gelişme göstermiş ve son yıllarda yeni zaman serisi yöntemleri önerilmiştir. Zaman serisi sınıflandırma yöntemleri sayesinde de çeşitli sensörlerden gelen veriler ile yüksek doğrulukla aktivite tanımlamak mümkün hale gelmiştir.

Takip uygulamaları çeşitli giyilebilir ya da taşınabilir cihazlar ile yapılsa da genellikle telefon, akıllı saat veya akıllı bileklik kullanılarak yapılmaktadır. Bu cihazlar ile birçok takip uygulamasında yüksek başarı elde edilebilmektedir. Ancak spesifik olarak aktivite takibi düşünüldüğünde yaptığımız birçok aktivitede (koşma, yürüme, spor vs.) ayaklarımızı kullanırız. Bu yüzden ayakkabıya yerleştirilecek bir cihaz üzerinden aktiviteleri takip etmek daha doğru aktivite takibi yapmamızı sağlayabilir ve birçok aktiviteyi tanıyarak aktiviteye göre kalori hesabı, atılan adım sayısı vb. hesaplar daha doğru bir şekilde yapılabilir. Bu çalışma içerisinde de ayakkabıya yerleştirilecek bir cihazdan gelen veriler zaman serisi sınıflandırma yöntemleri kullanılarak kullanıcıların aktivitelerini tanımlaması amaçlanmıştır.

Literatür incelendiğinde aktivite tanıma üzerine yapılan çalışmalarda vücudun çeşitli bölgelerine yerleştirilen farklı cihazlar kullanılsa da kolay ulaşılabilir olması nedeniyle genellikle akıllı telefonların algılayıcılarından yararlanıldığı görülmektedir. Bao ve Intille (2004) tarafından hazırlanan makalede, kullanıcının sağ kalçasına, el bileğine, üst koluna, ayak bileğine ve uyluğuna takılan beş çift eksenli ivmeölçeri kullanarak 20 kullanıcıdan veri topladılar. Karar tabloları ve Naive Bayes sınıflandırıcılarını kullanarak 20 farklı etkinliği tanıyabildiler. Çalışmalarından elde edilen sonuçlar, bir bireyin kalçasına ve baskın bileğine yerleştirilen ivmeölçerin ADL (günlük yaşam aktiviteleri)' yi tanımda en fazla güce sahip olduğunu göstermektedir. Sharma, vd. (2008) belirli bir tür üç eksenli ivmeölçer kullandı ve sinir ağları aracılığıyla %84 doğruluk elde ettiler. Yang vd. (2008), ivmeölçer ile yürüme, koşma, ayakta durma, bilgisayardan çalışma, ovma, oturma, temizlik yapma ve dış fırçalama aktiviteleri tespiti gerçekleştirmiştir. Çalışma aktivitelerin sınıflandırılmasında %90'ın üzerinde başarı oranı sağlanmıştır. Kwapisz, vd. (2010) altı temel etkinliği tanımak için kullanıcının pantolonunun cebine yerleştirilen bir akıllı telefonda tek bir üç eksenli ivme ölçer kullanılmıştır. Elde ettikleri sonuçlar, J48 karar ağaçlarının sabit aktiviteleri kolaylıkla çıkarabildiğini, çok katmanlı algılayıcıların ise hareket edilirken ki aktiviteleri yüksek doğrulukta tanıdığını görülmüştür. Tek bir sınıflandırıcının tüm faaliyetler için iyi sonuçlar vermediğini ve yukarı ve aşağı hareketlerin oldukça kafa karıştırıcı olduğunu görülmektedir. Lee ve Cho (2011) ayrıca Hierarchical Probabilistic Modelleri (HMM'ler) ile eylemleri ve etkinlikleri çıkarmak için gönüllü 4 denek tarafından elle tutulan bir akıllı telefon ivmeölçerini kullandılar. Akıllı telefon, akıllı saatler vb. sensör içeren akıllı cihazların herkes tarafından erişilebiliyor olması bu cihazların aktivite tanımda kullanımını artırmıştır. Kwapisz vd. (2011), 29 kişiden akıllı telefonlar kullanarak yürüme, hafif tempoda koşma, merdiven çıkma ve inme, oturma ve ayakta durma eylemlerini %91,7 oranında doğru sınıflandırmıştır. Khan, (2013) ile dört farklı gündelik eylemi (uzanma, oturma, koşma ve yürüme) ayırt etmek için Wii Remote'u ve gömülü 3D ivmeölçeri (belin arkasına yerleştirilmiş) kullandı ve J48 sınıflandırıcılarını kullanarak yüksek doğrulukta sonuçlar elde etti. Riboni ve Bettini (2011), çalışmasında, dış fırçalama, yürüme, bisiklet sürme, düşük tempoda koşma, ayakta durma, gezinme, merdivenden inme, merdivenden çıkma ve tahtaya yazı yazma eylemlerini tespit etmek için Android tabanlı akıllı telefon ve giyilebilir cihazları kullanılmıştır. Wu, vd. (2012) bir İpod Touch kullanarak 16 katılımcı ile 13 aktivite (oturma, yürüme, koşma ve farklı hızlarda yukarı ve aşağı yürüme) jiroskop ve ivmeölçer verisi toplayarak farklı sınıflandırıcılar üzerinde çalışma gerçekleştirdiler. Sonuçta jiroskopun eklenmesinin sadece ivmeölçer okumalarına güvenmekten daha faydalı olduğu sonucuna vardılar. De la Vega vd. (2013), tarafından yapılan çalışmada akıllı telefonda bulunan dahili ivmeölçer, jiroskop ve manyetometre sensörlerini kullanarak aktivite tanıma yapmışlardır. Sağbaş ve Ballı (2016) Android işletim sistemine sahip akıllı telefonun dahili ivmeölçer, jiroskop ve GPS algılayıcıları kullanarak ulaşım türü tespiti gerçekleştirilmiştir. Yürüme, koşma, bisiklet sürme ve araba veya otobüs ile seyahat etme eylemlerinin farklı makine öğrenmesi yöntemleriyle ile değerlendirmiş ve en iyi sonucu Random Forest yönteminden elde etmişlerdir. Ayrıca çalışmada sadece ivmeölçer yerine jiroskop ve gps algılayıcılarının kullanılması veriyi zenginleştirerek doğruluk oranını artırdığı gözlemlenmiştir.

Bu çalışmada insan hareketlerinin tespitini gerçekleştirmek için arduino platformu ile oluşturulan cihazdan elde edilen gyroscope ve ivmeölçer verileri kullanılmıştır. Cihaz ayakkabı üzerine yerleştirileceği için özellikle dış mekanda yapılan ve bireylerin ayak hareketleri ön plana çıktığı yürüme, koşma, merdiven inme ve merdiven çıkma aktiviteleri tercih edilmiştir. Toplanan veriler çeşitli makine öğrenim algoritmaları kullanılarak test edilmiş ve performansları karşılaştırılmıştır.

## YÖNTEM

Bu bölüm içerisinde de kullanılan algoritmalar ve kullanılan diğer yöntemlere yer verilmiştir. Öncelikle verileri toplamak için bir cihaz oluşturulmuştur. Ardından toplanan veriler çeşitli veri ön işleme aşamalarından geçirilerek zaman serisi sınıflandırılmasında da sıklıkla kullanılan Naive Bayes, K-Nearest Neighbor, Support Vector Machine, Decision Tree ve Random Forest makine öğrenim algoritmaları aktivite sınıflandırılması için kullanılmıştır.

### Naive Bayes Classifier (NBC)

NBC Sınıflandırıcısı güçlü nitelik bağımsızlığı varsayımına sahip, olasılığa dayalı basit bir algoritmadır. NBC Sınıflandırıcısı test verisi üzerinden öğrenir ve en yüksek orandaki örneğini sınıfa dâhil eder.  $C$  bir örneğin sınıfını belirten rastgele değişken olsun.  $X < X_1, X_2, \dots, X_m >$  değerleri ise gözlemlenen nitelik değerlerinin bir vektörü olsun.  $C$  bilinen bir sınıf etiketini,  $x < x_1, x_2, \dots, x_m >$  ise bilinen ve gözlemlenen niteliklerin değerlerini temsil etsin.  $x$  test verisine göre sınıfı tahmin etmek için Bayes Teoremi olasılık hesaplar:

$$p(C = c|X = x) = \frac{p(C = c)p(X = x|C = c)}{p(X = x)} \quad (1)$$

Sonrasında en yüksek olasılık ile sınıfı tahmin eder. Bu örnekte  $X = x$  durumu  $X_1 = x_1 \cap X_2 = x_2 \cap \dots \cap X_m = x_m$  ifade eder.  $p(X = x)$  sınıflar arasında değişme göstermediği durumda ihmal edilir ve (1) numaralı denklem aşağıdaki şekle gelir.

$$p(C = c|X = x) \propto p(C = c)p(X = x|C = c) \quad (2)$$

$p(C = c)$  ve  $p(X = x | C = c)$  öğrenme verilerinden tahmin edilir.  $x_1, x_2, \dots, x_m$  nitelikleri birbirlerine koşullu olarak bağımsızdırlar. Bu durumda (2) numaralı denklem aşağıdaki şekli alır:

$$p(C = c|X = x) \propto p(C = c) \prod_{i=1}^m p(X_i = x_i|C = c) \quad (3)$$

NBC denilen (3) numaralı denklemi kullanarak test örneklerini hesaplamak ve öğrenme verilerinden tahminde bulunmak daha kolaydır (Chandra ve diğerleri, 2007).

### K-Nearest Neighbor (KNN)

K-NN (En Yakın K Komşu) algoritması, 1967 yılında T. M. Cover ve P. E. Hart tarafından önerilen basit ve çok kullanılan bir makine öğrenme algoritmasıdır. (Cover vd., 1967) Sınıfları belli olan bir veri kümesindeki verilerden yararlanmaktadır. Örnek veri setine katılacak olan yeni verinin, mevcut verilere göre uzaklığı hesaplanıp, k sayıda yakın komşuluğuna bakılır. (Mitchell, 1997) Uygulaması basit ve gürlütlü eğitim verilerine karşı dayanıklı olmasından dolayı sınıflandırma uygulamalarında çok tercih edilmektedir. (Bhatia ve Vandana, 2010) Her zaman, K'nin değerini belirlenmesi ve tüm eğitim örnekleri için veri noktaları arasındaki mesafenin hesaplanması gerekir. Bu durum veri seti büyüdükçe kullanılan belleğin artmasına, harcanan sürenin uzamasına ve bunların sonucunda hesaplama maliyetini yükselmesine sebep olur. (Liu vd., 2012) Veriler arası mesafenin belirlenmesinde denklem-4'deki formül kullanılmaktadır.

$$d_{(i,j)} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (4)$$

KNN algoritmasında sınıflandırma yapılması için yeni bir veri eklendiğinde K değerini bakılır. Eşitlik olmaması için K tek sayı olarak seçilir. Yeni eklenen veri ile mevcut veriler arasındaki uzaklığı hesaplamak için Öklid, Manhattan veya Kosinüs yöntemleri kullanılır. (Mitchell, 1997)

## Support Vector Machine (SVM)

SVM (Destek Vektör Makineleri) ilk olarak Vapnik (1995) tarafından tanıtılmıştır. SVM regresyon, sınıflandırma ve genel örüntü tanıma için çok etkili yöntemlerden biri olmuştur. (Ristianini, 2000) Girdi uzayının boyutu çok büyük olsa bile önsel bir bilgi eklemeye gerek kalmadan yüksek genelleme performansına sahip olması nedeniyle iyi bir sınıflandırıcı olarak kabul edilir. SVM' de amaç eğitim verilerindeki iki sınıf ın üyeleri arasında ayırım yapabilme için en iyi sınıflandırma fonksiyonunu bulmaktır. Doğrusal olarak ayrılabilir bir veri kümesi için; doğrusal bir sınıflandırma fonksiyonu  $f(x)$ , iki sınıfı ayıran ve ortasından geçen bir ayırıcı hiper düzleme karşılık gelir. Bu fonksiyon belirlendikten sonra yeni veri örneği  $x_n$  basitçe  $f(x)$  fonksiyonu üzerinde test edilerek sınıflandırılabilir;  $f(x) > 0$  ise  $x_n$  pozitif sınıfa aittir.

Veri setinde bu tür birçok doğrusal hiper düzlem olduğunda, SVM, iki sınıf arasındaki marjı maksimize ederek en iyi  $f(x)$  fonksiyonunu bulur. Sezgisel olarak, marj, hiper düzlem tarafından tanımlanan iki sınıf arasındaki boşluk miktarı olarak tanımlanır. Geometrik olarak, marj, en yakın veri noktaları ile hiper düzlemdeki bir nokta arasındaki en kısa mesafeye karşılık gelir. Bu geometrik tanıma sahip olmak, marjı nasıl maksimize edeceğimizi keşfetmemizi sağlar. Böylece sonsuz sayıda hiper düzlem olsa bile, sadece birkaçı SVM' nin çözümü olarak nitelendirilebilir. Maksimum marj hiper düzlemlerinin gerçekten bulunduğundan emin olmak için, bir SVM sınıflandırıcısı aşağıdaki işlevi  $w$  ve  $b$ 'ye göre maksimize etmeye çalışır:

$$L_p = \frac{1}{2} ||w|| - \sum_{i=1}^t \alpha_i y_i (w \cdot x_i + b) + \sum_{i=1}^t \alpha_i \quad (5)$$

Burada  $t$ , eğitim örneklerinin sayısı ve  $\alpha_i, i = 1, \dots, t$  " $L_p$ 'ye göre türevleri sıfır olacak şekilde negatif olmayan sayılardır.  $L_p$ , Lagrange çarpanlarıdır ve  $L_p$ , Lagrangian olarak adlandırılır. Bu denklemde vektörler ve  $b$  sabiti hiper düzlemi tanımlar.

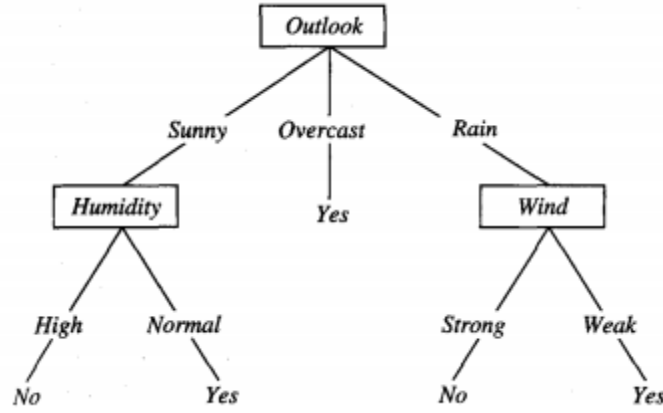
SVM gibi bir makine öğrenme algoritması, bazı temel parametrelere dayalı bir fonksiyon sınıfı olarak modellenebilir. Farklı işlev sınıfları, VC boyutu olarak bilinen bir  $h$  parametresi ile temsil edilen farklı öğrenme kapasitesine sahip olabilir. (Vapnik, 1995) VC boyutu, bu noktalardaki herhangi bir sınıf etiketi ataması için eğitim verilerinde sıfır hata oranları elde ederek işlev sınıfının düzgün bir şekilde öğrenmek için kullanılabileceği maksimum eğitim örneği sayısını ölçer. Gelecekteki verilerdeki gerçek hatanın iki terim toplamı ile sınırlandırıldığı kanıtlanabilir. İlk terim eğitim hatasıdır ve ikinci terim, VC boyutunun  $h$  karekökü ile orantılı ise  $h$ 'yi en aza indirebilirsek, gelecekteki hatayı en aza indirebiliriz, eğitim hatasını da en aza indirdiğimiz sürece, SVM sayısal hesaplamalar yapmak için kolayca genişletilebilir.

SVM iki şekilde genişletilebilir. Diğeri ise SVM'yi, hedefin hedef fonksiyona yaklaşabilen doğrusal bir fonksiyon üretmek olduğu regresyon analizi yapmak üzere genişletmek. Hata modellerinin seçimine dikkatle bakılır; vektör regresyonunu veya SVR'yi desteklemede, gerçek ve tahmin edilen değerler arasındaki farklar epsilon miktarı içinde olduğunda hata sıfır olarak tanımlanır. Aksi takdirde, epsilon hatası doğrusal olarak büyür. Destek vektörleri daha sonra Lagrangian'ın en aza indirilmesi yoluyla öğrenilebilir. Destek vektörü regresyonunun bir avantajı, aykırı değerlere karşı duyarsızlığıdır. SVM' nin ilk dezavantajlarından biri ise hesaplama verimsizliğidir. Ancak bu soruna farklı yaklaşımlar ile çözüm getirilmiştir. Yaklaşımlardan biri, büyük bir optimizasyon problemini bir dizi küçük probleme bölmektir; burada her problem, optimizasyonun verimli bir şekilde yapılabilmesi için dikkatle seçilmiş birkaç değişkeni içerir. İşlem, tüm ayrıştırılmış optimizasyon sorunları başarıyla çözümlene kadar yinelenir. (Kumar, 2012)

## Decision Tree

Decision Tree (Karar ağaçları) veri madenciliğinde sınıflandırma için kullanılan popüler makine öğrenme algoritmalarından biridir. Karar ağaçları istatistiksel varsayımlar yapılamayan ve öngörülerin birkaç mantıksal if-then şartıyla sunulan parametrik olmayan bir yöntemdir. (Nisbet vd., 2009) Bir karar ağacı algoritması gürültülü ve bilinmeyen verilerin karşı dayanıklıdır. En önemli tercih sebeplerinden biri ise oluşturulan modellerin yorumlanması kolay olmasıdır. Karar ağaçları düğümler dallar ve yapraklardan oluşur. (Dong-Peng vd., 2008)

Karar ağaçları, verileri kökten bazı yaprak düğümlerine doğru sıralayarak sınıflandırır. Ağaçtaki her düğüm, verinin bazı öznelitliklerinin bir testini içerir ve bu düğümden inen her dal, bu öznelitlik için olası değerlerden birine karşılık gelir. Bir veri, ağacın kök düğüminden başlayarak, bu düğüm tarafından belirtilen öznelitliği test ederek, ardından verilen örnekteki öznelitliğin değerine karşılık gelen ağaç dalından aşağı inerek sınıflandırılır. Bu işlem daha sonra yeni düğümden köklenen alt ağaç için tekrarlanır. (Mitchell, 1997)



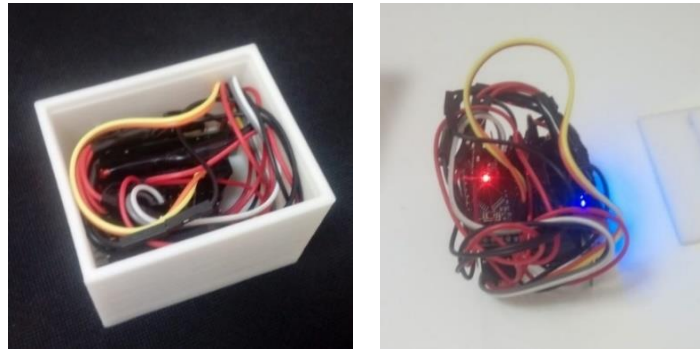
Şekil 1 tipik bir öğrenilmiş karar ağacını göstermektedir. Bu karar ağacı, Cumartesi sabahlarını tenis oynamaya uygun olup olmadıklarına göre sınıflandırır.

### Random Forest (RF)

Random Forest (Rastgele Orman) Leo Breiman ve Adele Cutler (2001) tarafından geliştirilen ve içerisinde oylama metodunu barındıran bir sınıflama yöntemidir. Birçok karar ağacının biraya gelmesiyle oluşur ve bireysel ağaçlar tarafından oylanarak kazanan sınıf belirlenir. Rastgele orman içerisinde birçok karar ağacı içermesi sebebiyle doğruluk değerini çok fazla arttıran bir yöntemdir.

### Veri Toplama

Bu çalışma için gerekli olan verileri toplamamızı sağlayacak bir cihaz oluşturulmuştur. Cihaz üzerinde Arduino ara yüzünde yazılan kodlar ile sensörler ve modüller tanıtılarak uygun formatta veri gönderilmesi sağlanmıştır. Cihaz sayesinde Bluetooth bağlantısı aracılığıyla bilgisayara gyroscope ve ivmeölçer verileri aktarılmıştır. Aktarılan bu veriler python yazılım dilinde oluşturulan kodlar ile işlenmiş ve csv formatında kaydedilmiştir.



Şekil 2. Cihaz

Ayakkabı üzerine yerleştirilen cihaz çalışmaya başladıktan sonra sürekli olarak bilgisayara saniyede 10 veri (10 Hz) göndermektedir. Bilgisayarda oluşturulan yazılım gelen verileri takip ederek kullanıcı harekete başladığını ve durduğunu algılaması sağlanıp sadece hareket anındaki veriler toplanmıştır. Ayrıca veriler kaydedilmeden önce kalman filtresinden geçirilerek daha doğru veriler elde edilmesi sağlanmıştır.

Bu çalışmada 4 farklı hareket verisi toplanmıştır. Bunlar;

- Yürüme (Walk)
- Koşma (Run)
- Merdiven inme (Downstairs)
- Merdiven çıkma (Upstairs)

Hareket verileri toplanmadan önce cihaz her zaman sağ ayağa yerleştirilmiş ve her zaman harekete sağ ayak ile başlanmıştır. Her hareket için 1, 3, 5, 11 adımlık veriler 10 sefer toplanmıştır. 3 adımlık koşma verisi için örnek vermek gerekirse

sırasıyla sağ, sol, sağ adımlar atılarak 10 defa tekrarlanmıştır. Veri toplama sürecinin sonucunda zaman uzunlukları farklı 150 adet veri elde edilmiştir.

Verilerde dosya isimleri <aktivite-adı><adım-sayısı>step-<iterasyon-sayısı> biçiminde İngilizce olarak isimlendirilmiştir. Örnek vermek gerekirse 5 adımlık yürüme verisinin 3.üncü iterasyonu walk5step-3 olarak isimlendirilmiştir.

Her hareket verisi 6 sütuna sahiptir. Bunlar;

- time: 100 ms saniyede bir gelen verilerin zaman etiketi
- ax: x ekseninde ivmeölçer verisi
- ay: y ekseninde ivmeölçer verisi
- az: z ekseninde ivmeölçer verisi
- gx: x ekseninde jiroskop verisi
- gy: y ekseninde jiroskop verisi
- gz: z ekseninde jiroskop verisi

	time	ax	ay	az	gx	gy	gz
0	0.0	5007.142857	-3275.714286	-6902.857143	1310.714286	-680.000000	-8410
1	100.0	3828.877551	-4298.673469	-11727.551020	3658.673469	4745.714286	-8112
2	200.0	672.849854	-5014.861516	-15700.568513	5959.147230	6347.959184	-8410
3	300.0	-3216.025094	-2483.839546	-14716.079758	5295.523219	7908.688047	-3901
4	400.0	-8303.158989	-638.182565	-11570.336988	811.407784	4422.728030	707

Tablo 1. Veri örneği

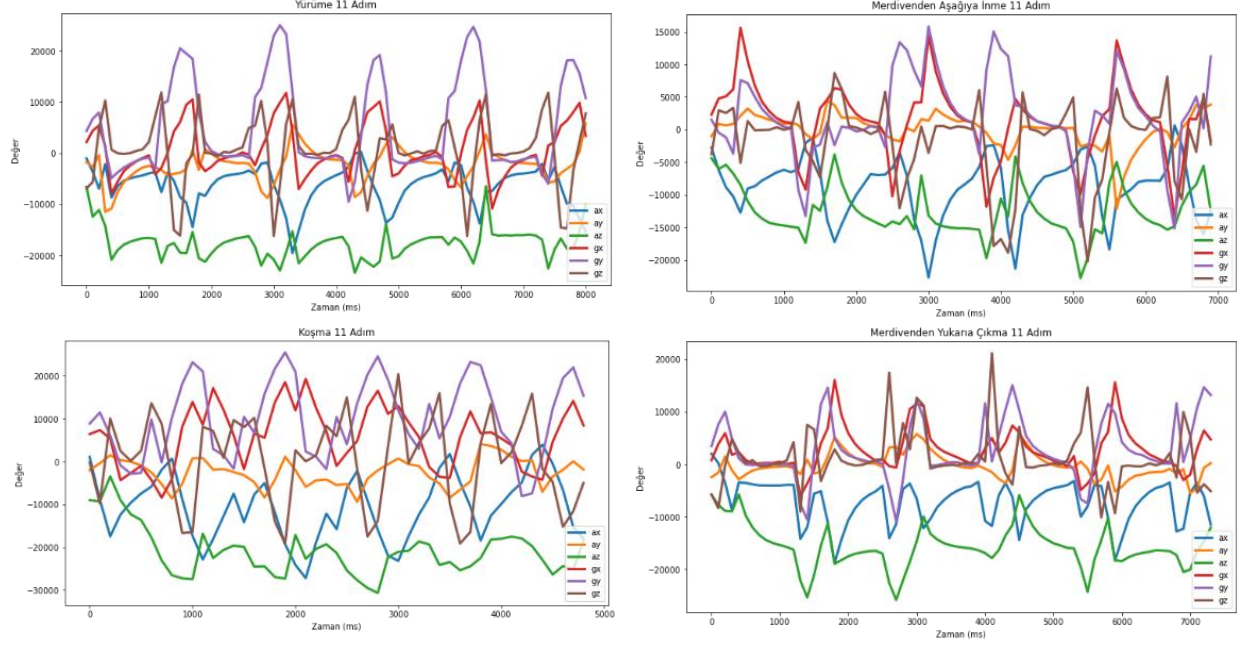
### Veri Ön İşleme

Elde edilen veriler eğitime uygun olmadığından ve bazı hatalar içerdiğinde dolayı bazı ön işlemlerden geçmesi gerekmektedir. Bu yüzden elde edilen verilere sırasıyla aşağıdaki adımlar uygulanmıştır.

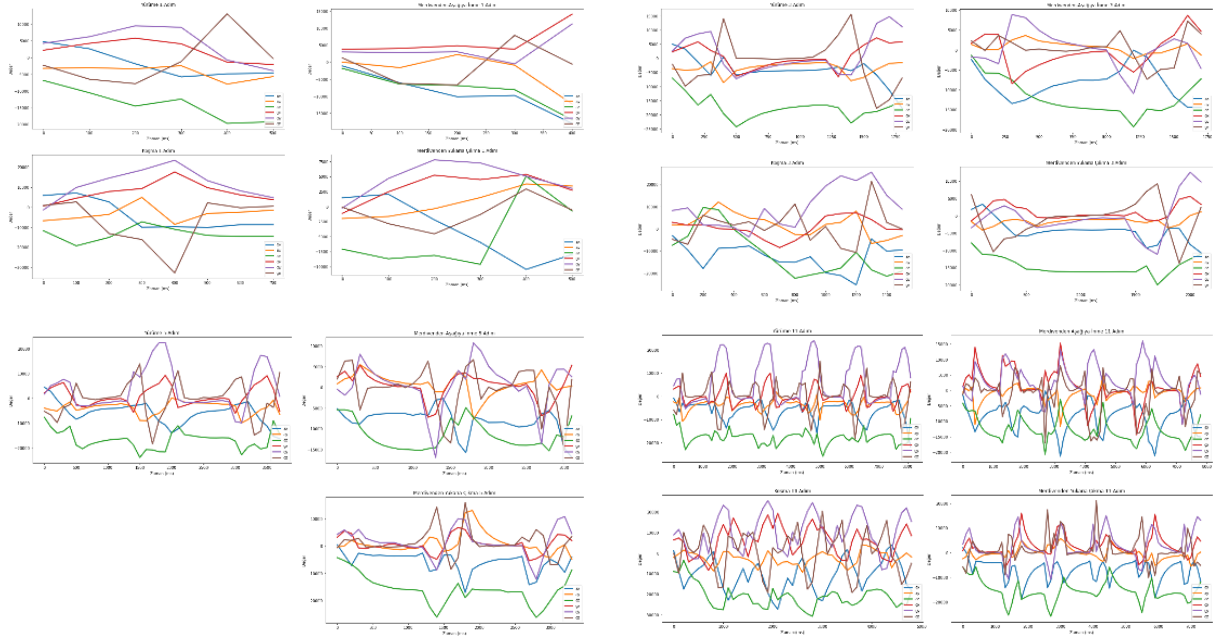
1. Bir aktivitedeki aynı adım sayısındaki verilerinin uzunluklarının ortalamaları alınıp eşit uzunluğa getirilmesi
2. Veriler farklı uzunlukta ve sıradan sınıflandırma algoritmalarının zaman serisi sınıflandırmasını uygun olmamasından dolayı verileri tek satıra indirgenmesi gerekmektedir. Bu konuda birden fazla yöntem uygulanabilir. Bunlardan biri ise istatistikte de sık olarak kullanılan veriyi açıklamaya yardımcı olan ortalama, standart sapma, minimum ve maksimum noktalar, medyan vb. özelliklerin her sütuna uygulanarak ayrı sütunlar halinde tek satıra indirgenmesidir. Bu çalışmada da ax, ay, az, gx, gy, gz, sütunlarının sırasıyla ortalaması, standart sapması, medyanı, minimum ve maksimum noktaları alınarak yan yana sütunlar halinde birleştirilmiştir.
3. Ardından her verinin uzunluğunu içeren sütun ve modeli eğitirken de etiket olarak kullanılacak aktivite sütunu da ekleyip bütün veriler tek bir dosyada birleştirilmiştir.
4. Modeli eğitirken herhangi bir sütunun diğer sütunlar üzerinde etkisinin daha fazla olmaması için veriler normalize edilmiştir.
5. Son olarak veriler bağımlı değişkenler ve bağımsız değişken olarak ayrıca eğitim ve test olarak parçalanmıştır. Bu işlem sonucunda x\_train, x\_test, y\_train, y\_test olarak 4 farklı veri elde edilmiştir. Tüm verinin %20 si test verisi olarak kullanılmıştır.

## BULGULAR

Cihaz üzerinden gelen ham jiroskop ve ivmeölçer verilerinin aktivite ve adım sayılarına göre ayrı ayrı görselleştirmeleri yapılmış ve adım sayıları eşit verilerin diğer aktivitelerle yan yana getirilmesiyle kıyaslamalar yapılarak çizgi grafikler üzerinde farklar gözlemlenmiştir. Oluşturulan grafikler bakıldığında aktiviteler arası farklılıklar ve adımlar çok belirgin olmasa da gözlemlenebilmektedir.



Şekil 3. 11 adımlık yürüme verisi çizgi grafiği

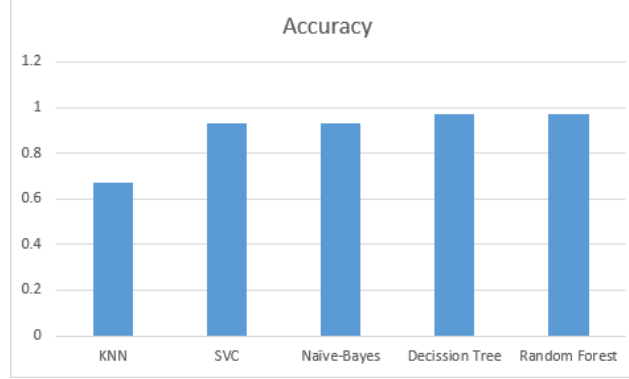


Şekil 4. 1,3,5,11 adımlık yürüme, koşma, merdiven inme ve merdiven çıkma aktivitelerinin verilerinin karşılaştırmalı çizgi grafiği

## Model Eğitimi

Veriler hazırlandıktan sonra sklearn kütüphanesi yardımıyla KNN, SVC, Naïve Bayes, Decission Tree ve Random Forest algoritmalarında modeller oluşturularak test edilmiş ve performansları karşılaştırılmıştır.

Şekil 10 da ki modellerin accuracy değerlerine bakıldığında yapılan testler sonucunda KNN, Decission Tree ve Random Forest algoritmaları %97 lik bir oranla en yüksek doğruluk oranına sahip olduğu görülmektedir.

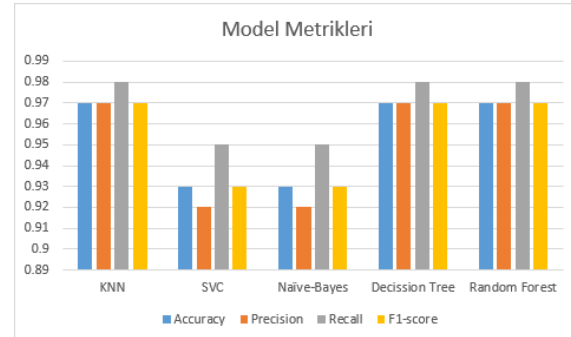


Şekil 5. Modellerin Accuracy değerleri grafiği

Tablo 2 ve Şekil 6 incelendiğinde ise model metriklerinde yine en yüksek başarımlı oranı KNN, Decission Tree ve Random Forest algoritmalarına ait olduğu görülmektedir.

	Accuracy	Precision	Recall	F1-score
<b>KNN</b>	0.97	0.97	0.98	0.97
<b>SVC</b>	0.93	0.92	0.95	0.93
<b>Naïve-Bayes</b>	0.93	0.92	0.95	0.93
<b>Decission Tree</b>	0.97	0.97	0.98	0.97
<b>Random Forest</b>	0.97	0.97	0.98	0.97

Tablo 2. Modellerin Metrik Tablosu



Şekil 6. Model Metrikleri grafiği

Tablo 3 de modellerin confusion matrisleri kıyaslanmıştır. Kıyaslama sonucunda genellikle modelleri aynı veriler üzerinde hata yaptığı söylenebilir.

KNN						SVM					
		Gerçek Değer						Gerçek Değer			
		Downsta	Upstairs	Walk	Run			Downsta	Upstairs	Walk	Run
Tahmin Değeri	Downsta	8	0	0	0	Tahmin Değeri	Downsta	7	0	0	0
	Upstairs	0	7	0	0	Upstairs	0	7	0	0	
	Walk	0	1	10	0	Walk	1	0	10	0	
	Run	0	0	0	4	Run	0	0	0	0	
Naive Bayes						Decision Tree					
		Gerçek Değer						Gerçek Değer			
		Downsta	Upstairs	Walk	Run			Downsta	Upstairs	Walk	Run
Tahmin Değeri	Downsta	7	0	0	1	Tahmin Değeri	Downsta	8	0	0	0
	Upstairs	0	7	0	0	Upstairs	0	7	0	0	
	Walk	1	0	10	0	Walk	1	0	10	0	
	Run	0	0	0	4	Run	0	0	0	0	
Random Forest											
		Gerçek Değer									
		Downsta	Upstairs	Walk	Run						
Tahmin Değeri	Downsta	8	0	0	0						
	Upstairs	0	7	0	0						
	Walk	1	0	10	0						
	Run	0	0	0	4						

Tablo 3. Modellerin Confusion Matrix karşılaştırma tablosu



## SONUÇ

Bu çalışmanın konusu olan akıllı ayakkabı projesinde verileri toplayabilmek için küçük ve maliyeti düşük bir cihaz oluşturulmuştur. Bu cihaz sayesinde 3 eksen ivmeölçer ve 3 eksen jiroskop olmak üzere 6 eksenli zaman etiketleri ile koşma, yürüme, merdivenden inme ve çıkma olarak veriler toplanmıştır. Toplanan bu veriler incelenmiş çeşitli veri düzenleme ve temizleme işlemlerinden geçirilerek bir dosya içerisinde aktivite etiketleri de eklenerek birleştirilmiştir.

Birleştirilen bu veriler ile farklı sınıflandırma algoritmalarında sınıflandırma performansları incelenmiş ve kıyaslanmıştır. Yapılan incelemeler ve kıyaslamalar sonucu genel olarak algoritmaların sınıflandırma başarısı %93'ün üzerinde olduğu görülmüştür.

Bu çalışmanın sonucunda ayakkabı üzerinden aktivite tanımlaması yapılabileceği ortaya konulmuştur. Veriler tek bir kişi üzerinden toplandığı için bir başkasının verileri ile model test edildiğinde aynı başarıyı göstermeyebilir. Bu sebepten dolayı daha fazla kişi üzerinden veri toplanması gerekmektedir. Veri sayısı arttığı zamanda veri madenciliğindeki basit sınıflandırma algoritmaları yeterli kalmayabilir. Daha fazla veri toplandıktan sonra deep learning yöntemleriyle model başarısı iyileştirilebilir. Tabi ki buraya kadar bahsettiklerim hala bir akıllı ayakkabı olabilmesi için yeterli değildir. Bu çalışmada piyasadaki hazır geliştirme kart ve modülleri ile oluşturulan cihaz, küçük olması için ekstra çaba gösterilmesine rağmen ayakkabı üzerinde boyutundan dolayı çok fazla dikkat çekmektedir. Cihazın bilgisayar ortamında baştan profesyonel olarak tasarlanıp tek bir devre kartı olarak basılması gerekmektedir. Ayrıca yazılımsal olarak da bir mobil uygulama ile desteklenerek kullanıcının verilerini takip edebilmesi sağlanabilir ve elde edilen veriler ile kalori hesabı, gidilen mesafe vb. hesaplarda yapılarak kullanıcıya daha fazla veri sunulabilir.

Özellikle son yıllarda sağlık ve spor alanında takip üzerine geliştirilen uygulamalar çok fazla talep edildiği görülmektedir. Bu proje kapsamında ele alınan akıllı ayakkabı ile de bu alanda güzel bir alternatif oluşturularak çeşitli sağlık ve spor alanındaki sorunlara çözüm getirebilmesi mümkündür.

## KAYNAKÇA

A. M. Khan, (2013). "Recognizing physical activities using Wii remote," International Journal of Information and Education Technology, vol. 3, no. 1, pp 60-62.

A. Sharma, Y.-D. Lee, and W.-Y. Chung, (2008). "High accuracy human activity monitoring using neural network," International Conference on Convergence and Hybrid Information Technology, pp 430-435.

Bhatia, N. and Vandana, (2010). "Survey of nearest neighbor techniques", International Journal of Computer Science and Information Security, 8(2):302-305.

Breiman L. Random Forests. (2001). Machine learning, 45(1): 5–32.

Cover, T.M. and Hart, P.E., (1967). "Nearest neighbor pattern classification". IEEE Transactions on Information Theory, IT13(1):21–27.

Liu, H. and Zhang, S., (2012). "Noisy data elimination using mutual k-nearest neighbor for classification mining", Journal of Systems and Software, 85(5):1067-1074.

De la Vega, L.G.M., Raghuraman, S., Balasubramanian, A., and Prabhakaran, B. (2013). Exploring unconstrained mobile sensor based human activity recognition. 3rd International Workshop on Mobile Sensing, 8-11 April 2013, Philadelphia USA.

Chandra B, Gupta M, Gupta MP. (2007). "Robust approach for estimating probabilities in Naive-Bayes classifier". Pattern Recognition and Machine Intelligence. Kolkata, India, 18-22 December.

Dong-Peng Yang, Li Jin-Lin, Lun Ran and Chao Zhou, (2008). Applications of Data Mining Methods in the Evaluation of Client Credibility, Applications of Data Mining in E-Business and Finance C. Soares et al. (Eds.), IOS Press, Amsterdam, pp.35-43

Jennifer R. Kwapisz, Gary M. Weiss, Samuel A. Moore. (2010). Activity Recognition using Cell Phone Accelerometers. Fordham University

- Kwapisz, J. R., Weiss, G. M. and Moore, S. A. (2011). Activity recognition using cell phone accelerometers. *ACM SigKDD Explorations Newsletter*, 12(2), 74-82.
- L. Bao and S. Intille, (2004) "Activity recognition from user-annotated acceleration data," *Pervasive Computing, Lecture Notes in Computer Science*, vol. 3001, pp. 1-17.
- Mitchell, T., (1997). "Machine Learning". McGraw Hill, New York.
- Nisbet, R., Elder, J., and Miner, G. (2009). *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier Inc, Burlington.
- Raj Kumar, Dr. Rajesh Verma; (2012). *Classification Algorithms for Data Mining: A Survey*; Jind Institute of Engg. & Technolog, Jind, Haryana, India.
- Riboni, D., and Bettini, C. (2011). COSAR: hybrid reasoning for context-aware activity recognition. *Personal and Ubiquitous Computing*, 15(3), 271-289.
- Ristianini N, Shawe-Taylor J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge, UK: Cambridge University Press.
- S.-B. Cho and Y.-S. Lee, (2011). Activity recognition using hierarchical hidden Markov models on a smartphone with 3D accelerometer," *Hybrid Artificial Intelligent Systems*, IOS Press, vol. 6678, pp. 460-467.
- Sağbaş E.A, Ballı S. (2016). Akıllı Telefon Algılayıcıları ve Makine Öğrenmesi Kullanılarak Ulaşım Türü Tespiti. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, Baskıda.
- Vapnik V. (1995). *The nature of statistical learning theory*. Springer, New York.
- W. Wu, S. Dasgupta, E.E. Ramirez, C. Peterson, and G.J. Norman, (2012). Classification accuracies of physical activities using smartphone motion sensors, *Journal of Medical Internet Research*, vol. 14, no. 5.
- Yang, J. Y., Wang, J. S. and Chen, Y. P. (2008). Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers. *Pattern recognition letters*, 29(16), 2213-2220.