

Uyku Sağlığı ile Yaşam Tarzı Arasındaki İlişkinin PCA, Naive Bayes ve Rastgele Orman Ağaçları Yöntemleri ile İncelenmesi ve Karşılaştırılması

Discovery and Comparison of the Relationship between Sleep Health and Lifestyle with PCA, Naive Bayes and Random Forest Trees Methods

Serkan AYAN¹ 

Turgay Tugay BİLGİN² 

DOI:10.33461/uybisbbd.1415925

Öz

Makale Bilgileri

Makale Türü:

Araştırma Makalesi

Geliş Tarihi:

07.01.2024

Kabul Tarihi:

29.02.2024

©2023 UYBISBBD
Tüm hakları saklıdır.



Uyku, günlük yaşamın temel bir unsuru olarak kabul edilir ve genel sağlık ile refahın sürdürülmesinde önemli bir rol oynar. Bu araştırma, Kaggle platformundan elde edilen "Uyku Sağlığı Yaşam Tarzı" veri setini kullanarak bir tahmin modeli oluşturmayı, bu modeli Principal Component Analysis (PCA) yöntemi, Naive Bayes yöntemi ve Rastgele Orman Ağaçları yöntemiyle değerlendirmeyi ve görselleştirmeler gerçekleştirmeyi amaçlamaktadır. İncelenen veri seti, KNIME platformunda PCA modülü ile boyut azaltma işlemine tabi tutulmuş ve elde edilen çıktılar sunulmuştur. Uyku kalitesini etkilediği düşünülen öznitelikler arasındaki ilişkiler, korelasyon hesaplamaları ile belirlenmiştir. Ayrıca, veri seti Naive Bayes ve Rastgele Orman Ağaçları yöntemleriyle analiz edilmiş, tahmin sonuçları KNIME ortamında değerlendirilmiştir. Elde edilen sonuçlar tablolar halinde sunulmuştur. Bu karşılaştırmaların dağılım matrisleri, KNIME platformundaki Scatter Plot modülü kullanılarak görselleştirilmiştir. Bu çalışmanın en önemli katkısı uyku verileri içeren veri setlerinde kullanılacak metodolojiler arasında en etkili olan yöntemi belirlemektir. Bulgular, tartışma ve sonuçlar bölümünde detaylı bir şekilde ele alınmıştır.

Anahtar Kelimeler: Uyku Sağlığı, Tahmin Modelleme, Temel Bileşen Analizi, Veri Görselleştirme.

Abstract

Article Info

Paper Type:

Research Paper

Received:

07.01.2021

Accepted:

29.02.2024

©2023 UYBISBBD
All rights reserved.



Sleep is considered a fundamental element of daily life and plays a crucial role in maintaining overall health and well-being. This study aims to develop a predictive model using the "Sleep Health Lifestyle" dataset downloaded from the Kaggle platform. The model is constructed using Principal Component Analysis (PCA), Naive Bayes, and Random Forest methods, and its performance is evaluated. Additionally, the dataset undergoes dimensionality reduction through the PCA module in the KNIME platform, and the results are presented. Relationships between attributes that influence sleep quality are determined through correlation calculations. Furthermore, the dataset is analyzed using the Naive Bayes and Random Forest methods, and the prediction results are assessed using the KNIME platform. The results are presented in tabular form. The scatter matrices of these comparisons are visualized using the Scatter Plot module in the KNIME platform. The primary contribution of this study is to identify the most effective methodology for mining datasets containing sleep-related information. The findings are discussed in the conclusion section.

Keywords: Sleep Health, Predictive Modeling, Principal Component Analysis, Data Visualization.

Atıf/ to Cite (APA): Ayan, S. & Bilgin T.T. (2024). Uyku Sağlığı ile Yaşam Tarzı Arasındaki İlişkinin PCA, Naive Bayes ve Rastgele Orman Ağaçları Yöntemleri ile İncelenmesi ve Karşılaştırılması. Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi, 8(1), 41-56. DOI: 10.33461/uybisbbd.1415925

¹ Araştırma Görevlisi, Mühendislik ve Doğa Bilimleri Fakültesi, Bursa Teknik Üniversitesi, serkan.ayan@btu.edu.tr, Bursa, Türkiye.

² Prof. Dr., Mühendislik ve Doğa Bilimleri Fakültesi, Bursa Teknik Üniversitesi, turgay.bilgin@btu.edu.tr, Bursa, Türkiye.

1. GİRİŞ

Uyku günlük yaşamımızın vazgeçilmez bir bileşenidir ve genel sağlık ve refahımızı korumada önemli bir rol oynar. Yaşam tarzı seçimleri de dahil olmak üzere çeşitli faktörler, uyku kalitesi ve süresi üzerinde önemli etkilere sahip olabilir. Uyku sağlığı ile yaşam tarzı alışkanlıkları arasındaki ilişkiyi anlamak, uykuyla ilgili sonuçları iyileştirmek için desenleri ve potansiyel müdahaleleri belirlemek açısından önemlidir. Buna ek olarak, uyku sağlığı ile yaşam tarzı arasındaki ilişkilere benzer olabilecek Brink-Kjaer vd (2022)'deki çalışma da araştırmacılar, uyku apnesinin kesin teşhisi olan polisomnogramlar (PSG'ler) aracılığıyla yaş ve ölüm riskini tahmin eden derin sinir ağları üzerine bir takım sonuçlar sunmuştur. Bu bağlamda, uyku sağlığının kapsamlı bir şekilde ele alınması ve tanımlanması, yaşam tarzı seçimlerinin ve çevresel faktörlerin uyku kalitesi üzerindeki etkilerini daha iyi anlamamıza ve dolayısıyla bireysel ve toplumsal sağlık sonuçlarını iyileştirmek için etkili stratejiler geliştirmemize olanak tanır (Buysse, 2014).

Uyku sağlığının çok boyutlu doğası ve sosyoekolojik faktörlerin etkileşimi, uyku kalitesi ve sağlık sonuçları arasındaki karmaşık ilişkilerin anlaşılmasında kritik bir rol oynar. Ghose vd (2023) tarafından gerçekleştirilen çalışma, uyku sağlığı ve kendine yeterlilik arasındaki ilişkiyi inceleyerek, uyku sağlığının çok boyutlu doğasını ve bunun bireylerin yaşam tarzı seçimleri üzerindeki etkilerini ortaya koymuştur. Bu çalışma, uyku sağlığı ile ilgili farklı alanlarda kendine yeterliliğin rolünü belirleyerek, uyku kalitesi ve süresi üzerinde önemli etkileri olabilecek yaşam tarzı seçimlerinin anlaşılmasına katkıda bulunur. Diğer yandan, uyku sağlığının çok boyutlu yapısı ve insanların hem kendi aralarında hem de fiziksel çevreleri arasındaki uyum üzerine yapılan kapsamlı çalışma gösteriyor ki uyku sağlığı ve onun geniş çaplı sağlık sonuçları üzerindeki etkisinin daha iyi anlaşılmasına olanak tanır (Hale vd., 2020). Buna ek olarak bu çalışmada incelenen kişiler ile onların uyku sağlığını etkileyecek parametreler arasındaki ilişkilerin doğrudan keşfedilmesi, birtakım zorluklar içermektedir. Bu bağlamda, veri setindeki latent desenleri ve yapısal özellikleri ortaya çıkarmak, uyku ve yaşam tarzı profilleri bakımından benzer grupları saptamak ve öznitelikler arasındaki ilişkileri görsel bir biçimde sergilemek için keşifsel veri analizinin bir parçası olarak Temel Bileşen Analizi (PCA) gibi boyut indirgeme yöntemlerinin uygulanması öngörülmüştür. PCA, çok değişkenli veri setlerinin temel bileşenlerini çıkararak, ana veri yapısını daha az sayıda bileşenle temsil etmeyi ve böylelikle veri seti içerisindeki karmaşıklığı azaltmayı hedeflemektedir. Bu yöntemle, veri setindeki değişkenler arası ilişkilerin ve grupların daha açık bir şekilde tanımlanması mümkün hale gelmiştir.

PCA, çok boyutlu verilerdeki kalıpları bulmak için kullanılan bu teknik, gözlemlenen değişkenlerin olası korelasyonlarını ortadan kaldırarak doğrusal olarak bağımsız değişkenler kümesine dönüştürmek için bir dizi matematiksel işlem yapmaktadır. Bu işlem sonucunda, veri setindeki boyut sayısı azaltılabilir ve verinin çoğu bilgisi korunabilir. Yani bu yeni değişkenlerin varyansını maksimize ederek en önemli bilgileri korur ve diğer değişkenleri atar. Bu nedenle, daha az boyutlu bir veri seti elde edilirken, orijinal verinin büyük bir kısmı korunur. Bro vd (2014) ve Karamizadeh vd (2013) çalışmalarında yazarlar PCA'nın nasıl anlaşılacağı, kullanılacağı ve yorumlanacağı konusunda kapsamlı bir rehber sunmaktadır. Ayrıca Ricciardi vd (2020) tarafından yapılan çalışmalarda kalp damar hastalığı olan veya şüphesi olan hastaların sınıflandırılması için veri madenciliği uygulamasının kullanılmasını test etmektedir. Bu amaçla, başka bir veri madenciliği algoritması olan doğrusal diskriminant analizi (LDA) algoritması ve burada ilgilendiğimiz temel bileşen analizi (PCA) kullanılmıştır. PCA yöntemi kullanılarak veri setinde boyut azaltma yapılmış ve anlamlı bileşenlerin çıkarılması sağlanmıştır. Ayrıca verideki genel varyansa katkılar incelenmiş ve bilgilendirici grafikler ve tablolar aracılığıyla değişkenler arasındaki ilişkiler görselleştirilmiştir. Bu analiz çalışmasında, KNIME platformunun kullanımı, veri analizi sürecini daha kolay ve verimli hale getirmiştir. Bu çalışmanın bulguları, uykuyla ilgili konuların anlaşılmasına katkıda bulunabilir ve sağlıklı uyku alışkanlıklarının teşvik edilmesine yönelik hedefe odaklı müdahalelerin geliştirilmesine rehberlik edebilir.

2. MATERYAL METOT

Bu çalışmada, veri seti KNIME veri platformunda işlenmiş ve PCA yöntemi uygulanmadan önce normalize edilmiştir. Daha sonra, veri görselleştirme işlemleri KNIME'in gömülü düğümleri kullanılarak gerçekleştirilmiştir. Ek olarak, çalışmada Naive Bayes ve Rastgele Orman Algoritması gibi makine öğrenmesi yöntemleri de kullanılmıştır. Bu yöntemler, veri setindeki ilişkileri daha fazla analiz etmek ve sınıflandırma yapmak için kullanılan yaygın algoritmalar arasındadır. Bu yöntemler, veri setindeki öznitelikler arasındaki ilişkileri anlamak ve uyku sağlığıyla ilişkilendirmek için kullanılabilir.

Naive Bayes ve Rastgele Orman Ağaçları sınıflandırma gibi makine öğrenmesi yöntemleri kullanarak veri setindeki öznitelikler arasındaki ilişkiler, verilen bu üç veri madenciliği yöntemlerinin karşılaştırılmasıyla birlikte analiz edilmiştir.

2.1. Veri Seti

Bu çalışmada, bireylerin uyku düzeni ve günlük alışkanlıklarına dair önemli içgörüler sunan Uyku Sağlığı ve Yaşam Tarzı veri seti ele alınmaktadır. Uyku Sağlığı ve Yaşam Tarzı veri seti, 374 satır ve 13 sütundan oluşan uyku ve günlük alışkanlıklara ilişkin değişkenlerin kapsamlı bir koleksiyonunu içermektedir.

Tablo 1. Veri Setinden bir kesit.

Kişi Kimliği	Cinsiyet	Yaş	Meslek	Uyku Süresi	Uyku Kalitesi	Fiziksel Aktivite Seviyesi	Stres Seviyesi	BMI Kategorisi	Kan Basıncı	Kalp Atış Hızı	Günlük Adımlar	Uyku Bozukluğu
1	Male	27	Yazılım Mühendisi	6,1	6	42	6	Fazla kilolu	126/83	77	4200	Hiçbiri
2	Male	28	Doktor	6,2	6	60	8	Normal	125/80	75	10000	Hiçbiri
3	Male	28	Doktor	6,2	6	60	8	Normal	125/80	75	10000	Hiçbiri
4	Male	28	Satış Temsilcisi	5,9	4	30	8	Obez	140/90	85	3000	Uyku Apnesi
5	Male	28	Satış Temsilcisi	5,9	4	30	8	Obez	140/90	85	3000	Uyku Apnesi
6	Male	28	Yazılım Mühendisi	5,9	4	30	8	Obez	140/90	85	3000	Uykusuzluk
7	Male	29	Öğretmen	6,3	6	40	7	Obez	140/90	82	3500	Uykusuzluk
8	Male	29	Doktor	7,8	7	75	6	Normal	120/80	70	8000	Hiçbiri
9	Male	29	Doktor	7,8	7	75	6	Normal	120/80	70	8000	Hiçbiri
10	Male	29	Doktor	7,8	7	75	6	Normal	120/80	70	8000	Hiçbiri
11	Male	29	Doktor	6,1	6	30	8	Normal	120/80	70	8000	Hiçbiri
12	Male	29	Doktor	7,8	7	75	6	Normal	120/80	70	8000	Hiçbiri
13	Male	29	Doktor	6,1	6	30	8	Normal	120/80	70	8000	Hiçbiri
14	Male	29	Doktor	6	6	30	8	Normal	120/80	70	8000	Hiçbiri
15	Male	29	Doktor	6	6	30	8	Normal	120/80	70	8000	Hiçbiri

Veri seti, cinsiyet, yaş, meslek, uyku süresi, uyku kalitesi, fiziksel aktivite düzeyi, stres seviyeleri, BMI kategorisi, kan basıncı, kalp atış hızı, günlük adım sayısı ve uyku bozukluklarının varlığı ya da yokluğu gibi bilgileri içermektedir. Bu değişkenler, uyku sağlığı ile çeşitli yaşam tarzı faktörleri arasındaki ilişkileri araştırmak için zengin bir veri kaynağı sunmaktadır (Tharmalingam, 2023). Ele alınan Uyku Sağlığı ve Yaşam Tarzı veri seti, bireylerin uyku düzeni ve günlük alışkanlıklarına ilişkin verileri derinlemesine incelemeyi amaçlamaktadır. Seçilen 374 kişilik örneklem, hem demografik çeşitliliği hem de uyku ve yaşam tarzı değişkenlerinin geniş bir yelpazesini temsil etmektedir. Bu özel örneklem, uyku kalitesi ve süresi ile yaşam tarzı faktörleri arasındaki ilişkileri analiz etmek ve sağlık sonuçları üzerindeki potansiyel etkilerini aydınlatmak için belirlenmiştir. Nihai amacı, uyku düzenini ve yaşam tarzını iyileştirmek için müdahaleler geliştirirken, bu faktörlerin birbiriyle olan karmaşık etkileşimlerini ortaya koymaktır. Yapılan analizler ve bulgular, uyku sağlığını destekleyen

yaşam tarzı değişikliklerinin tanımlanmasına ve kişiye özel sağlık önerilerinin geliştirilmesine katkı sağlayabilir.

Bu çalışmanın ilgili bölümünde, toplamda 374 birey üzerinde yürütülen araştırmanın bulgularına derinlemesine bir inceleme gerçekleştirilmiştir. Ancak, makalenin görsel materyallerine entegre edilen tablo, bu bireyler arasından seçilen ilk 15'ini kapsayacak şekilde özel bir sınırlamaya tabi tutulmuştur. Söz konusu tablo, katılımcıların mesleki bilgileri, uyku düzenleri, fiziksel aktivite düzeyleri, stres seviyeleri, vücut kitle indeksleri, kan basıncı ölçümleri, kalp atış hızları ve günlük olarak attıkları adım sayıları gibi çok yönlü sağlık ve yaşam tarzı parametrelerini detaylandırmaktadır. Bu öncü 15 kişilik örneklem, genel popülasyon içerisinde sağlık profilleri ve yaşam alışkanlıkları üzerine yapılan kapsamlı analizin doğruluğunu ve genelleştirmesini temsil etmek üzere özenle seçilmiştir. Araştırma sonuçlarının genel geçerliliği ve kapsamı açısından, bu örneklem grubunun temsilci bir alt küme oluşturduğu kabul edilmektedir. Bu veri kesiti, çalışmanın metodolojik dizaynını ve analitik süreçlerini kavramak açısından kritik bir öneme sahiptir. Araştırmanın ilerleyen kısımlarında, bu özel örneklem üzerinden elde edilen bulgular, genel veri setinin geri kalanıyla karşılaştırmalı bir biçimde ele alınmakta ve bu karşılaştırma sonuçları ayrıntılı bir şekilde tartışılmaktadır.

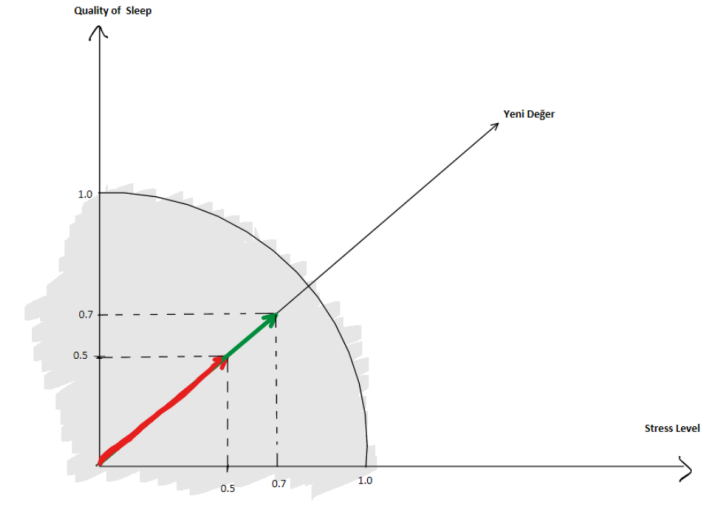
Ek olarak, bu çalışmada, Uyku Sağlığı ve Yaşam Tarzı veri setinin seçilmesinin temel nedeni, kapsamlı ve multidisipliner özellikleri barındırmasıdır. Bu zengin veri seti, uyku sağlığı ile yaşam tarzı arasındaki karmaşık ilişkileri detaylı bir şekilde analiz etme ve bu ilişkilerin sağlık sonuçları üzerindeki potansiyel etkilerini ortaya koyma fırsatı sunmaktadır. Ayrıca, demografik çeşitliliği temsil eden geniş bir örneklem grubunu içermesi, bulguların genel bir popülasyona uygulanabilirliğini artırırken, aynı zamanda çeşitli alt gruplar arasında özgül ilişkilerin incelenmesine olanak tanımaktadır.

2.2. Temel Bileşen Analizi (Principal Component Analysis - PCA)

Temel Bileşen Analizi (Principal Component Analysis-PCA), çok değişkenli veri setlerindeki değişkenler arasındaki ilişkileri anlamak ve veriyi daha az sayıda öz değerlerle temsil etmek için kullanılan bir istatistiksel analiz yöntemidir. Temel Bileşen Analizi, bir veri setindeki değişkenler arasındaki varyansın büyük bir kısmını açıklayan ve değişkenler arasındaki korelasyon yapısını ortaya koyan temel bileşenleri bulmayı amaçlar. Burada veri setindeki değerleri modelleyen matematiksel matrise $X_{D \times N}$ diyelim. X matrisinin x_j , ($j = 1, \dots, N$) sütun değerlerine D boyutlu uzayın vektörleri denir.

$$t = w_1x_1 + w_2x_2 + \dots + w_Nx_N$$

eşitliğine bu x değerlerinin bir lineer kombinasyonu denir ve bu denklem sistemini temsil eden ifade $t = Xw$ eşitliğiyle verilir. X matrisi problemle ilgili varyasyonu içerdiğinden, bu varyasyonun mümkün olduğunca büyük bir kısmının t 'de de bulunması makul görünmektedir. t 'deki bu varyasyon miktarı kayda değer ise, o zaman x değişkenlerinin iyi bir özeti olarak hizmet edebilir. Dolayısıyla, X 'in on üç değişkeni, ilgili bilgilerin çoğunu koruyan tek bir t değişkeni ile değiştirilebilir. Eğer t 'deki bu varyasyon miktarı kayda değer ise, o zaman x değişkenlerinin iyi bir özeti olarak hizmet edebilir. Dolayısıyla, X 'in on dört değişkeni, ilgili bilgilerin çoğunu koruyan tek bir t değişkeni ile değiştirilebilir. t 'deki varyasyon, varyansı olan $var(t)$ ile ölçülebilir, istatistikte olağan şekilde tanımlanmıştır. O zaman problem, w_1, \dots, w_N optimal ağırlıklarını seçerek bu varyansı maksimize etmeye dönüşür. Bununla birlikte, bir uyarı vardır, çünkü optimal bir w değerini keyfi büyük bir sayı ile çarpmak t 'nin varyansını da keyfi büyük yapacaktır. Dolayısıyla, düzgün bir problem için ağırlıkların normalize edilmesi gerekir. Bu, normlarının, yani kareler toplamı değerlerinin bir olmasını gerektirerek yapılır (bkz. Şekil 1). Şekil Microsoft One Note kullanılarak Bro ve Smilde, (2014) makalesinden motivasyon alınarak çizilmiştir.



Şekil 1. Birim Vektör Konsepti

Burada ele alınan uzaklıkları ifade etmek için $\| \cdot \|$ Öklid normu kullanılmıştır. PCA, öncelikle veri setindeki değişkenler arasındaki korelasyon matrisini oluşturur:

$$var(\mathbf{x}) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T \quad 2.1$$

\bar{x} , i -inci sınıftaki özneliklerin ortalamasıdır. Orijinal uzaydaki (2.1) Korelasyon matrisi, değişkenlerin birbirleriyle nasıl ilişkilendiğini gösteren bir matristir. Daha sonra, bu matris üzerinde yeni uzaya izdüşürülen veriler üzerine eigenvalue-eigenvector analizi uygulanır. Böylece biçimsel problem

$$\operatorname{argmax}_{\|w\|=1} var(\mathbf{t}) \quad 2.2$$

eşitliğiyle incelenir. Burada ifade t varyansını maksimize eden w vektörünün bulunması olarak okunmalıdır ($\|w\|=1$) olması gerektiğine dikkat edin). argmax fonksiyonu, maksimizasyon fonksiyonunun w argümanını döndürmek için kullanılan matematiksel gösterimdir. Bu, $t = Xw$ eşitliğini kullanarak daha açık hale getirilebilir:

$$\operatorname{argmax}_{\|w\|=1} (\mathbf{t}^T \mathbf{t}) = \operatorname{argmax}_{\|w\|=1} \mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}. \quad 2.3$$

Bu ifadeyi özetleyecek olursak, eski uzaydaki verinin yeni uzaya izdüşümünü almak için uygun bir w vektörünü bulunarak optimizasyon problemi çözülmektedir, burada; yeni eksene yapılan izdüşüm sonrası varyans, çözülen denklem sisteminde elde edilen özdeğerdir. Bu sayede burada bir özdeğer problemi ile karşılaşmış oluyoruz.

2.3. Naive Bayes

Naive Bayes sınıflandırıcısı, bir öğenin hangi sınıfa ait olduğunu belirlemek için kullanılan bir makine öğrenmesi algoritmasıdır. Bu sınıflandırıcı, Bayes teoremi temelinde çalışır ve her özellik için sınıf olasılıklarını hesaplar. Ardından, bu olasılıkları kullanarak en yüksek olasılığa sahip sınıfı tahmin eder. "Naive" terimi, algoritmanın her özelliğin diğer özelliklerden bağımsız olduğunu varsayması nedeniyle kullanılır. Bu varsayım gerçekte doğru değildir, ancak uygulamada yüksek başarı oranları elde edilir. Naive Bayes Sınıflandırıcısı, aşağıdaki denklemi ifade eden Bayes Teoreminden esinlenmiştir:

$$P(A | B) = \frac{P(B | A)P(A)}{P(B)} \quad 2.4$$

Bu denklem, anlaşılmasını kolaylaştırmak için X (girdi değişkenleri) ve Y (çıkıktı değişkeni) kullanılarak yeniden yazılabilir. Daha basit bir ifadeyle, bu denklem X girdi özellikleri göz önüne alındığında Y olasılığını çözmektedir.

$$P(Y | X) = \frac{P(X | Y)P(Y)}{P(X)} \quad 2.5$$

Sınıf göz önüne alındığında değişkenlerin bağımsız olduğu Naive varsayımı (dolayısıyla adı) nedeniyle, $P(X|Y)$ 'yi aşağıdaki gibi yeniden yazabiliriz:

$$P(X | Y) = P(X_1 | Y)P(X_2 | Y) \dots P(X_n | Y) \quad 2.6$$

Naive Bayes'in amacı maksimum olasılıkla Y sınıfını seçmektir. (2.2) denkleminde de ifade edildiği gibi argmax basitçe, bir hedef fonksiyondan maksimum değeri veren argümanı bulan bir fonksiyondur. Bu durumda, maksimum Y değerini bulmak için

$$Y = \underset{Y}{\operatorname{argmax}} [P(Y) \prod_{i=1}^n P(X_i | Y)] \quad 2.7$$

eşitliğiyle verilen denklem kullanılır (Rish., J., 2001). Bu çalışmada ele alınan veri setiyle birlikte, KNIME ortamında Naive Bayes modülü kullanılarak verilerin test edilmesi ve tahmin edilmesi ile elde edilen sonuçlar analiz edilmiştir.

2.4. Rastgele Orman Ağaçları (ROA)

Rastgele Ormanlar sınıflandırma, regresyon ve diğer görevler için kullanılan popüler bir makine öğrenimi algoritmasıdır. Breimann, L. (2001) tarafından yapılan çalışma gösteriyor ki birden fazla karar ağacını birleştirerek çalışır ve bu sayede tek bir karar ağacının karşılaşılabileceği aşırı uyum (overfitting) sorununu azaltır. Rastgele Ormanlar, hem özellik uzayını hem de eğitim örneklerini rastgele seçerek bir "orman" oluşturur. Bu, her bir karar ağacının eğitim sırasında farklı veri kümelerinden ve özelliklerden öğrenmesini sağlar. Bununla birlikte, Hastie vd (2009) tarafından yapılan çalışmada ağaç öğrenimi veri madenciliği için kullanıma hazır bir prosedür olarak hizmet etme gereksinimlerini karşılamaya en yakın yöntemdir, çünkü ölçekleme ve özellik değerlerinin diğer çeşitli dönüşümleri altında değişmezdir, ilgisiz özelliklerin dahil edilmesine karşı dayanıklıdır ve denetlenebilir modeller üretir.

Çok derin yetiştirilen ağaçlar oldukça düzensiz modeller öğrenme eğilimindedir: eğitim setlerine aşırı uyum sağlarlar, yani düşük önyargıya, ancak çok yüksek varyansa sahiptirler. Rastgele ormanlar, varyansı azaltmak amacıyla aynı eğitim setinin farklı bölümleri üzerinde eğitilen birden fazla derin karar ağacının ortalamasını almanın bir yoludur. Bu, önyargıda küçük bir artış ve yorumlanabilirlikte bir miktar kayıp pahasına gerçekleşir, ancak genellikle nihai modeldeki performansı büyük ölçüde artırır.

Bu sınıflandırıcı, eğitim verilerinin bootstrapped örneği ve rastgele seçilmiş özelliklerin alt kümesi kullanılarak büyütülen bir karar ağacı topluluğu olarak düşünülebilir. Sınıflandırıcı daha sonra ağaçların en popüler sınıf için oy kullanmasına izin verir, bu da arazi örtüsü sınıflandırması için artan doğruluk sağlar. ROA sınıflandırıcısı, diğer sınıflandırıcılarla karşılaştırıldığında daha az aşırı uyum eğilimindedir ve birçok girdi değişkenini ele alabilir. Ayrıca yüksek doğruluk oranına sahiptir ve ikili ve çoklu sınıflandırma problemleri için kullanılabilir. Ek olarak, rastgele orman sınıflandırıcısı hesaplama açısından verimlidir ve eksik verileri imputasyon yapmadan ele alabilir Pal, M. (2005). Burada yapılan araştırma çalışmasında da ele alınan veri seti üzerine, KNIME ortamında ROA modülü uygulanarak sonuçlar elde edilmiş ve diğer sınıflandırma modelleriyle kıyaslanmıştır.

2.5. KNIME

KNIME, açık kaynaklı bir yazılımdır ve kullanıcıların veri analiz süreçlerini görsel bir arabirim üzerinden sürükle-bırak yöntemiyle tasarımlarına olanak tanır. Fillbrunn (2017) tarafından yapılan çalışmada, KNIME'in yaşam bilimleri verilerinin çapraz alan analizinde tekrarlanabilirliği sağlama kapasitesi detaylandırılmıştır. Bu, KNIME'in görsel programlama aracı olarak veri madenciliği ve

makine öğrenmesi iş akışlarını yönetmede etkin bir çözüm sunduğunu gösterir, bu sayede veri madenciliği ve makine öğrenmesi iş akışlarını etkili bir şekilde yönetmeyi sağlar. Sınıflandırma, regresyon, kümeleme, boyut azaltma ve daha fazlası gibi popüler algoritmaları uygulamak için önceden oluşturulmuş bileşenler mevcuttur. Ayrıca, kullanıcılar kendi özel algoritmalarını entegre etmek için açık API'leri kullanabilir. Maraza-Quispe vd (2022) tarafından yapılan çalışmada yazarlar öğrenme analitiği temelli bir tahmin modelinin KNIME üzerinde uygulandığına dair bir araştırma sunmuşlardır. Bu çalışma, sanal öğrenme ortamlarında zamanında karar alma için öğrenme analitiğinin kullanımının önemini vurgulamaktadır. Ek olarak Dietz vd (2020) tarafından yapılan çalışmada KNIME, kimyo-informatik ve kütle spektrometrisi gibi çeşitli topluluklarda karmaşık uçtan uca talepleri karşılamada başarılı olduğu gösterilmiştir. Yazarlar biyogörüntü analizi topluluğundaki benzer ihtiyaçları, ImageJ'i KNIME Analytics Platforma entegre eden KNIME Image Processing uzantısını oluşturmuştur.

Bu platform, veri hazırlama, makine öğrenimi modellemesi, veri görselleştirme ve analizi gibi bir dizi işlevselliği kapsayan uçtan uca çözümler sunar. Özellikle, Python, R gibi programlama dillerine hakimiyet gerektirmeksizin, makine öğrenimi modelleri oluşturabilme imkanı sağlar. Bu, KNIME'in kullanıcı dostu Grafik Kullanıcı Arayüzü (GUI) ile mümkün kılınmaktadır; arayüz, çeşitli kullanım senaryoları için gerekli olan makine öğrenimi algoritmalarının geniş bir yelpazesini destekler. KNIME'in GUI'si, veri bilimi süreçlerini kolaylaştıran ve kullanıcıların veri bilimi iş akışlarını görsel olarak oluşturabilmelerine olanak tanıyan düğümler (node) adı verilen yapı taşlarına dayanır. Bu düğümler, veri ön işleme, veri temizleme, özellik mühendisliği, makine öğrenimi algoritmalarının uygulanması ve istatistiksel analizler gibi çeşitli işlevleri yerine getirebilir. Kullanıcılar, bu düğümleri bir araya getirerek karmaşık veri işleme ve analiz iş akışlarını, kod yazmadan grafiksel bir arayüz aracılığıyla kolayca oluşturabilirler. KNIME, veri bilimcilerin ve analistlerin .txt, .csv, .arff, .table gibi uzantılı dosyalardan veri setlerini yüklemelerine, ön işlemlerine, biçimlendirmelerine ve analiz etmelerine olanak tanıyan modüller sunar Kaya ve Özel (2014). Bu modüller, kullanıcıların veri üzerinde kapsamlı manipülasyonlar yapmasını sağlayarak, ham veriden anlamlı bilgiler çıkarmalarına yardımcı olur.

3. VERİ SETİNİN KNIME PLATFORMUNDA ANALİZİ

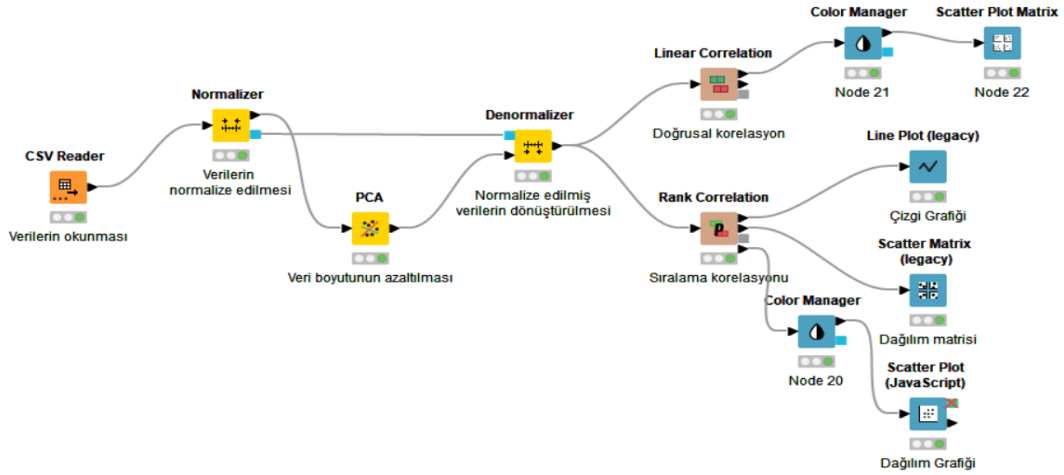
KNIME platformu, kullanıcıların özel düğümler oluşturmasına ve böylece platformun işlevselliğini kendi ihtiyaçlarına göre genişletmesine olanak tanır. Bu özelleştirme yeteneği, KNIME'i çeşitli endüstri ve araştırma uygulamaları için uygun hale getirmektedir. Özetle, KNIME platformu, kullanıcılara veri analizi ve modelleme süreçlerini uçtan uca, esnek ve etkili bir şekilde gerçekleştirme olanağı sunar. Platformun sunduğu geniş düğüm yelpazesi, kullanıcıların karmaşık veri setleri üzerinde analizler yapmalarına ve güçlü tahmin modelleri geliştirmelerine olanak tanır. Bu özellikler, KNIME'i sadece veri bilimcileri ve analistler için değil, aynı zamanda araştırmacılar için de değerli bir araç haline getirir.

Bu bölümde veri setinden elde edilen bulguların KNIME platformu üzerindeki incelemeleri detaylandırılmıştır. Veri setinin incelenmesi, iş akışının oluşturulması ve elde edilen sonuçların incelenerek görselleştirilmesi süreçleri, KNIME platformunun çok yönlülüğünü ve uygulanabilir bir arayüzünün olduğunu yansıtmaktadır.

3.1. İş Akışının Oluşturulması

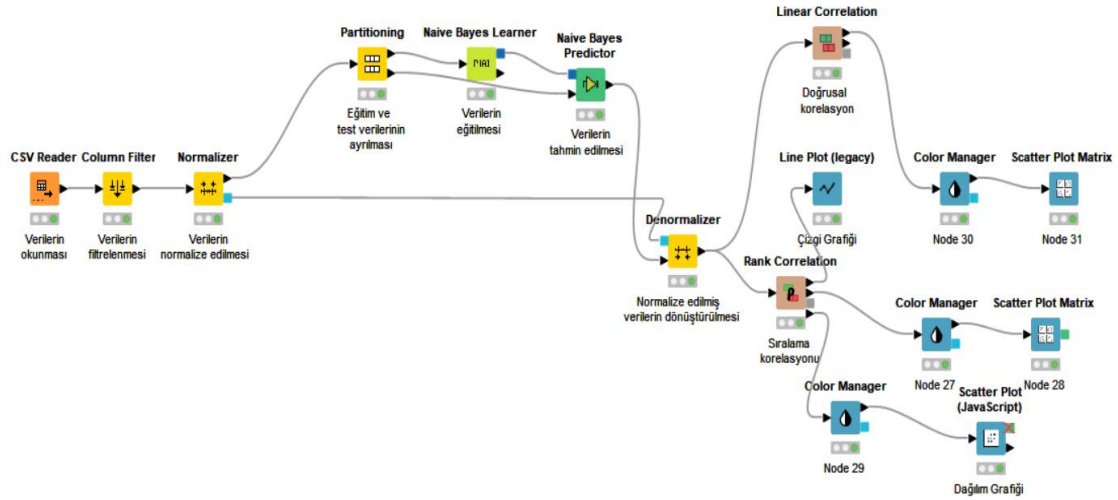
Şekil 2, 3 ve 4, veri işleme ve makine öğrenimi süreçlerini gösteren akış şemalarını temsil etmektedir. Her biri, veri setini işlemek ve çeşitli tahmin modelleri oluşturmak için kullanılan adımları ve kullanılan araçları göstermektedir.

Veri seti, KNIME iş akışı platformunda CSV Reader kullanılarak aşağıdaki görsellerde olduğu gibi okunmuştur. Bu işlem, veri setinin KNIME platformunda kullanılmak üzere hazır hale getirilmesini sağlar.



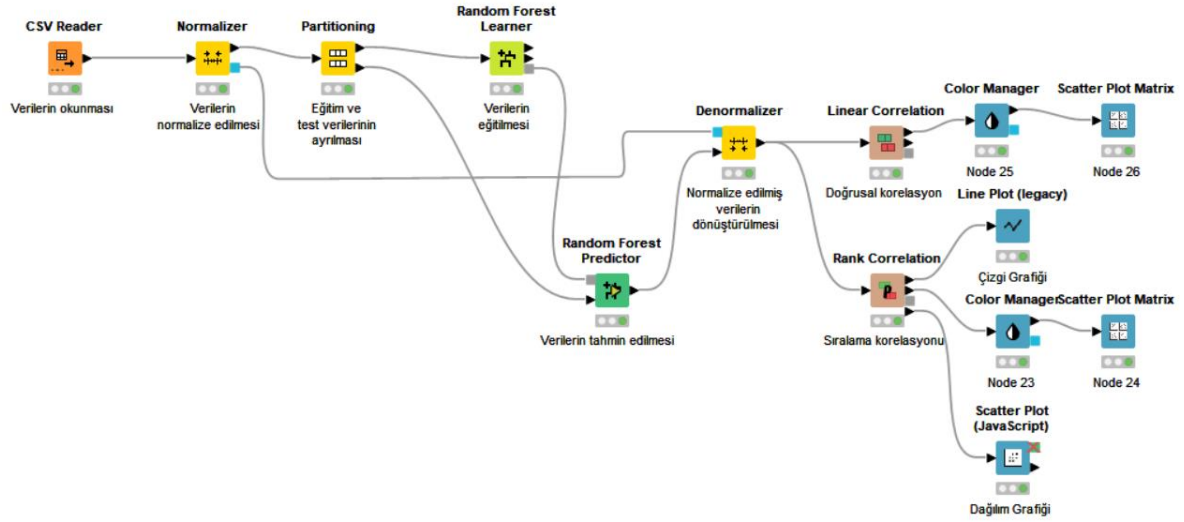
Şekil 2. KNIME Platformunda Veri Setinde PCA uygulaması.

Şekil 2’de, veri setini bir CSV dosyası olarak okuyan ve verileri normalleştiren bir akış başlatılmaktadır. Daha sonra, Temel Bileşen Analizi (PCA) ile veri boyutunun azaltılması gerçekleşiyor. Sonrasında ise, verilerin normalleştirme işleminden önceki hallerine dönüştürülmesi (denormalize edilmesi) sağlanmaktadır. Aynı zamanda, veriler üzerinde doğrusal ve sıralama korelasyonları gibi çeşitli istatistiksel analizler yapılmakta ve sonuçlar çeşitli görselleştirme yöntemleri ile sunulmaktadır. Şekil 3’te, veri seti CSV okuyucu ile okunduktan sonra, sütun filtresi uygulanarak belirli veriler seçiliyor ve normalleştirme işlemi yapılıyor.



Şekil 3. Veri Seti Üzerine Naive Bayes Yönteminin Uygulaması.

Bu süreçten sonra, veriler eğitim ve test verileri olarak bölünüyor (partitioning) ve Naive Bayes sınıflandırıcı algoritması ile bir model eğitilmektedir. Modelin tahmin performansı, denormalize edilmiş veriler üzerinde değerlendirilerek çeşitli istatistiksel analizler ve görselleştirmeler uygulanmaktadır.



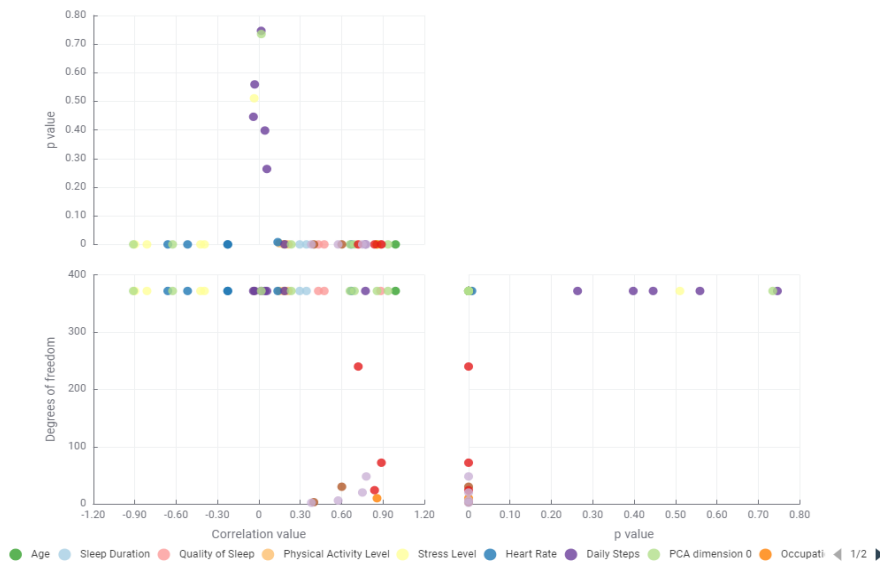
Şekil 4. Veri Setinde ROA uygulaması

Şekil 4’te de benzer bir süreç izleniyor: Veri seti CSV okuyucu ile yüklenerek, normalleştiriliyor ve eğitim/test olarak bölünüyor. Ancak bu sefer, model olarak ROA algoritması kullanılıyor. Model tahminlerinin değerlendirilmesi sonrasında denormalizasyon yapılıyor ve yine çeşitli istatistiksel analizler ve görselleştirmeler sunuluyor.

4. BULGULAR

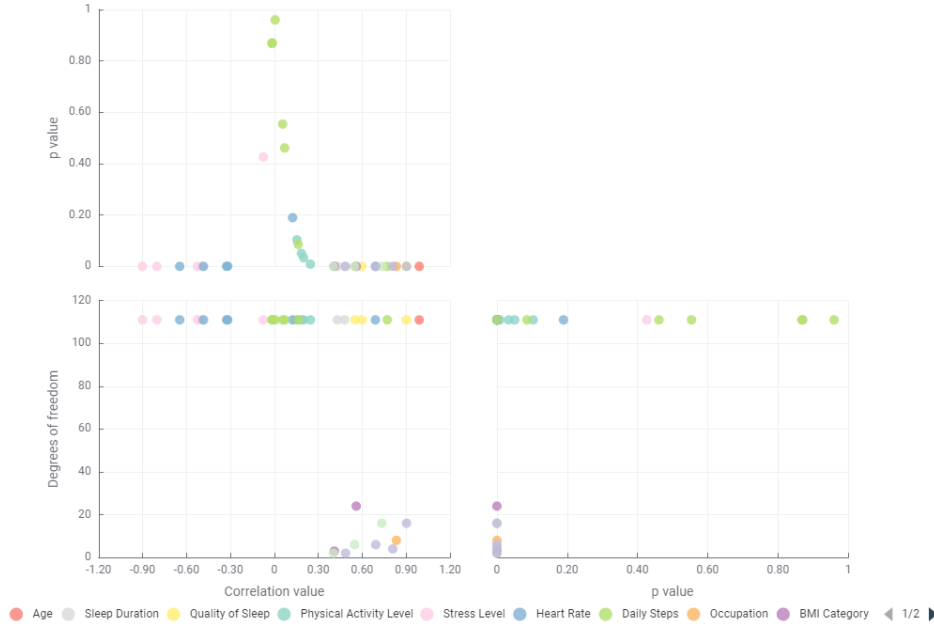
Veri setine PCA, Naive Bayes ve ROA yöntemleri kullanılıp veri üzerinde denormalizasyon uygulandıktan sonra doğrusal ve rank korelasyon sonuçlarının analizi sayesinde uyku sağlığıyla yaşam kalitesi arasındaki bağlantılı olduğu düşünülen parametreler arasındaki ilişkinin görselleştirilmesi bu bölümde elde edilmiştir. Şekil 5’te denormalize edilen verinin linear korelasyon uygulanarak p değeri ve korelasyon değeriyle birlikte elde edilen dağılım matrisi verilmiştir. P değeri, hiçbir etki veya fark olmadığı varsayımı (sıfır hipotezi) altında, gerçekte gözlemlenene eşit veya daha uç bir sonuç elde etme olasılığı olarak tanımlanır. P olasılık anlamına gelir ve gruplar arasında gözlenen herhangi bir farkın şansa bağlı olma olasılığını ölçer.

Şekil 5, 6 ve 7’de PCA, Naive Bayes ve ROA ile elde edilen Doğrusal Korelasyon değerleri ile P değerlerinin verildiği görseller KNIME platformundaki Scatter Plot Matrix modülü ile veri setindeki çoklu değişkenlerin birbiriyle olan ilişkileri görselleştirilmiştir.



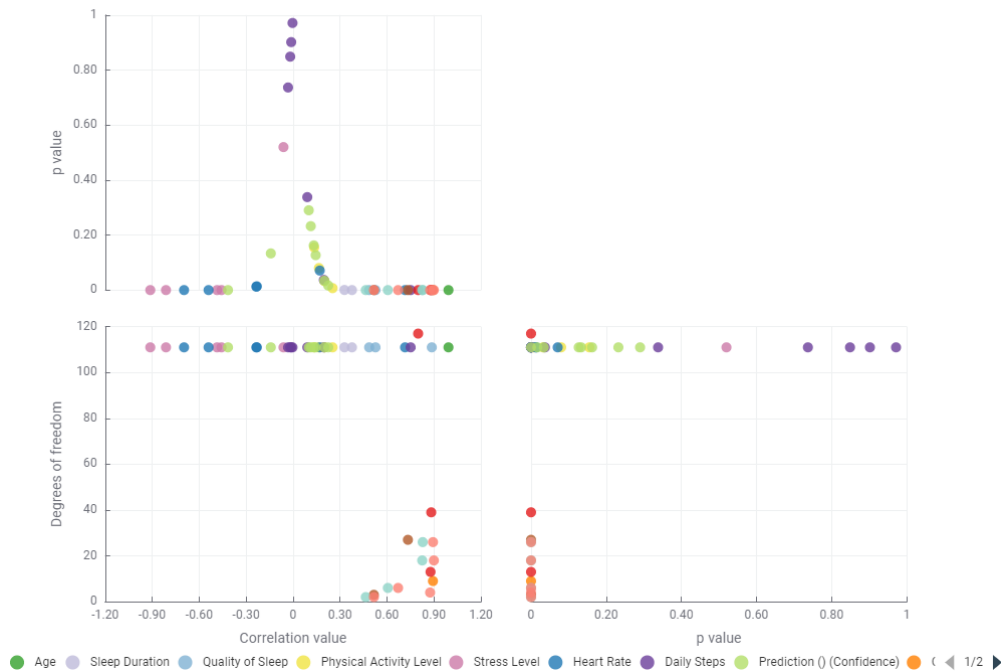
Şekil 5. PCA Sonucunda Elde Edilen Dağılım Matrisi

Birinci ve ikinci sütun isimleri olarak Person ID, Gender, Age, Occupation, Sleep Duration, Quality of Sleep, Physical Activity, Stress Level, BMI Category, Blood Pressure, Heart Rate ve Daily Steps ifadeleri PCA yöntemi uygulandıktan sonra doğrusal korelasyon ölçülerinden elde edilen karşılaştırmaların grafiği olarak verilmiştir. Şekil 6 ve 7'de Naive Bayes ve ROA yöntemleri için benzer dağılım matrisleri farklı doğrusal korelasyon değerleri için görülmektedir.



Şekil 6. Naive Bayes Yöntemiyle Elde Edilen Dağılım Matrisi

Elde edilen bu korelasyon sonuçlarına göre uyku sağlığı ve yaşam tarzı ile ilgili verilen öznitekler arasındaki ilişkilerin karşılaştırılması Tablo 2’de verilmiştir.



Şekil 7. ROA ile Elde Edilen Dağılım Matrisi

Buradaki bilgilere göre birbirleri ile ilişkili olan sütun başlıkları arasında korelasyon değeri pozitif olanların P değerinin sıfır ya da sıfıra çok yakın olduğu gözlemlenmiştir. Tablo 2, üç farklı istatistiksel yöntem (PCA, Naive Bayes ve ROA) kullanılarak elde edilen korelasyon ve p-değerleri karşılaştırmasını sunmaktadır. Sütun isimleri altında verilen parametreler arasındaki ilişkiler, PCA

(Ana Bileşen Analizi), Naive Bayes ve ROA (Rastgele Orman Ağaçları Analizi) yöntemleri ile incelenmiştir. Her bir parametre çifti için, bu üç yöntem kullanılarak hesaplanan korelasyon değerleri tabloda gösterilmiştir. Ayrıca, bu ilişkinin istatistiksel anlamlılığını belirlemek için hesaplanan p-değerleri de ilgili sütunlarda yer almaktadır.

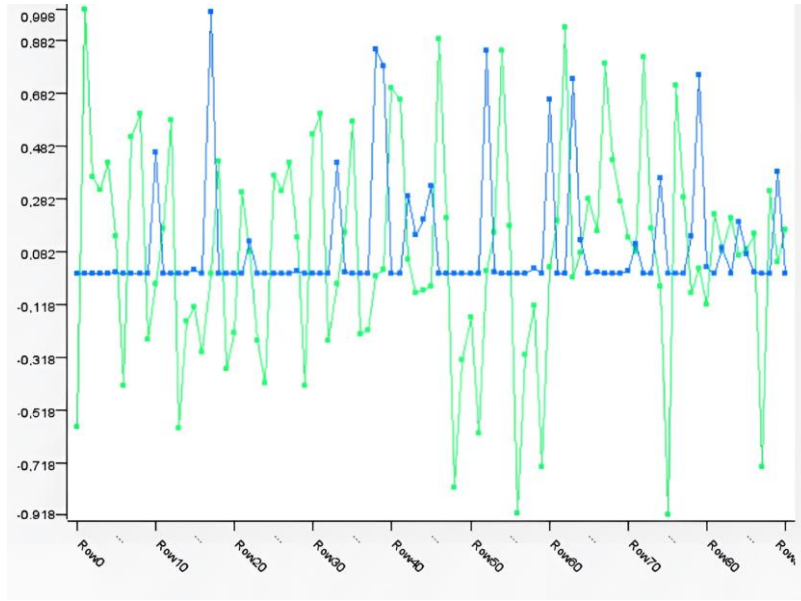
Tablo 2. PCA, Naive Bayes ve ROA Sonuçlarının Karşılaştırılması

Sütun isimleri	Korelasyon değeri			P değeri		
	PCA	Naive Bayes	ROA	PCA	Naive Bayes	ROA
Cinsiyet - Kan Basıncı	0.8385	-	0.8779	0	-	5.0259E-13
Kişi Kimliği - Yaş	0.9905	0.9892	0.9924	0	0	0
Yaş - Günlük Adımlar	0.0579	0.0701	-0.0034	0.2634	0.4611	0.9714
Meslek- Uyku Bozukluğu	0.7505	0.7339	0.82472	0	0	0
Uyku Süresi - Stres Seviyesi	-0.8110	-0.8030	-0.8117	1.2378E-88	1.0456E-26	1.0979E-27
Uyku Kalitesi – Stres Seviyesi	-0.8988	-0.9010	-0.9120	2.8801E-135	4.6215E-42	9.0116E-45
Fiziksel Aktivite Seviyesi – Günlük Adımlar	0.7727	0.7707	0.7509	0	0	0
Stres Seviyesi - Kalp Atış Hızı	0.67002	0.6898	0.7152	0	0	0
Kalp Atış Hızı – Günlük Adımlar	-0.0303	-0.0158	0.0909	0.5590	0.8677	0.3380

Örneğin, 'Cinsiyet - Kan Basıncı' parametre çiftinde PCA yöntemi ile elde edilen korelasyon değeri 0.8385 iken, ROA yöntemi ile bu değer 0.8779 olarak hesaplanmıştır. PCA için p-değeri verilmemişken, ROA yöntemi için çok düşük bir p-değeri ($5.0259E - 13$) gözlemlenmiştir, bu da bu iki değişken arasında güçlü bir ilişki olduğunun istatistiksel olarak anlamlı bir göstergesidir. Kişi Kimliği - Yaş ve Yaş - Günlük Adımlar gibi bazı parametreler için, PCA yöntemi ile elde edilen korelasyon değerleri sırasıyla 0.9905 ve 0.0579 olarak bulunmuştur. Naive Bayes ve ROA yöntemleri ise bu ilişkiler için farklı korelasyon değerleri ve p-değerleri sağlamıştır. Uyku Süresi - Stres Seviyesi ve Uyku Kalitesi - Stres Seviyesi gibi negatif ilişkiler gösteren parametre çiftleri, üç yöntem tarafından benzer şekilde yüksek negatif korelasyonlar ile işaretlenmiş ve bu ilişkilerin istatistiksel olarak son derece anlamlı olduğu gösterilmiş ($p < 0.05$). Fiziksel Aktivite Seviyesi - Günlük Adımlar parametresi pozitif bir ilişki sergilemiş ve PCA, Naive Bayes, ve ROA yöntemlerinin her biri için korelasyon değerleri sırasıyla 0.7727, 0.7707 ve 0.7509 olarak hesaplanmıştır, ancak bu ilişki için p-değeri sonuçlandırılmamıştır. Son olarak, Stres Seviyesi - Kalp Atış Hızı ve Kalp Atış Hızı - Günlük Adımlar gibi parametreler arasında düşük korelasyon değerleri gözlemlenmiş ve bu ilişkilerin p-değerleri, ilişkinin istatistiksel olarak anlamlı olmadığını göstermektedir.

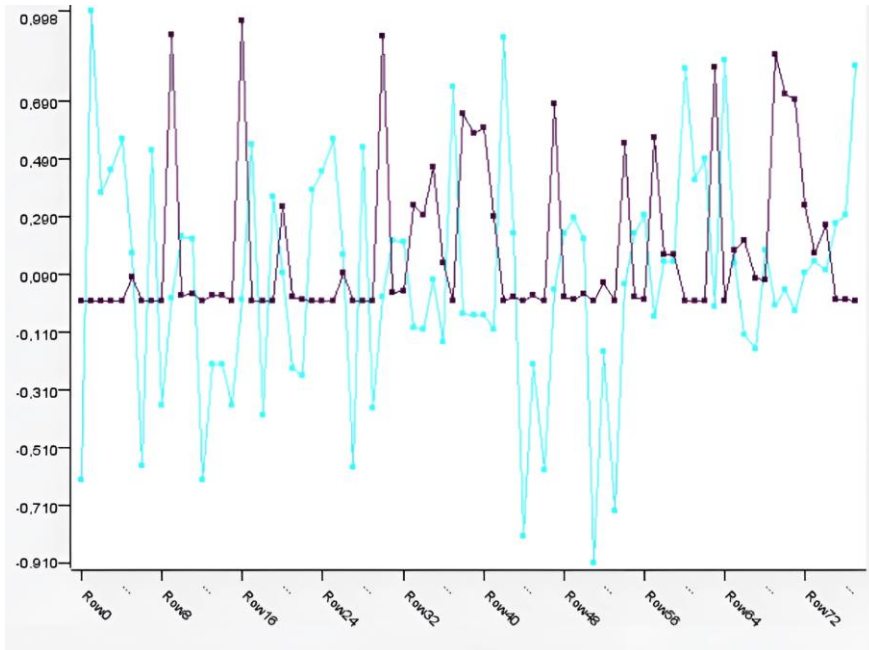
Burada bu iki öznelik için negatif korelasyonun olduğunu ROA yöntemi daha iyi sonuçlar vermiştir. Ayrıca bu bulunan sonuçların P değerine bakılacak olursa ROA yönteminden elde edilen sonuç diğer yöntemlere göre daha iyi olduğu görülmektedir. Doğrusal korelasyon incelemesinin yanında bu veri seti için Rank Korelasyon ile veri analizi de yapılmıştır. Rank korelasyonu, verilerin sıralama düzenini kullanarak ilişkiyi ölçer. Bu korelasyon yöntemi lineer olmayan ilişkileri yakalamak için daha uygundur. Örneğin, değişkenler arasında monoton bir ilişki varsa, ancak bu ilişki tam olarak doğrusal değilse, rank korelasyonu daha doğru bir ölçüm sağlar.

Aşağıda üç yöntem için Rank Korelasyon modülü kullanılarak P ve Korelasyon değerlerinin KNIME ortamındaki Line Plot modülü ile elde edilen görüntüleri verilmektedir.



Şekil 8. PCA Rank Korelasyon P değeri ve Korelasyon Sonucu Karşılaştırılması

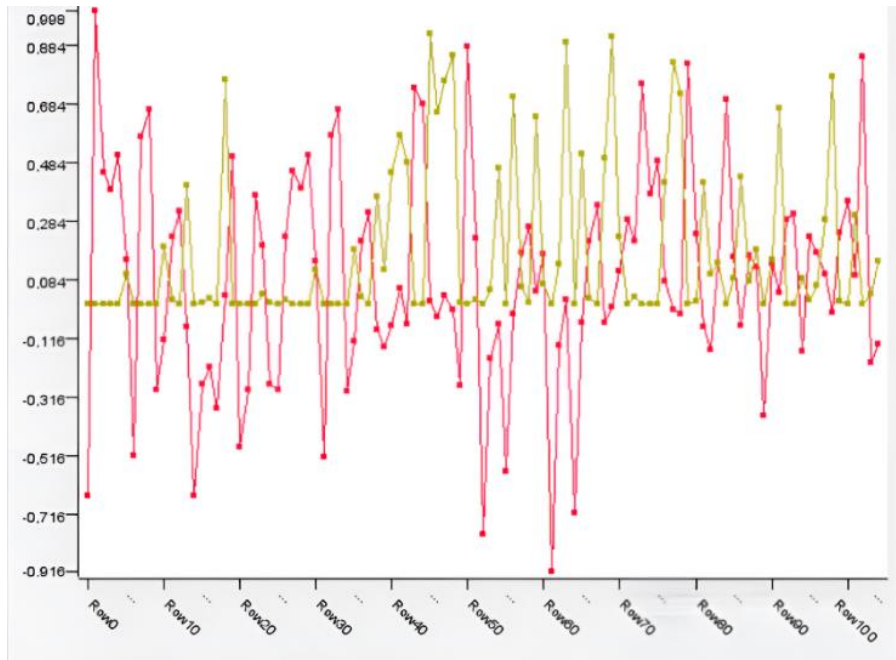
Şekil 8’da PCA yöntemi ile elde edilen sonuçların görselleştirilmiş hali verilmiştir. Scatter Plot modülü ile elde edilen bu grafikte eksik olarak verilen verilerin tahmini interpolasyon yöntemi yapılarak elde edilmiştir. Şekil 9 da Naive Bayes yöntemiyle elde edilen sonuçları içeren bölümü görülmektedir. Burada dikey eksen P ve Korelasyon değerlerini iki boyutlu uzayda gösterirken, yatay eksen öznitelikleri temsil eden satırları göstermektedir. Dikkat edileceği üzere, P değeri karşılaştırılan öznitelikler arasındaki korelasyon sonucunun şansa bağlı olup olmadığını kontrol ettiği için elde edilen sonuçlar $[0,1]$ kapalı aralıktadır.



Şekil 9. Naive Bayes Rank Korelasyon P değeri ve Korelasyon Sonucu Karşılaştırılması

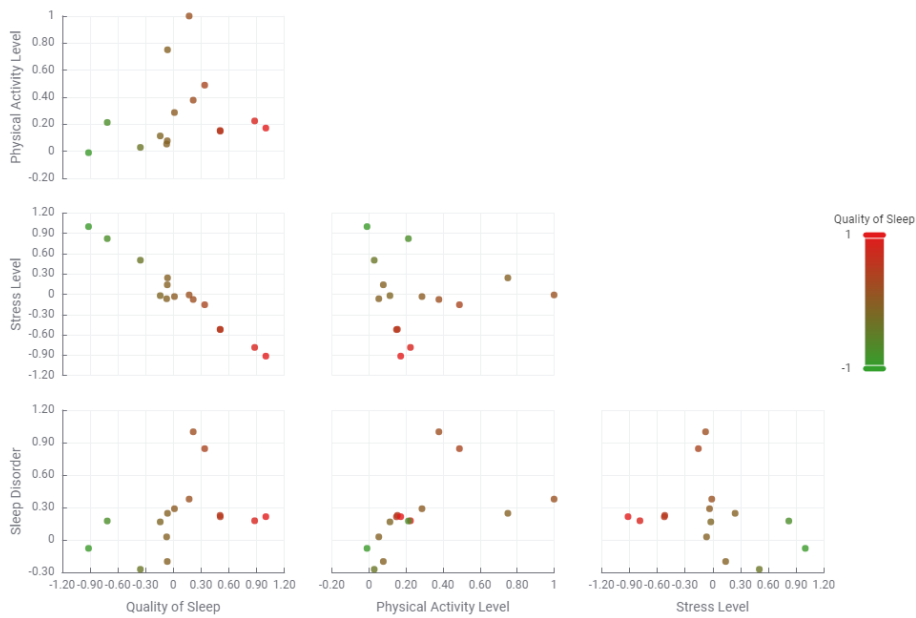
Son olarak aynı yöntemin incelendiği ROA için Şekil 8 ve Şekil 9'a benzer şekilde grafiğini aşağıda inceleyelim. Şekil 10’da görüleceği üzere veri noktalarının neredeyse yüzde 65 ve üstünde olan bir kısmının P-değeri 0 sonucunu vermektedir. Bu ise bu dağılımdan elde edilecek sonuçların gerçekte gözlenme olasılığının bu veri noktaları için sıfır olduğu anlamına gelmektedir. Grafikte veri olarak verilmemiş noktalardaki değerler interpolasyon işlemiyle tahmin edilmiştir. Burada incelenen Rank Korelasyon değerlerinin Line Plot modülü ile elde edilmesinin yanısıra Scatter Matrix modülüyle üç yöntem için de uyku sağlığını olumlu ya da olumsuz yönde etkilediği düşünülen

öznitelikler için dağılım matrisleri belirlenmiştir. Şekil 11’de ROA yöntemi ile elde edilen sonuçlardan biri olan Quality of Sleep ile Sleep disorder arasında pozitif korelasyon yönünde bir gözlem söz konusudur.



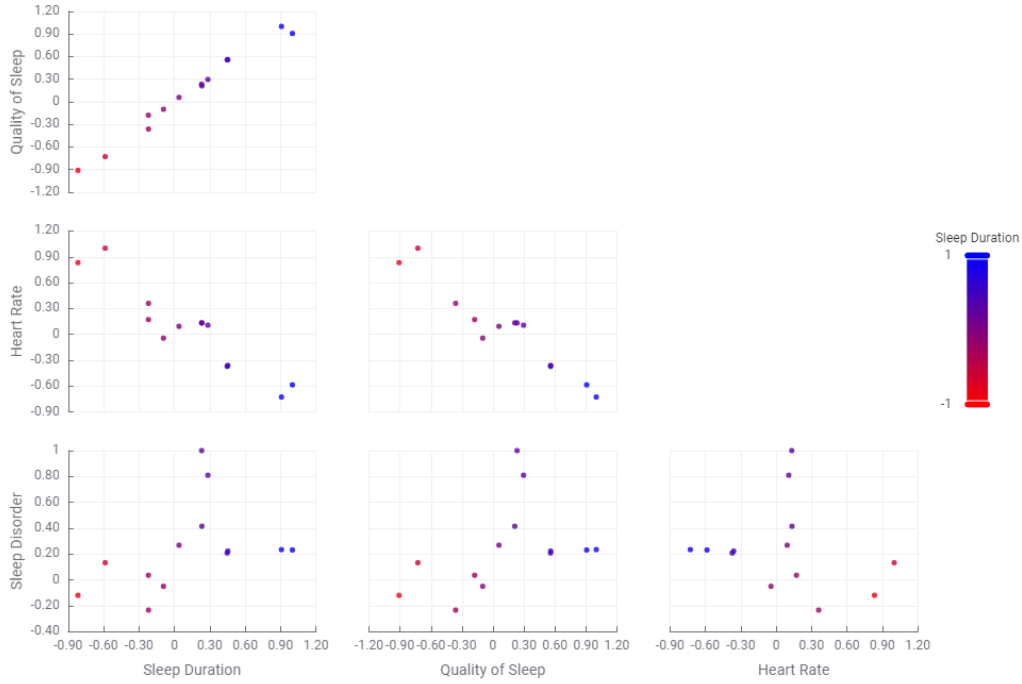
Şekil 10. ROA Rank Korelasyon P değeri ve Korelasyon Sonucu Karşılaştırılması

Ancak Physical Activity Level ile Quality of Sleep arasındaki dağılım diğerlerine göre daha dengelidir. Buna benzer bir dağılım Stress Level ile Physical Activity Level arasında da söz konusudur.



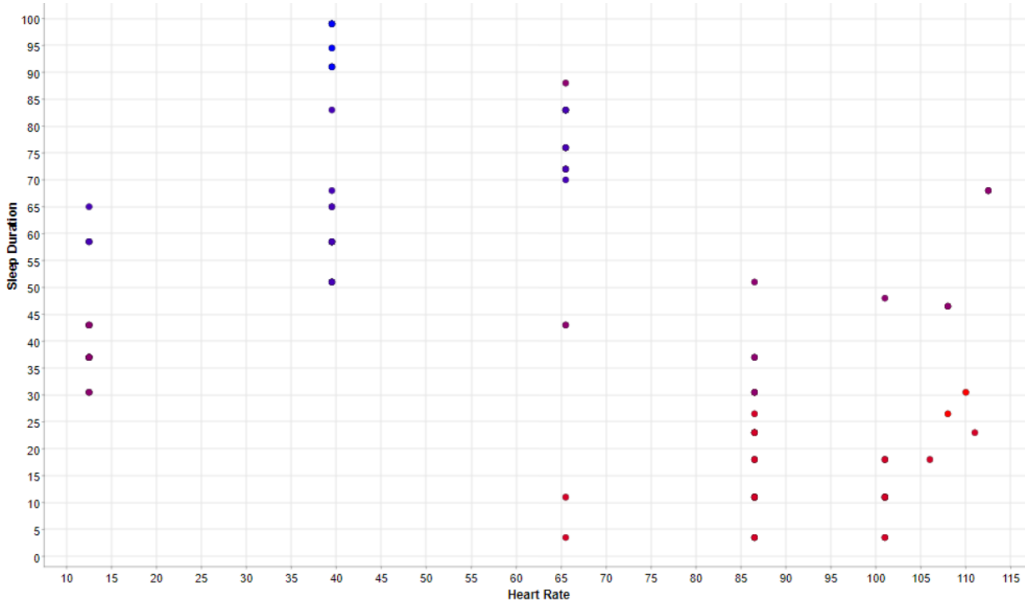
Şekil 11. ROA Rank Korelasyon Dağılım Matrisi

Elde edilen bilgilerin daha özel sonuçlarını yine KNIME ortamında yer alan Scatter Plot Matrix modülüyle elde edebiliriz. Bunu vermeden önce Şekil 12 ile verilen dağılım matrisini Naive Bayes yöntemi için yorumlanırsa, elde edilen sonuçlara göre Sleep Duration ile Heart Rate arasında negatif korelasyon gözlemlenmektedir. Ayrıca Quality of Sleep ile Sleep Duration arasında pozitif korelasyon görülmektedir.



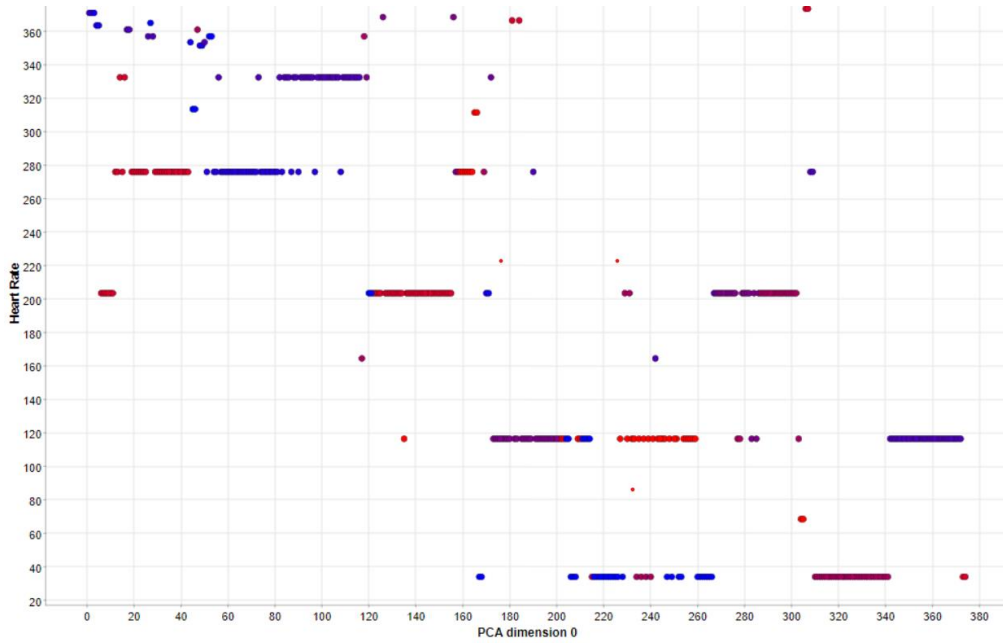
Şekil 12. Naive Bayes Rank Korelasyon Dağılım Matrisi

Bu üç yöntem ile elde edilen sonuçların Scatter Plot Matrix ile incelenmesi, veri setindeki öznitelikler arasındaki ilişkiyi korelasyon sonuçlarına göre yorumlamayı sağlamaktadır. Bunun yanında daha özel sonuçları öğrenmek KNIME ortamında kullanılan Scatter Plot modülü kullanılabilir. Şekil 13'te Naive Bayes yöntemi ile analiz edilen veri setinin rank korelasyon sonuçları kullanılarak Sleep Duration ile Quality of Sleep arasındaki ilişkinin dağılımını gösteren bir grafik KNIME ortamındaki Scatter Plot modülüyle elde edilmiştir.



Şekil 13. Sleep Duration - Quality of Sleep

Kullanılan diğer yöntemlerden biri olan Temel Bileşen Analizi ile veri setinde boyut azaltma yapıp istenen özneliğin yeni uzaydaki izdüşümüyle birlikte görselleştirmesi yapılabilmektedir. Bu durum Şekil 14 üzerinde görülmektedir.



Şekil 14. PCA Dimension – Occupation

5. TARTIŞMA VE SONUÇ

Bu çalışmanın bulguları, uyku sağlığı ve yaşam tarzı arasındaki ilişkileri keşfetmek için PCA, Naive Bayes ve ROA yöntemlerinin etkinliğini göstermektedir. PCA yöntemi, veri setindeki değişkenler arasındaki varyansı açıklayan ve korelasyon yapısını ortaya koyan temel bileşenleri bulmada başarılı olmuştur. Bu yöntem, veri setindeki gizli desenleri ve yapısını belirlemeyi, benzer uyku ve yaşam tarzı profillerine sahip grupları tanımlamayı ve değişkenler arasındaki ilişkileri görselleştirmeyi mümkün kılmıştır. PCA yöntemi, veri setindeki boyut sayısını azaltırken, verinin çoğu bilgisini korumuştur. Bu sayede, veri setinin daha kolay ve verimli bir şekilde analiz edilmesini sağlamıştır.

Naive Bayes ve ROA yöntemleri, veri setindeki değişkenler arasındaki ilişkileri daha fazla analiz etmek ve sınıflandırma yapmak için kullanılan yaygın algoritmalar arasındadır. Bu yöntemler, veri setindeki öznitelikler arasındaki ilişkileri anlamak ve uyku sağlığıyla ilişkilendirmek için kullanılabilir. Sonuçlar, bu yöntemlerin yüksek doğruluk oranlarına sahip olduğunu ve uyku sağlığı ile yaşam tarzı arasındaki ilişkileri tahmin etmede etkili olduğunu göstermiştir.

Bu çalışmanın en önemli katkısı, uyku verileri içeren veri setlerinde kullanılabilecek metodolojiler arasında en etkili olan yöntemi belirlemektir. Çalışmanın bazı sınırlamaları vardır. Örneğin, veri seti sadece 374 gözlem içermektedir, bu da sonuçların genelleştirilebilirliğini sınırlayabilir. Ayrıca, veri seti sadece bireylerin kendilerinin bildirdiği uyku ve yaşam tarzı değişkenlerini içermektedir, bu da ölçüm hatalarına veya yanlılıklara neden olabilir. Bu nedenle, daha büyük ve daha kapsamlı veri setleri kullanarak ve uyku sağlığını objektif olarak ölçen cihazlarla destekleyerek bu çalışmanın gelecekteki çalışmalarda geliştirilmesi önerilir.

Bunun yansısı, gelecekteki çalışmalarda, sadece daha geniş veri setlerini değil, aynı zamanda farklı demografik özelliklere sahip grupları, çeşitli coğrafyalardan bireyleri ve farklı yaşam tarzı faktörlerini kapsayacak şekilde veri seti çeşitlendirilerek analizler zenginleştirilebilir. Ayrıca, uyku sağlığına etki eden yeni değişkenleri keşfetmek, gelişmiş makine öğrenmesi tekniklerini kullanmak ve uyku ölçümlerini daha objektif yöntemlerle değerlendirmek gibi metodolojik iyileştirmeler üzerine de yoğunlaşılabilir. Bu yaklaşımlar, çalışmamızın genelleştirilebilirliğini ve uygulanabilirliğini artırarak, uyku sağlığı ve yaşam tarzı arasındaki ilişkileri daha kapsamlı bir şekilde anlamamıza olanak tanıyacaktır.

KAYNAKÇA

- Breiman, L. (2001). Random forests. *Machine learning*, 45, 5-32.
- Brink-Kjaer, A., Leary, E. B., Sun, H., Westover, M. B., Stone, K. L., Peppard, P. E., ..., Mignot, E. (2022). Age estimation from sleep studies using deep learning predicts life expectancy. *NPJ digital medicine*, 5(1), 103.
- Bro, R., Smilde, A. K., 2014. Principal component analysis. *Analytical methods*, 6(9), 2812-2831.
- Buysse, D. J. (2014). Sleep health: can we define it? Does it matter?. *Sleep*, 37(1), 9-17.
- Dietz, C., Rueden, C. T., Helfrich, S., Dobson, E. T., Hom, M., Eglinger, J., ..., Eliceiri, K. W. (2020). Integration of the ImageJ ecosystem in Knime analytics platform. *Frontiers in computer science*, 2, 8.
- Fillbrunn, A., Dietz, C., Pfeuffer, J., Rahn, R., Landrum, G. A., Berthold, M. R. (2017). KNIME for reproducible cross-domain analysis of life science data. *Journal of biotechnology*, 261, 149-156.
- Ghose, S. M., Dzierzewski, J. M., Dautovich, N. D. (2023). Sleep and self-efficacy: The role of domain specificity in predicting sleep health. *Sleep Health*, 9(2), 190-195.
- Hale, L., Troxel, W., Buysse, D. J. (2020). Sleep health: an opportunity for public health to address health equity. *Annual review of public health*, 41, 81-99.
- Hastie, T., Tibshirani, R., Friedman, J. H., Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758)*. New York: springer.
- Karamizadeh, S., Abdullah, S.M., Manaf, A.A., Zamani, M. and Hooman, A., 2013. An overview of principal component analysis. *Journal of Signal and Information Processing*, 4(3B), 173-175.
- Kaya, M., Özel, S. A. (2014). Açık kaynak kodlu veri madenciliği yazılımlarının karşılaştırılması. *Akademik Bilişim*, 1-8.
- Maraza-Quispe, B., Valderrama-Chauca, E. D., Cari-Mogrovejo, L. H., Apaza-Huanca, J. M., Sanchez-Ilabaca, J. (2022). A predictive model implemented in knime based on learning analytics for timely decision making in virtual learning environments. *International Journal of Information and Education Technology*, 12(2), 91-99.
- Pal, M. (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, 26(1), 217-222.
- Ricciardi, C., Valente, A. S., Edmund, K., Cantoni, V., Green, R., Fiorillo, A., ..., Cesarelli, M. (2020). Linear discriminant analysis and principal component analysis to predict coronary artery disease. *Health informatics journal*, 26(3), 2181-2192.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, 3(22), 41-46.
- Tharmalingam, L. 2023. Sleep Health and Lifestyle Dataset. <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>