

Araştırma Makalesi

Makine Öğrenimi Teknikleriyle Uyku Bozukluklarına Yönelik Derinlemesine Analiz ve Sınıflandırma

Yunus Emre GÜR * , Bilal SOLAK , Mesut TOĞAÇAR 

Gönderim: 29.04.2024

Kabul: 1.06. 2024

Fırat Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Elazığ, Türkiye; yegur@firat.edu.tr; bsolak@firat.edu.tr; mtogacar@firat.edu.tr

*Sorumlu yazar

Özet: Bu çalışma, uyku kalitesi ile yaşam tarzı faktörleri arasındaki bağlantıyı derinlemesine incelemektedir. Araştırma, cinsiyet, yaş ve meslek gibi demografik özellikler ile uyku süresi, kalitesi, fiziksel aktivite düzeyleri ve stres gibi yaşam tarzı değişkenlerinin uyku bozukluklarına etkisini makine öğrenimi teknikleri kullanarak analiz etmektedir. Çalışmada, Lojistik Regresyon, En yakın komşular, Naive Bayes, Rastgele Orman, AdaBoostClassifier ve Destek Vektör Makinesi (SVM) gibi çeşitli makine öğrenimi modelleri uygulanmıştır. Özellikle Rastgele Orman ve SVM modelleri, yüksek doğruluk oranları sayesinde uyku bozukluklarını belirlemede etkili oldukları gözlemlenmiştir. Ayrıca, meslek ve stres düzeyleri ile uyku bozuklukları arasındaki ilişkiler üzerine detaylı analizler yapılarak, uyku sağlığının iyileştirilmesi için öneriler sunulmuştur.

Anahtar Kelimeler: Uyku sağlığı; uyku bozuklukları; makine öğrenimi; Random Forest; destek vektör makinesi

In-depth Analysis and Classification of Sleep Disorders with Machine Learning Techniques

Abstract: This study investigates the link between sleep quality and lifestyle factors in depth. The study analyses the effect of demographic characteristics such as gender, age, and occupation and lifestyle variables such as sleep duration, quality, physical activity levels, and stress on sleep disorders using machine learning techniques. In the study, various machine learning models such as logistic regression, nearest neighbors, naive bayes, random forest, adaBoost classifier, and support vector machine (SVM) were applied. In particular, Random Forest and SVM models were found to be effective in identifying sleep disorders due to their high accuracy rates. In addition, detailed analyses on the relationships between occupation, stress levels, and sleep disorders were performed, and recommendations were presented to improve sleep health.

Keywords: Sleep health; sleep disorders; machine learning; Random Forest; support vector machine

1. Giriş

Uyku, insan hayatının temel bir yönüdür ve kaliteli bir yaşam sürdürebilmek için yeterli ve rahat bir uyku şarttır. Yapılan araştırmalar, uyku kalitesinin ve süresinin, bireylerin fiziksel, zihinsel ve duygusal sağlığı üzerinde derin bir etkisi olduğunu göstermektedir [1-3]. Ancak, günümüz toplumunda birçok birey, çeşitli nedenlerle uyku bozuklukları yaşamakta ve bu durum genel sağlık üzerinde olumsuz etkilere sahip olmaktadır. Uyku bozukluklarının arka planındaki faktörleri anlamak ve etkili müdahaleler geliştirmek, bu nedenle, kamu sağlığı için büyük bir öncelik haline gelmiştir [4].

Bu araştırma, uyku sağlığı ve yaşam tarzı arasındaki karmaşık ilişkileri detaylı bir şekilde incelemeyi amaçlamaktadır. Özellikle, cinsiyet, yaş, meslek, uyku süresi, uyku kalitesi, fiziksel aktivite seviyesi, stres seviyesi, BMI kategorisi, kan basıncı, kalp atış hızı ve günlük adımlar gibi faktörlerin uyku bozuklukları üzerindeki etkilerine odaklanılmaktadır. Ayrıca, demografik faktörlerin - cinsiyet, yaş, meslek - uyku bozuklukları üzerindeki etkileri de araştırmanın önemli bir parçasını oluşturmaktadır. Bu bağlamda, araştırmamız makine öğrenmesi algoritmaları kullanarak uyku bozukluklarını sınıflandırmakta ve bu bozuklukların cinsiyet, yaş, meslek gibi çeşitli demografik ve yaşam tarzı faktörleriyle ilişkisini ortaya koymaktadır.

Makine öğrenimi algoritmaları kullanılarak uyku bozukluğu sınıflandırması son araştırmalarda önemli ilerlemeler kaydetmiştir. Çeşitli çalışmalar, otomatik uyku bozukluğu sınıflandırması için makine öğrenimi tekniklerinin uygulanmasını araştırmıştır. Örneğin, Urtnasan vd. [5] uykusuzluk, periyodik bacak hareketi, REM uykusu davranış bozukluğu ve gece frontal lob epilepsisi gibi başlıca uyku bozukluklarının otomatik sınıflandırılması için yapay zekâ destekli bir algoritma olan uyku bozukluğu ağı (SDN) geliştirmiştir. Bu, yapay zekânın uyku bozukluklarını doğru bir şekilde kategorize etme potansiyelini göstermektedir. Ayrıca Kwon vd. [6], hassas uyku evresi sınıflandırması için evrişimli derin sinir ağlarını kullanarak makine öğrenimi algoritmalarının sınıflandırma doğruluğunu artırmadaki etkinliğini vurgulamıştır. Ek olarak, Rakhonde vd. [7], uyku evresi sınıflandırması için bir makine öğrenimi algoritması olarak Stokastik Gradyan İnişi (SGD) kullanmış ve bu alandaki farklı makine öğrenimi yaklaşımlarının çok yönlülüğünü göstermiştir. Bununla birlikte, Pradeepkumar vd. [8], derin öğrenme tabanlı algoritmaların uyku evresi sınıflandırmasında geleneksel makine öğrenimi yöntemlerinden daha iyi performans gösterdiğini göstererek, gelişmiş sonuçlar için gelişmiş tekniklerden yararlanmanın önemini vurgulamıştır. Sundararajan vd. [9], 2021 yılında, bileğe takılan ivmeölçer verileri ve rastgele ormanlar kullanarak uyku sınıflandırması üzerine bir çalışma yürütmüştür. Delimayanti vd. [10], beyin dalgalarını ve EEG sinyallerinden çıkarılan yüksek boyutlu FFT özelliklerini kullanarak uyku evrelerinin sınıflandırılmasını araştırmıştır. Yulita vd. otomatik uyku evresi sınıflandırması için hızlı bir konvolüsyonel yöntem geliştirerek uyku bozukluğu sınıflandırmasında makine öğrenimi algoritmaları alanına katkıda bulunmuştur. Yulita vd. [11], otomatik uyku evresi sınıflandırması için hızlı bir konvolüsyonel yöntem geliştirerek uyku bozukluğu sınıflandırmasında makine öğrenimi algoritmaları alanına katkıda bulunmuştur. Cho vd. [12], uyku skorlama uygulamaları için çeşitli makine öğrenimi yöntemleri uygulayarak farklı uyku aşamalarını ayırt etmede rastgele orman, torbalama ve kstar sınıflandırıcı gibi algoritmaların çok yönlülüğünü sergilemiştir.

Bu gelişmeler, uyku bozukluğu sınıflandırması için daha sofistike algoritmalara doğru bir kaymaya işaret etmektedir. Dolayısıyla, makine öğrenimi algoritmalarıyla uyku bozukluğu sınıflandırmasına ilişkin literatür taraması, uyku evresi sınıflandırmasının doğruluğunu ve verimliliğini artırmak için yapay zeka ve derin öğrenme tekniklerinden yararlanmaya yönelik bir eğilim olduğunu ortaya koymaktadır.

Bu çalışmanın bir diğer önemli yönü, meslek türleri ve stres seviyeleri ile uyku bozuklukları arasındaki ilişkiyi derinlemesine analiz etmektir. İstatistiksel analiz yöntemleri kullanılarak, cinsiyet ve meslek türlerine göre uyku bozukluklarının ve stres seviyelerinin analizi yapılmaktadır. Bu analizlerin amacı, belirli meslek gruplarında uykusuzluk problemi yaşama olasılığının daha yüksek olduğunu belirlemek ve meslek ile stres seviyeleri arasındaki ilişkiyi ortaya koymaktadır. Bu değerlendirmeler, uyku bozukluklarına yönelik önleme ve müdahale stratejilerinin geliştirilmesinde kritik öneme sahiptir. Araştırma, uyku sağlığına yönelik farkındalığı artırmayı, uyku bozukluklarının önlenmesine yönelik etkili stratejiler geliştirmeyi ve böylece bireylerin yaşam kalitesini iyileştirmeyi hedeflemektedir. Bu çalışma, uyku sağlığını iyileştirmeye yönelik müdahalelerin geliştirilmesine katkıda bulunarak, bu alandaki bilgi boşluklarını doldurmayı ve toplum sağlığını olumlu yönde etkilemeyi amaçlamaktadır.

Çalışmanın ilerleyen bölümlerinde, araştırmanın metodolojisi açıklanmış, ardından analize ilişkin bulgulara yer verilmiş ve sonuç ve tartışma bölümünde araştırmanın temel bulguları, bu bulguların mevcut literatürle karşılaştırılması, çalışmanın sınırlılıkları ve gelecek çalışmalar için önerilerde bulunulmuştur.

2. Materyal ve Metod

Bu araştırma, uyku sağlığı ve yaşam tarzı arasındaki ilişkileri kapsamlı bir şekilde inceleyen, makine öğrenmesi ve istatistiksel analiz yöntemleri kullanılarak gerçekleştirilmiş bir çalışmadır. Araştırmanın temel odak noktası, uyku bozuklukları, stres seviyeleri ve meslek türleri arasındaki dinamik ilişkilerin derinlemesine analizi ve uyku bozukluklarının makine öğrenmesi yöntemleri ile sınıflandırılmasıdır.

2.1. Veri Seti

Araştırmada kullanılan "Uyku Sağlığı ve Yaşam Tarzı" veri seti, Kaggle platformundan elde edilmiştir [13]. Bu veri seti, 374 katılımcının bilgilerini içermekte olup, cinsiyet, yaş, meslek gibi demografik bilgiler ile uyku alışkanlıkları, fiziksel aktivite düzeyleri, stres seviyeleri ve daha birçok yaşam tarzı faktörüne dair geniş bir yelpazede veri sunmaktadır. Veri toplama süreci, katılımcıların gönüllü katılımı esas alınarak ve katılımcıların gizliliğini koruyacak şekilde yürütülmüştür. Kullanılan veri setinin içerdiği değişkenler ve bu değişkenlerin açıklamaları Tablo 1'de gösterilmektedir.

Tablo 1. Veri Setinde Kullanılan Değişkenler ve Açıklamaları

Değişkenler	Açıklama	Tür Sayısı
Kişi Sayısı	Çalışmaya katılan katılımcılar	374
Cinsiyet	Erkek/Kadın	2
Yaş	Kişinin yıl cinsinden yaşı	31
Meslek	Meslek Türleri	11
Uyku süresi	(saat): Kişinin günde kaç saat uyuduğu	27
Uyku kalitesi	(ölçek: 1-10): Uyku kalitesinin 1 ila 10 arasında değişen öznel bir değerlendirmesi.	6
Fiziksel aktivite seviyesi	(dakika/gün): Kişinin günlük fiziksel aktivite yaptığı dakika sayısı.	16

Stres seviyesi	(ölçek: 1-10): Kişinin yaşadığı stres düzeyinin 1 ila 10 arasında değişen öznel bir değerlendirmesi.	6
Vücut Kitle İndeksi (BMI) Kategorisi	Kişinin BMI kategorisi (örneğin, Zayıf, Normal, Aşırı Kilolu).	4
Kan Basıncı	Kişinin kan basıncı ölçümü, diyastolik basınç üzerinden sistolik basınç olarak gösterilir.	25
Kalp atış hızı	Dakikadaki atım sayısı olarak kişinin dinlenme kalp atış hızı.	19
Günlük adımlar	Kişinin günde attığı adım sayısı.	20
Uyku Bozukluğu (Bağımlı Değişken)	Kişide uyku bozukluğu olup olmaması (Yok, Uykusuzluk, Uyku Apnesi).	3

Bu araştırma, uyku sağlığı ve yaşam tarzı arasındaki ilişkileri kapsamlı bir şekilde incelemeyi amaçlamakta olup, çeşitli veri keşfi ve gruplama teknikleri, veri ön işleme, analiz ve sınıflandırma model değerlendirme işlemleri Python programlama dili kullanılarak gerçekleştirilmiştir. Python, bilimsel hesaplamalar ve veri analizi için kapsamlı kütüphanelere (pandas, NumPy, scikit-learn, matplotlib) sahip olduğu için tercih edilmiştir.

2.2. Veri Keşfi ve Gruplama Teknikleri

Bu çalışmada, uyku bozuklukları, meslek türleri, stres seviyeleri ve cinsiyet gibi değişkenler arasındaki ilişkileri incelemek için çeşitli veri keşfi ve gruplama teknikleri kullanılmıştır. Analiz süreci, Python programlama dili ve bu dilin güçlü veri işleme kütüphaneleri olan Pandas, NumPy, Seaborn ve Matplotlib kullanılarak gerçekleştirilmiştir. “.value_counts()” ve “.groupby()” fonksiyonları kullanılarak, uyku bozuklukları, meslek türleri, stres seviyeleri ve cinsiyet gibi kategorik değişkenlerin dağılımları ve birbirleriyle olan ilişkileri analiz edilmiştir. Tablo 2’de bu kategorik değişkenlerin dağılımlarına ilişkin veri tipleri sıralanmıştır. Bu analizler, veri setindeki farklı grupların uyku bozuklukları üzerindeki potansiyel etkilerini anlamak için yapılmıştır.

Tablo 2. *Veri setinde kullanılan kategorik değişkenlerin veri tipleri*

Sütun	Null Olmayan Sayısı	Veri Tipi
Kişi Kimliği	374	int64
Cinsiyet	374	object
Yaş	374	int64
Meslek	374	object
Uyku Süresi	374	float64
Uyku Kalitesi	374	int64
Fiziksel Aktivite Seviyesi	374	int64
Stres Seviyesi	374	int64
BMI Kategorisi	374	object
Kan Basıncı	374	object
Kalp Atış Hızı	374	int64
Günlük Adım Sayısı	374	int64
Uyku Bozukluğu	374	object

2.3. Veri Ön İşleme

İlk adımda, uyku sağlığı ve yaşam tarzı veri setinden elde edilen kategorik değişkenler (cinsiyet, yaş, meslek vb.), Python'un pandas kütüphanesi kullanılarak işlenmiştir. Bu dönüşüm, sklearn kütüphanesinin LabelEncoder fonksiyonu kullanılarak gerçekleştirilmiştir. Kişisel kimlik numarası gibi analiz için gerekli olmayan sütunlar veri setinden çıkarılmıştır. Ek olarak, kan basıncı verileri iki kategoriye dönüştürülmüştür: normal ve anormal. Önceden belirlenen değerler (örneğin, '120/80', '126/83', '125/80' vb.) normal kan basıncı olarak kabul edilmiş ve bu değerlere sahip olan gözlemler "0" ile etiketlenmiştir. Bu değerlerin dışındaki tüm kan basıncı değerleri ise anormal olarak kabul edilerek "1" ile etiketlenmiştir. Bu işlem, kan basıncı verilerinin analizde daha etkin bir şekilde kullanılmasını sağlamaktadır. Bu işlemler sonrasında, özellikler (X) ve hedef değişken (y) olarak iki ayrı veri seti oluşturulmuştur.

2.4. Model Eğitimi ve Testi

Model eğitimi ve test işlemleri için scikit-learn kütüphanesinden faydalanılmıştır. Veri seti, "train_test_split" fonksiyonu ile eğitim ve test setlerine ayrılmıştır. Eğitim seti (X), modelin öğrenmesi için kullanılırken, test seti (y) modelin genelleştirme performansını değerlendirmek için kullanılmıştır. Oluşturulan X ve y veri setleri, eğitim ve test setleri olmak üzere ikiye ayrılmıştır. Bu ayrım, veri setinin %80'i eğitim, %20'si test olacak şekilde ayarlanmıştır. Eğitim seti, modelin öğrenmesi için; test seti ise modelin performansının değerlendirilmesi için kullanılmıştır.

2.5. Model Değerlendirme

Lazy Predict, sınıflandırma görevlerinde model seçimi ve değerlendirme sürecini otomatikleştiren bir makine öğrenimi kütüphanesidir. Kapsamlı manuel yapılandırmaya gerek kalmadan hızlı prototip oluşturma ve birden fazla modeli karşılaştırma için özellikle kullanışlıdır. Sınıflandırmada Lazy Predict, eğitim verilerine birden fazla model uydurarak ve doğruluk, kesinlik, geri çağırma ve F1 puanı gibi performans ölçümlerinin bir özetini sağlayarak kullanıcıların veri kümeleri için en uygun modeli belirlemelerini kolaylaştırır [14]. Lazy Predict, modelin eğitim aşamasında eğitim verilerinden aktif olarak ayırt edici bir işlev öğrenmediği tembel öğrenme ilkesine göre çalışır. Bunun yerine, bir tahmin gerekene kadar öğrenme sürecini erteler, bu noktada en yakın komşuları hesaplar veya tahmin yapmak için depolanmış örnekleri kullanır. Bu yaklaşım, eğitim sırasında tek bir küresel model oluşturan ve büyük

veri kümeleri için hesaplama açısından daha az verimli olabilen istekli öğrenme yöntemleriyle tezat oluşturmaktadır [15]. Bu çalışmada, LazyPredict kütüphanesi aracılığı ile LazyClassifier kullanılarak Lojistik Regresyon, KNN, Naive Bayes, Random Forest, AdaBoostClassifier, DummyClassifier ve SVM makine öğrenmesi modellerinin performansları karşılaştırılmış ve en iyi sonucu veren modeller genel doğruluk oranları ile belirlenmiştir.

Makine öğrenimi modellerinin matematiksel formülasyonları ve algoritmaların nasıl çalıştığına dair detaylı açıklamaların eksikliği, okuyucuların kullanılan modellerin iç işleyişini tam olarak anlamasını zorlaştırmaktadır. Bu nedenle, her bir makine öğrenimi modeli için matematiksel formülasyonlar ve algoritma adımları detaylandırılmalıdır. Lojistik regresyonun hipotez fonksiyonu sigmoid fonksiyonudur ve şu şekilde tanımlanır:

$$\mathbf{h}_{\theta}(\mathbf{x}) = \frac{1}{1+e^{-\theta^T \mathbf{x}}} \quad (2.1)$$

Maliyet fonksiyonu (cost function) şu şekilde ifade edilir:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [\mathbf{y}^{(i)} \log(\mathbf{h}_{\theta}(\mathbf{x}^{(i)})) + (1 - \mathbf{y}^{(i)}) \log(1 - \mathbf{h}_{\theta}(\mathbf{x}^{(i)}))] \quad (2.2)$$

Burada, m toplam örnek sayısını, $\mathbf{y}^{(i)}$ ise i 'inci örneğin gerçek sınıf etiketini temsil eder.

Bununla birlikte, SVM'nin amacı, iki sınıfı en geniş marjinle ayıran bir hiper düzlem bulmaktır. Hiper düzlem denklemi şu şekilde ifade edilir:

$$\mathbf{f}(\mathbf{x}) = \mathbf{W}^T \mathbf{x} + \mathbf{b} \quad (2.3)$$

Burada \mathbf{W} ağırlık vektörünü, \mathbf{b} ise bias terimini temsil etmektedir. Ek olarak, SVM optimizasyon problemi şu şekilde tanımlanmaktadır:

$$\min_{\mathbf{w}, \mathbf{b}} \frac{1}{2} \|\mathbf{w}\|^2 \quad \text{subject to } \mathbf{y}^{(i)} (\mathbf{w}^T \mathbf{x}^{(i)} + \mathbf{b}) \geq 1 \quad \forall i \quad (2.4)$$

Bu problem Lagrange çarpanları kullanılarak dual formülasyona dönüştürülür:

$$\max_{\mathbf{a}} \sum_{i=1}^m \mathbf{a}_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \mathbf{a}_i \mathbf{a}_j \mathbf{y}^{(i)} \mathbf{y}^{(j)} (\mathbf{x}^{(i)})^T \mathbf{x}^{(j)} \quad (2.5)$$

Burada \mathbf{a}_i Lagrange çarpanlarıdır.

Öte yandan, KNN algoritması, sınıflandırma veya regresyon için en yakın \mathbf{k} komşuyu bulur. Herhangi bir veri noktası \mathbf{x} için, sınıflandırma aşağıdaki gibi yapılır:

$$\mathbf{y} = \mathbf{mode}(\{\mathbf{y}^{(i)} | \mathbf{x}^{(i)} \in \mathbf{N}_k(\mathbf{x})\}) \quad (2.6)$$

Burada $\mathbf{N}_k(\mathbf{x})$, \mathbf{x} noktasının en yakın \mathbf{k} komşularını temsil eder ve mode ise en yaygın sınıf etiketini belirtir. Bununla birlikte, Naive Bayes sınıflandırıcısı, Bayes teoremi üzerine kuruludur ve şu şekilde ifade edilmektedir:

$$\mathbf{P}(\mathbf{y}|\mathbf{x}) = \frac{\mathbf{P}(\mathbf{x}|\mathbf{y})\mathbf{P}(\mathbf{y})}{\mathbf{P}(\mathbf{x})} \quad (2.7)$$

Burada, $\mathbf{P}(\mathbf{y}|\mathbf{x})$ \mathbf{x} gözlemi verildiğinde \mathbf{y} sınıfının olasılığıdır. Gaussian Naive Bayes için, her özellik \mathbf{x}_i normal dağılıma sahiptir ve şu şekilde hesaplanır:

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma_y^2}\right) \quad (2.8)$$

Son olarak, Random Forest, birden çok karar ağacının ensemble (birlikte) kullanılması ile oluşturulan bir öğrenme yöntemidir. Her bir ağaç bağımsız olarak eğitilir ve sınıflandırma için çoğunluk oylaması ile karar verilir:

$$\hat{y} = \mathbf{mode}(\{h_1(x), h_2(x), \dots, h_B(x)\}) \quad (2.9)$$

Burada, $h_i(x)$, i 'inci karar ağacının tahminini temsil eder ve B toplam ağaç sayısıdır. Bu tür matematiksel ifadeler ve algoritma adımları, okuyucuların modellerin teorik temellerini ve uygulama süreçlerini daha iyi anlamalarını sağlayacaktır.

Bununla birlikte, modellerin sınıflandırma performansları aşağıdaki metriğe formülasyonları kullanılarak değerlendirilmiştir. Burada, TP: True Positives (Doğru Pozitifler), TN: True Negatives (Doğru Negatifler), FP: False Positives (Yanlış Pozitifler) ve FN: False Negatives (Yanlış Negatifler) değerlerini temsil etmektedir.

Doğruluk, tüm doğru tahminlerin toplam örnek sayısına oranıdır.

$$\mathbf{Doğruluk (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.10)$$

Kesinlik, pozitif olarak tahmin edilen değerlerin gerçekten pozitif olma oranıdır.

$$\mathbf{Kesinlik (Precision)} = \frac{TP}{TP+FP} \quad (2.11)$$

Geri çağırma, gerçek pozitif değerlerin ne kadarının doğru tahmin edildiğini gösterir.

$$\mathbf{Geri Çağırma (Recall)} = \frac{TP}{TP+FN} \quad (2.12)$$

F1 skoru, kesinlik ve geri çağırmanın harmonik ortalamasıdır ve dengesiz sınıflarda performansı değerlendirmek için kullanılır.

$$\mathbf{F1 Skoru (F1 Score)} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.13)$$

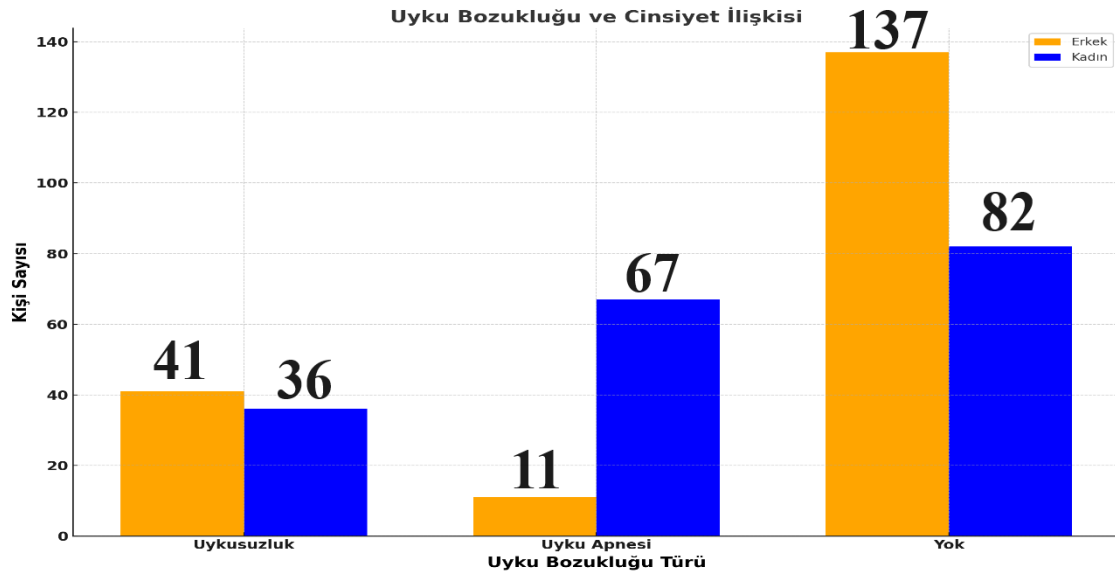
Bu formülasyonlar, makine öğrenimi modellerinin performansını değerlendirmek için yaygın olarak kullanılan metrikleri matematiksel olarak tanımlamaktadır.

2.6. Aykırı Değerlerin Ele Alınması

Makine öğrenmesi algoritmalarının doğruluğunu arttırmak adına, aykırı değerler ele alınmıştır. Aykırı değerlerin ele alınması için, yaş, uyku süresi, fiziksel aktivite seviyesi, stres seviyesi, kalp atış hızı ve günlük adım sayısı gibi sayısal değişkenler için IQR yöntemi [16-17] kullanılmıştır. Standartlaştırma için scikit-learn'ün StandardScaler fonksiyonu kullanılmıştır. Aykırı değerler temizlendikten sonra, kalan veri üzerinde standartlaştırma işlemi gerçekleştirilmiş ve ilk aşamada en iyi sonucu veren modeller yeniden eğitilmiştir. Modellerin performansı, çapraz doğrulama yöntemi ile değerlendirilmiş ve sınıflandırma raporu oluşturulmuştur.

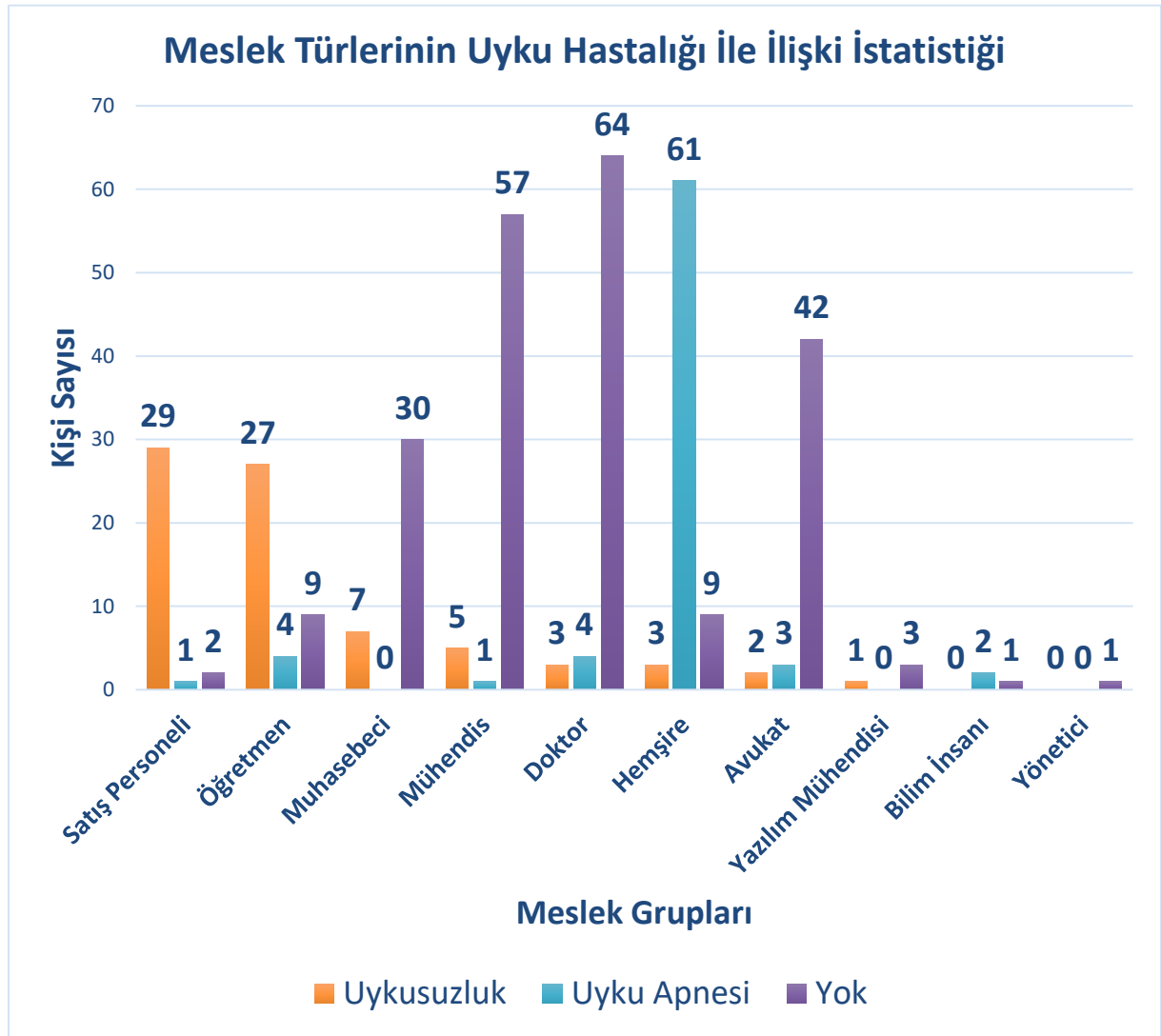
3. Bulgular ve Tartışma

Bu bölümde, analiz yöntemleri kullanılarak elde edilen bulgular sunulmaktadır. Analizde, uyku bozuklukları, meslek türleri, stres seviyeleri ve cinsiyet arasındaki ilişkiler üzerinde odaklanmıştır. Elde edilen veriler, uyku bozukluğu prevalansının demografik faktörlerle nasıl ilişkili olduğunu ortaya koymaktadır. İlk olarak, Şekil 1’deki grafikte uyku bozukluğu ve cinsiyet arasındaki ilişki gösterilmektedir. Bu sütun grafiği, çeşitli uyku bozukluklarının erkekler ve kadınlar arasında nasıl dağıldığını göstermektedir. Uykusuzluk durumuna bakıldığında, erkeklerde 41 kişi ve kadınlarda 36 kişi bu sorunu yaşamaktadır. Uyku apnesi söz konusu olduğunda ise cinsiyetlere göre dağılım önemli derecede farklılık göstermekte; 11 erkek uyku apnesi sorunu yaşarken, bu sayı kadınlarda 67’ye çıkmaktadır. “Yok” kategorisi ise uyku bozukluğu yaşamayan bireyleri ifade etmekte ve burada erkekler açık bir çoğunluktadır; 137 erkeğe karşın 82 kadın bu kategoriye girmektedir. Grafiğin genel gösterimi, uyku apnesinin kadınlarda daha yaygın olduğunu, ancak diğer kategorilerde ve toplamda uyku bozukluğu yaşamayanların sayısının erkeklerde daha fazla olduğunu göstermektedir.



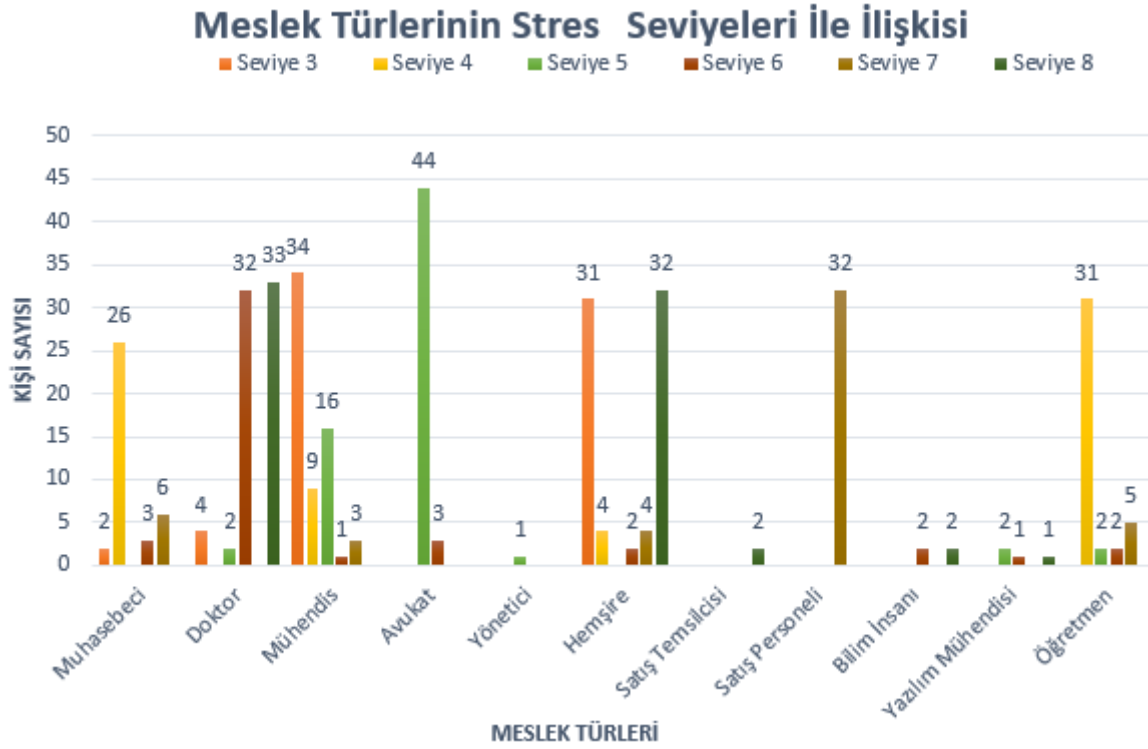
Şekil 1. Uyku bozukluğu ve cinsiyet arasındaki ilişki

Bununla birlikte Şekil 2’deki sütun grafiğinde, çeşitli meslek gruplarının üç farklı uyku bozukluğu durumuyla olan ilişkisi gösterilmektedir. Kırmızı sütunlar uykusuzluğu, mavi sütunlar uyku apnesini ve sarı sütunlar ise hiçbir uyku bozukluğu belirtisi olmayan kişileri temsil etmektedir. Satış personeline uykusuzluk en fazla olan meslek grubudur, 29 kişi ile bu durumu yaşayanlar arasında en yüksek sayıya sahiptir. Öğretmenlerde de uykusuzluk yaygındır; bu grupta 27 kişi etkilenmiştir. Uyku apnesi açısından bakıldığında, hemşireler toplamda 61 kişiyle bu uyku bozukluğunu en çok yaşayan meslek grubudur. Diğer meslek gruplarından öğretmen, doktor ve avukatlar da dahil olmak üzere uyku apnesi yaşayan kişiler var, ancak sayıları çok daha düşüktür. Grafiğin sarı sütunları, hiçbir uyku bozukluğu yaşamayan meslek gruplarını temsil etmekte ve doktorlar bu kategoride öne çıkmaktadır. Toplam 64 doktor uyku bozukluğu rapor etmemiştir. Mühendisler de büyük oranda uyku bozukluğu yaşamamakta, 57 kişiyle ikinci en yüksek sayıya sahiptirler. Avukatlar ise 42 kişi ile üçüncü sırada yer almaktadır. Genel olarak, bu grafik, farklı meslek grupları arasında uyku bozukluklarının dağılımının ne şekilde değişkenlik gösterdiğini görsel bir biçimde sergilemektedir. Bu tür veriler, mesleki stres seviyeleri veya çalışma saatleri gibi faktörlerin uyku kalitesi üzerindeki olası etkilerini anlamada yararlıdır.



Şekil 2. Meslek türleri ve uyku hastalığı ilişkisi

Buna ek olarak, satış personeli ve öğretmenler, uykusuzluk yaşayanların en büyük iki grubunu oluşturmakta, bu da bu meslek gruplarının potansiyel olarak yüksek stres düzeylerine veya düzensiz çalışma saatlerine sahip olabileceğini düşündürmektedir. Öte yandan, uyku apnesi özellikle hemşireler arasında yaygın görülmektedir. Bu durum, vardiya çalışmasının sık olduğu, dolayısıyla stresli bir meslek grubunda olmanın uyku düzeni üzerinde olumsuz bir etkisi olabileceğine işaret etmektedir. Bunun yanı sıra, doktorlar ve mühendisler, uyku bozukluğu rapor etmeyen en büyük iki gruptur. Bu, iş saatlerinin düzenli olabileceğini veya bu meslek gruplarının uyku sağlığını koruma konusunda daha fazla kaynağa ve bilince sahip olabileceğini göstermektedir. Dolayısıyla, uyku bozukluklarının prevalansı meslekler arası büyük farklılıklar göstermekte, bu da iş ortamının, iş yükünün ve stresin uyku üzerinde önemli bir etkisi olabileceğini göstermektedir. Uyku sağlığı genel sağlık ve iş performansı için hayati önem taşımaktadır. Uykusuzluk veya uyku apnesi gibi uyku bozukluklarının yüksek prevalansı, iş yerlerinin çalışanların uyku sağlığını desteklemeye yönelik politikalar geliştirmesinin önemini vurgulamaktadır. Bu grafik ayrıca, uyku bozukluklarının daha detaylı analizi için bir başlangıç noktası sağlamaktadır; örneğin, neden bazı meslek gruplarında bu sorunlar daha yaygın görülür ve iş yerinde bu sorunların önlenmesi veya azaltılması için neler yapılabilir gibi sorulara cevap aranabilir. Şekil 3'teki grafik, meslek türlerinin Stress seviyeleri ile ilişkisini göstermektedir. Bu grafik, farklı meslek gruplarının stres seviyelerine göre kişi sayılarını renk kodlu sütunlar halinde göstermektedir. Her renk, belirli bir stres seviyesini temsil ederken, sütunların yüksekliği her stres seviyesindeki kişi sayısını göstermektedir.



Şekil 3. Meslek türleri ve stres seviyesi ilişkisi

İlgili grafik incelendiğinde, ilk olarak, muhasebecilerdeki stres dağılımı, 4 stres seviyesinin en yüksek oranda olduğunu gösterirken, daha yüksek stres seviyeleri önemli ölçüde azdır. Bu, muhasebecilerin günlük işlerinde karşılaştığı zorlukların yönetilebilir olduğunu, ancak belirli dönemlerde (örneğin vergi dönemleri gibi) stresin artabileceğini gösterebilir. Doktorlar ve hemşirelerde görülen yüksek stres seviyeleri, sağlık sektörünün yoğun doğası ve bu mesleklerin karşılaştığı yüksek baskıyı yansıtır. Özellikle, 8 stres seviyesindeki yüksek oranlar, acil durumlar, uzun çalışma saatleri, yüksek sorumluluk düzeyleri ve hatta yaşam ve ölüm kararları gibi faktörlerle ilişkili olabilir. Mühendislerde stres seviyesinin daha düşük olduğunu gösteren 3 seviyesindeki yüksek sayılar, bu meslek dalının problem çözme, planlama ve öngörülebilir iş yükleri gibi stresi azaltan faktörleri içerebileceğini düşündürülebilir. Avukatların 5 stres seviyesinde yoğunlaşması, hukuk pratiğinin rekabetçi doğası ve karmaşık davalara hazırlık sürecindeki stresleri yansıtır olabilir. Öğretmenlerin ve satış personelinin yüksek stres seviyelerinde yoğunlaşması, bu mesleklerin genellikle yüksek insan etkileşimi, performans baskısı ve muhtemelen kaynak yetersizliklerinin birleşiminden kaynaklanan stresi göstermektedir. Bu verileri iş stresi teorileri ve modelleriyle ilişkilendirmek mümkündür. Örneğin, Karasek'in İş Talepleri-Kaynakları modeli [18] veya Siegrist'in Ödül Eşitsizliği modeli [19] gibi teorik çerçeveler, işin talepleri ve bireyin bu talepleri karşılamak için sahip olduğu kaynaklar arasındaki dengesizliğin stres yaratma potansiyeline işaret etmektedir [20-23]. Bu tür analizler, iş yerlerinin stresle başa çıkma stratejilerini geliştirmede, örneğin stres azaltıcı programların tasarımında, eğitim ve kaynak tahsisinde faydalı olabilir. Aynı zamanda, iş sağlığı ve güvenliği politikalarının şekillendirilmesinde ve meslek gruplarına yönelik özel destek sistemlerinin oluşturulmasında da kullanılabilir. Elde edilen bu bulgular, uyku bozukluğu prevalansının demografik faktörlerle nasıl ilişkili olduğunu ortaya koymaktadır. Özellikle, hangi meslek gruplarında uyku bozukluğunun daha yaygın olduğu, stres seviyesi ve uyku kalitesi arasındaki ilişki gibi detaylar burada sunulmuştur. Bu bilgilere ek olarak, bu araştırmada, uyku sağlığı ve yaşam tarzı veri seti üzerinde çeşitli makine öğrenmesi modelleri kullanılarak sınıflandırma analizleri gerçekleştirilmiştir. Analizin ilk aşamasında, LazyPredict kütüphanesi aracılığıyla LazyClassifier kullanılarak Lojistik Regresyon, KNN (k-Nearest Neighbors), Naive Bayes, Random Forest, Ada-BoostClassifier, DummyClassifier ve SVM (Support Vector Machine) makine öğrenmesi modellerinin

performansları karşılaştırılmış ve en iyi sonucu veren modeller genel doğruluk oranları ile ölçülmüştür. Bu modellere ilişkin sonuçlar, Tablo 3'te gösterilmektedir.

Tablo 3. *LazyClassifier Kullanılarak Gerçekleştirilen Makine Öğrenmesi Modellerine İlişkin Sonuçlar*

Modeller	Genel Doğruluk Oranı (%)
Lojistik Regresyon	0.91
KNN	0.89
Naive Bayes	0.89
Random Forest	0.91
AdaBoostClassifier	0.83
DummyClassifier	0.71
SVM	0.91

Tablo 3'deki sonuçlar incelendiğinde, Çalışmada incelenen makine öğrenmesi modelleri arasında, Lojistik Regresyon, Random Forest ve SVM modelleri %91'lik genel doğruluk oranı ile en iyi performans sergilemiştir. Bu sonuçlar, bu modellerin uyku bozukluklarının sınıflandırılmasında etkili olduğunu ve veri setindeki örüntüleri başarıyla yakalayabildiğini göstermektedir. KNN ve Naive Bayes modelleri de %89'lik genel doğruluk oranı ile yakın bir performans göstermiştir. Bu, bu iki modelin de uyku bozukluklarının sınıflandırılmasında kullanılabileceğini ancak en iyi modeller kadar olmasa da yüksek bir doğrulukla tahmin yapabildiğini belirtir. AdaBoostClassifier modeli, %83'lük bir doğruluk oranı ile daha düşük bir performans sergilemiştir. Bu sonuç, AdaBoostClassifier'ın veri setindeki örüntüleri diğer modellere kıyasla daha az etkili bir şekilde yakaladığını göstermektedir. DummyClassifier, %71'lik genel doğruluk oranı ile modeller arasında en düşük performansı sergileyen model olmuştur. DummyClassifier genellikle bir karşılaştırma noktası olarak kullanılır ve beklendiği üzere, daha sofistike modellerin performansını aşamamıştır. Daha sonra, makine öğrenmesi algoritmalarının doğruluğunu arttırmak adına, aykırı değerler ele alınmıştır. Aykırı değerlerin ele alınması için, yaş, uyku süresi, fiziksel aktivite seviyesi, stres seviyesi, kalp atış hızı ve günlük adım sayısı gibi sayısal değişkenler için IQR yöntemi kullanılmıştır. Aykırı değerler temizlendikten sonra, kalan veri üzerinde standartlaştırma işlemi gerçekleştirilmiş ve ilk aşamada en iyi sonucu veren Lojistik Regresyon, KNN, Naive Bayes, Random Forest ve SVM modelleri yeniden eğitilmiştir. Bu modellere ilişkin sınıflandırma sonuçları Tablo 4'te gösterilmektedir.

Aykırı değerlerin temizlenmesi ve veri üzerinde standartlaştırma işlemi gerçekleştirildikten sonra, makine öğrenmesi modellerinin performansında gözle görülür bir iyileşme olduğu gözlemlenmiştir. Aykırı değerlerin temizlenmesi ve verilerin standartlaştırılması, modelin daha doğru tahminler yapmasını sağlamış ve genel doğruluk oranlarında iyileşme sağlamıştır. Tablo 4'te gösterilen sonuçlara göre, Lojistik Regresyon, KNN ve Naive Bayes modellerinin genel doğruluk oranı %93'e yükselmiştir. Buna karşın, Random Forest ve SVM modellerinin performansı da %94'e çıkmıştır. Bu durum, modellerin uyku bozukluklarını sınıflandırmadaki etkinliğinin arttığını göstermektedir. Çapraz doğrulama, modelin sadece eğitim veri seti üzerinde değil, görmediği veriler üzerinde de iyi performans göstermediğini test etmeyi sağlamaktadır.

Tablo 4. *Aykırı değerler ele alındıktan sonra makine öğrenmesi modellerinin performansı*

Modeller	Genel Doğruluk Oranı (%)
Lojistik Regresyon	0.93
KNN	0.93
Naive Bayes	0.93
Random Forest	0.94
SVM	0.94

Bu, modelin genelleştirme kabiliyetini anlamak için kritik bir öneme sahiptir. Yani, model yalnızca eğitim verisine özgü özellikleri değil, genel örüntüleri öğrenmiş midir sorusuna yanıt aramaktadır [24]. Çalışmalar, çapraz doğrulamanın, özellikle probleme özgü bilgi eksikliğinin olduğu senaryolarda, sınıflandırma hata oranlarının doğru tahminlerini elde etmek için çok önemli olduğunu göstermiştir. En uygun sınıflandırma yönteminin deneysel olarak seçilmesine olanak tanıyarak tek bir sınıflandırma stratejisinin kullanılmasına kıyasla daha yüksek ortalama performansa yol açmaktadır [25-26]. Ayrıca çapraz doğrulama, eğitim setiyle aynı dağılımdan alınan ve görülmeyen verilere genellenebilen tahminlerin elde edilmesine yardımcı olmaktadır. Bu, modelin performansının eğitim verilerine aşırı uymamasını ve yeni verilere etkili bir şekilde uygulanabilmesini sağlamaktadır [27]. K-kat çapraz doğrulama veya tek başına çapraz doğrulama gibi çapraz doğrulama yönteminin seçiminin sınıflandırma modelinin doğruluğunu etkileyebileceğini unutmamak önemlidir. Çapraz doğrulamanın uygun şekilde uygulanması, önyargılardan kaçınmak ve model değerlendirmesinin geçerliliğini sağlamak için gereklidir [28]. Bu nedenlerle, çapraz doğrulama yöntemi, makine öğrenmesi modellerinin değerlendirilmesinde önemli bir rol oynar ve model seçimi, modelin robustluğu ve genelleştirme kabiliyeti hakkında kritik bilgiler sunmaktadır. Bu çalışmada, Lojistik Regresyon, KNN, Naive Bayes, Random Forest ve SVM modellerinin değerlendirilmesi için çapraz doğrulama yöntemi uygulanmıştır. Çapraz doğrulama için kat sayısı (cv) olarak 13 değeri belirlenmiştir. “cross_val_score” fonksiyonunun cv parametresi, veri setinin kaç farklı alt kümeye bölüneceğini ve modelin kaç farklı iterasyonla değerlendirileceğini belirtmektedir. Bu durumda, modeller 13 farklı alt küme kullanılarak değerlendirilmiş ve her bir alt küme için modellerin performansı test edilmiştir. Bu yöntemle, modellerin genel performansına ilişkin daha güvenilir ve genel geçer sonuçlar elde edilmiştir. Bu, çapraz doğrulamanın genellikle modellerin dayanıklılığını ve genelleştirme yeteneğini test etmek için kullanıldığının bir göstergesidir. Modellerin 13 katlı çaprazlama işleminin sonucunda elde edilen karmaşıklık matrisleri Şekil 4’te gösterilmiştir. Ek olarak, sınıflandırma raporları Tablo 5’te gösterilmiştir.

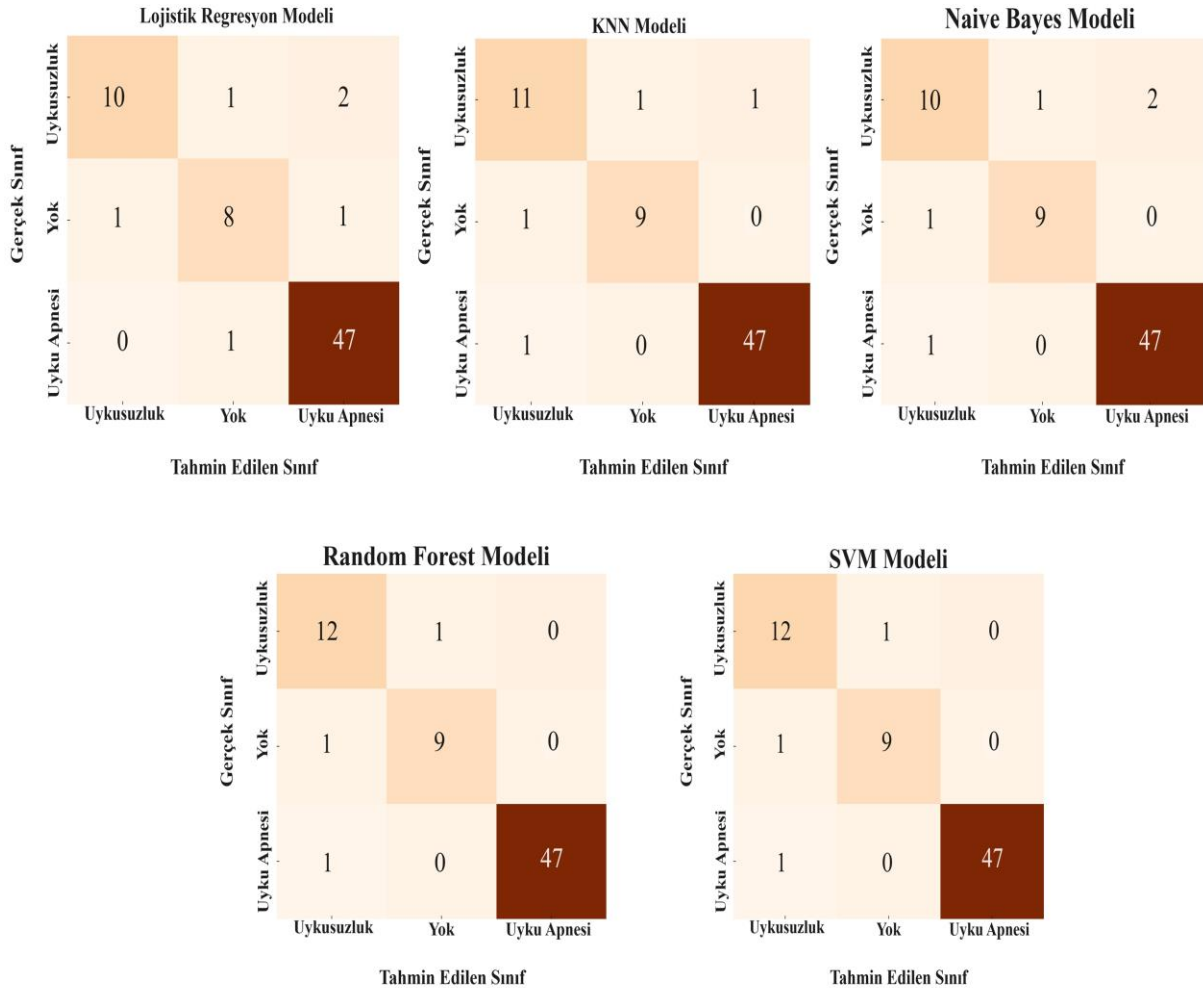
Tablo 5, çapraz doğrulama işlemi sonrasında uygulanan makine öğrenmesi modellerine ilişkin sınıflandırma raporunu göstermektedir. Bu rapor, her modelin uyku bozukluğu türlerine (Uykusuzluk, Uyku Apnesi, Yok) göre kesinlik, geri çağırma ve F-1 skoru gibi performans metriklerini içermektedir. Ayrıca, her modelin genel doğruluk oranı da belirtilmiştir. Bu sonuçlar, modellerin uyku bozukluklarını ne kadar iyi sınıflandırdığını ve her bir uyku bozukluğu türü için modelin performansını göstermektedir. Lojistik regresyon modeli, Uykusuzluk, Uyku Apnesi ve Yok durumları için yüksek kesinlik ve geri çağırma oranları elde etmiş, özellikle Uyku Apnesi durumunda %94 kesinlik ve %98 geri çağırma ile dikkat çekici performans sergilemiştir. Genel doğruluk %92 olup, uyku bozukluklarını sınıflandırmada etkili bir model olduğunu göstermektedir. Makro ve ağırlıklı ortalama değerleri, modelin genel olarak dengeli bir performans sergilediğini belirtir.

Tablo 5. *Çapraz doğrulama işleminin sonucunda modellere ilişkin sınıflandırma raporu*

Modeller		Kesinlik	Geri Çağırma	F-1 Skoru	Destek
Lojistik Regresyon	Uykusuzluk	0.91	0.77	0.83	13
	Uyku Apnesi	0.94	0.98	0.96	48
	Yok	0.80	0.80	0.80	10
	Doğruluk			0.92	71
	Makro Ortalama	0.88	0.85	0.86	71
	Ağırlıklı Ortalama	0.91	0.92	0.91	71
KNN	Uykusuzluk	0.85	0.85	0.85	13
	Uyku Apnesi	0.98	0.98	0.98	48
	Yok	0.90	0.90	0.90	10
	Doğruluk			0.94	71
	Makro Ortalama	0.91	0.91	0.91	71
	Ağırlıklı Ortalama	0.94	0.94	0.94	71
Naive Bayes	Uykusuzluk	0.83	0.77	0.80	13
	Uyku Apnesi	0.96	0.98	0.97	48
	Yok	0.90	0.90	0.90	10
	Doğruluk			0.93	71
	Makro Ortalama	0.90	0.88	0.89	71
	Ağırlıklı Ortalama	0.93	0.93	0.93	71
Random Forest	Uykusuzluk	0.91	0.77	0.83	13
	Uyku Apnesi	0.98	1.00	0.99	48
	Yok	0.82	0.90	0.86	10
	Doğruluk			0.94	71
	Makro Ortalama	0.90	0.89	0.89	71
	Ağırlıklı Ortalama	0.94	0.94	0.94	71
SVM	Uykusuzluk	0.91	0.77	0.83	13
	Uyku Apnesi	0.98	1.00	0.99	48
	Yok	0.82	0.90	0.86	10
	Doğruluk			0.94	71
	Makro Ortalama	0.90	0.89	0.89	71
	Ağırlıklı Ortalama	0.94	0.94	0.94	71

KNN modeli, tüm uyku bozukluğu türleri için Lojistik Regresyona kıyasla biraz daha yüksek performans göstermiştir. Özellikle Uyku Apnesi durumu için %98 kesinlik ve %98 geri çağırma

oranıyla mükemmel bir sonuç elde etmiştir. Genel doğruluk %94 ile yüksek değer elde etmiştir. Makro ve ağırlıklı ortalama değerlerinin yüksekliği, modelin tüm sınıflar arasında dengeli ve güçlü bir performans sergilediğini gösterir. Naive Bayes, Uykusuzluk ve Yok durumları için diğer modellere göre biraz daha düşük performans sergilemiş ancak Uyku Apnesi durumunda yüksek bir başarı oranı elde etmiştir. Genel doğruluk %93 ile oldukça yüksektir. Makro ve ağırlıklı ortalama değerleri, modelin genel olarak dengeli bir performans sergilediğini belirtir. Random Forest ve SVM, uyku bozukluklarını sınıflandırmada özellikle Uyku Apnesi durumunda mükemmel performans göstermiştir. Uyku Apnesi durumu için %98 kesinlik ve %100 geri çağırma oranıyla dikkat çekicidir. Her iki modelin de genel doğruluk oranı %94'tür, bu da onları bu veri seti için en etkili modellerden biri yapmaktadır. Makro ve ağırlıklı ortalama değerleri, modellerin genel olarak yüksek ve dengeli performans sergilediğini göstermektedir. Sonuç olarak, çapraz doğrulama sonuçları, incelenen makine öğrenmesi modellerinin uyku bozukluklarını sınıflandırmada yüksek ve dengeli bir performans sergilediğini göstermektedir. Bu bulgular, model seçimi ve uyku bozukluklarının tanı ve tedavisindeki potansiyel uygulamalar açısından önemli bilgiler sağlamaktadır. Modellerin yüksek genel doğruluk oranları ve dengeli performans metrikleri, uyku bozukluklarının etkili bir şekilde sınıflandırılmasında makine öğrenmesi yaklaşımlarının önemini vurgulamaktadır.



Şekil 4. Modellerin 13 katlı çaprazlama işleminin sonucunda elde edilen karmaşıklık matrisleri

Çapraz doğrulama sonuçlarına dayanarak, Lojistik Regresyon, KNN, Naive Bayes, Random Forest ve SVM modellerinin performansları ayrıntılı olarak incelenmiştir. Lojistik Regresyon modeli, uyku apnesi sınıfını oldukça yüksek doğrulukla tahmin ederken, uykusuzluk ve "yok" sınıflarında bazı karışıklıklar yaşamıştır; uykusuzluk sınıfında 10 doğru tahmin yaparken 1 kişiyi uyku apnesi ve 2 kişiyi

"yok" olarak yanlış sınıflandırmıştır. KNN modeli, uyku apnesi sınıfında benzer şekilde yüksek doğruluk sağlamış, uykusuzluk ve "yok" sınıflarında ise daha az hata yapmıştır. Naive Bayes modeli, uyku apnesi sınıfında benzer performans sergilemiş ancak uykusuzluk sınıfında diğer modellere göre daha fazla hata yapmıştır; 10 doğru tahmine karşılık 2 kişiyi uyku apnesi olarak yanlış sınıflandırmıştır. Random Forest modeli ise tüm sınıflarda yüksek doğruluk sağlamış ve özellikle "uykusuzluk" sınıfında en az hatayı yapmıştır; 12 doğru tahmin yaparken sadece 1 kişiyi uyku apnesi olarak yanlış sınıflandırmıştır. Benzer şekilde, SVM modeli de "uyku apnesi" ve "yok" sınıflarında başarılı olup, uykusuzluk sınıfında yüksek doğruluk göstermiştir.

Bu analizler, uyku apnesi sınıfının tüm modeller tarafından yüksek doğrulukla tahmin edildiğini, ancak "uykusuzluk" ve "yok" sınıflarında performans farklılıklarının bulunduğunu ortaya koymaktadır. Uykusuzluk sınıfında en başarılı modeller Random Forest ve SVM olup, her iki model de en az hatayı yapmıştır. Naive Bayes modeli ise uykusuzluk sınıfında diğer modellere kıyasla daha fazla hata yapmıştır. Genel performans açısından, Random Forest ve SVM modelleri tüm sınıflarda en yüksek doğruluğu sağlamış ve uyku hastalıklarının sınıflandırılmasında üstünlük göstermiştir. Bu sonuçlar, uyku hastalıklarının sınıflandırılmasında özellikle Random Forest ve SVM modellerinin etkinliğini vurgulamakta ve bu modellerin klinik uygulamalarda kullanılabilirliğini desteklemektedir.

4. Sonuç ve Tartışma

Bu çalışma, uyku sağlığı ve yaşam tarzı arasındaki ilişkileri analiz ederek, uyku bozukluğu sınıflandırmasında makine öğrenmesi yöntemlerinin etkinliğini kanıtlamaktadır. Araştırma sonuçları, cinsiyet, yaş ve meslek gibi demografik faktörlerin uyku bozuklukları üzerinde belirgin bir etkiye sahip olduğunu ortaya koymaktadır. Ayrıca, makine öğrenmesi algoritmalarının, özellikle Random Forest ve Support Vector Machine (SVM) modellerinin, yüksek doğruluk oranları ve dengeli sınıflandırma performansları ile bu bozuklukların tanımlanmasında güçlü araçlar olduğu gözlemlenmiştir. Çalışmada, uyku bozukluklarının prevalansının cinsiyet, yaş ve meslek gibi faktörlerle anlamlı bir ilişkisi bulunmuştur. Uykusuzluk durumunun, belirli meslek gruplarında daha sık rastlanan bir sorun olduğu belirlenmiştir. Ayrıca, model performansı açısından, aykırı değerlerin temizlenmesi ve standartlaştırma işlemlerinin ardından makine öğrenmesi modellerinin doğruluğu önemli ölçüde artmıştır. Bu, uyku bozukluğu sınıflandırmasında modellerin genelleştirme kabiliyetinin ve robustluğunun, uygun veri ön işleme teknikleri ile nasıl iyileştirilebileceğini göstermektedir. Ardından, çapraz doğrulama, modellerin eğitim verilerine aşırı uyum göstermeyerek, yeni ve görülmemiş verilere genelleme yapabilme yeteneğini başarıyla test etmiştir. Bu çalışmanın sonuçlarına dayanarak şu öneriler verilmektedir. Çalışmanın ortaya koyduğu bulguların eğitim alanına entegrasyonu önem arz etmektedir. Araştırmanın gösterdiği üzere, uyku bozukluğunun meslek grupları üzerindeki etkisi dikkate alındığında, işyerlerinde uyku sağlığının önemi konusunda bilinçlendirme ve eğitim programlarının geliştirilmesi kritik öneme sahiptir. Bu bağlamda, işyerlerine yönelik uygulanacak eğitim seminerleri, uyku hijyeninin önemini vurgulayarak çalışanların uyku kalitelerini artırmak için pratik adımlar atabilecekleri stratejiler sunulmalıdır. Ayrıca, uyku düzeni ve kalitesini etkileyen faktörler hakkında farkındalığı artırmak ve bu faktörleri yönetme becerilerini geliştirmek için kurumsal düzeyde eğitimler düzenlenmelidir. Eğitim kurumları, özellikle sağlık ve eğitim gibi yoğun stres altında çalışan meslek grupları için, uyku bozuklukları ve bunların iş performansı üzerindeki etkileri hakkında özel dersler veya atölye çalışmaları planlanmalıdır. Bu programlar, uyku düzenini iyileştirme yöntemleri, stres yönetimi teknikleri ve zaman yönetimi gibi konuları kapsayabilir. Eğitim müfredatına, özellikle tıp ve sağlık bilimleri öğrencileri için uyku tıbbı ve uyku bozuklukları konusunda derinlemesine dersler eklenmesi, geleceğin sağlık profesyonellerinin bu önemli sağlık sorununa karşı daha donanımlı olmalarını sağlayacaktır. Bunun yanı sıra, uyku bozukluklarının tanı ve tedavisinde çalışacak sağlık profesyonellerine yönelik sürekli profesyonel gelişim programları, makine öğrenimi ve yapay zekâ gibi yeni teknolojik araçların kullanımını içerecek şekilde güncellenmelidir. Bu tür eğitimler, sağlık hizmetlerinin verimliliğini artıracak ve hastalara sunulan bakım kalitesini yükseltecek potansiyel uygulamalara işaret etmektedir. Özellikle eğitim sektöründe, öğretmenlerin ve öğrencilerin sağlıklı uyku alışkanlıkları geliştirmelerine yardımcı olacak okul bazlı müdahalelerin tasarlanması gerekmektedir. Okullar, uyku eğitimi program-

larını sağlık ve refah müfredatlarının bir parçası haline getirerek, genç bireylerin erken yaşlarda sağlıklı uyku düzeni konusunda bilinçlenmelerini sağlamalıdır.

Çalışmadan elde edilen bulgular, literatürdeki benzer çalışmalarla karşılaştırılarak değerlendirildiğinde, öncelikle, Random Forest ve SVM modellerinin yüksek doğruluk oranları, uyku bozukluklarının sınıflandırılmasında etkili araçlar olduklarını göstermektedir. Bu sonuçlar, Sundararajan vd. [9] tarafından bileğe takılan ivmeölçer verileri kullanılarak yapılan çalışmada rastgele ormanların uyku sınıflandırmasında etkili olduğunu göstermesiyle uyumludur. Benzer şekilde, Kwon vd. [6] tarafından yapılan çalışmada evrişimli derin sinir ağlarının yüksek sınıflandırma doğruluğu sağladığı belirtilmiştir, bu da çalışmamızda kullanılan modellerin başarısını desteklemektedir. Meslek türleri ve stres seviyelerinin uyku bozuklukları üzerindeki etkisi, literatürde geniş bir şekilde ele alınmıştır. Örneğin, Karasek'in İş Talepleri-Kaynakları modeli, işin talepleri ve bireyin bu talepleri karşılamak için sahip olduğu kaynaklar arasındaki dengesizliğin stres yaratma potansiyeline işaret etmektedir. Çalışmamızda, yüksek stres seviyelerine sahip meslek gruplarının uyku bozuklukları yaşama olasılığının daha yüksek olduğunu bulduk. Bu bulgular, Yperen [20] ve Pelfrene vd. [21] tarafından yapılan çalışmalarda belirtilen iş stresi teorileriyle uyumludur. Ayrıca, çalışmamızda kullanılan makine öğrenimi modellerinin performansları, Rakhonde vd. [7] tarafından yapılan uyku evresi sınıflandırma çalışmasında Stokastik Gradyan İnişi (SGD) kullanılarak elde edilen sonuçlarla da karşılaştırılabilir. Çalışmamızda Naive Bayes modeli uyku apnesi durumunda yüksek doğruluk oranı sağlarken, Rakhonde vd. [7] çalışmasında SGD'nin benzer bir başarı elde ettiği görülmüştür. Bu çalışmanın literatüre katkısı, uyku bozukluklarının sınıflandırılmasında makine öğrenimi algoritmalarının etkinliğini gösterirken, aynı zamanda meslek türleri ve stres seviyelerinin bu bozukluklar üzerindeki etkilerini de vurgulamaktadır. Literatürde, uyku bozukluklarının belirli meslek grupları arasında farklılık gösterdiği ve stresin uyku kalitesi üzerindeki etkisi konusunda çeşitli görüşler mevcuttur. Çalışmamız bu konuda yeni veriler sunarak, özellikle uyku sağlığının iyileştirilmesine yönelik stratejilerin geliştirilmesinde önemli bir kaynak olabilir. Kullanılan yöntemlerin literatürdeki yeri ve önceki çalışmalardan nasıl farklılaştığı da detaylandırılmalıdır. Çalışmamızda, Random Forest ve SVM modellerinin yanı sıra Naive Bayes, KNN ve Lojistik Regresyon gibi çeşitli makine öğrenimi modelleri kullanılmıştır. Bu modellerin seçimi, literatürdeki benzer çalışmalarda elde edilen başarılar ve doğruluk oranları göz önünde bulundurularak yapılmıştır. Örneğin, Cho vd. [12] tarafından yapılan çalışmada rastgele orman ve torbalama yöntemlerinin uyku aşamalarını ayırt etmede etkili olduğu belirtilmiştir. Sonuç olarak, bu çalışmanın bulguları, literatürdeki mevcut bilgilerle karşılaştırıldığında uyku bozukluklarının sınıflandırılmasında makine öğrenimi algoritmalarının etkinliğini ve çeşitli meslek gruplarının uyku kalitesi üzerindeki etkisini ortaya koymaktadır.

Bu çalışmanın potansiyel kısıtlılıkları, araştırma tasarımının ve uygulamanın birtakım sınırlamaları etrafında şekillenmektedir. İlk olarak, veri setinin nispeten sınırlı boyutu, elde edilen bulguların genişletilebilirliği üzerinde doğal sınırlar oluşturmakta ve bu bulguların farklı popülasyonlara veya çeşitli demografik yapıya sahip topluluklara uygulanabilirliğini sınırlamaktadır. Çalışmada kullanılan veri seti, yalnızca 374 katılımcıyı içermektedir. Bu, makine öğrenmesi modellerinin genelleştirme yeteneğini sınırlayabilir ve sonuçların geniş popülasyonlara uygulanabilirliğini azaltabilir. İkinci olarak, uyku bozukluğunun ölçümünde kullanılan yöntemlerin öznel niteliği, ölçüm hatalarına ve yanlılıklara yol açabilecek faktörler arasında yer almaktadır. Ayrıca, veri setinin içerdiği değişkenlerin ve bu değişkenlerin ölçümlerinin tüm potansiyel etkileyici faktörleri kapsamaması veya yeterli derinliğe ulaşmaması, analizin kapsamını ve yorumlanabilirliğini sınırlayabilir. Seçilen makine öğrenimi modellerinin ve bunların hiperparametre ayarlarının sonuçlar üzerinde önemli bir etkisi bulunmakta olup, farklı modelleme tekniklerinin veya parametre ayarlarının farklı sonuçlar üretebileceğinin bilinciyle yaklaşılmalıdır. Ayrıca, kullanılan modellerin "kara kutu" doğası, modellerin karar verme süreçlerinin anlaşılmasını güçleştirmekte ve modellerin uygulanabilirliğine ilişkin soru işaretleri doğurabilmektedir. Model validasyonu için tercih edilen çapraz doğrulama yöntemi, modelin genelleştirme kabiliyetine dair sağladığı içgörüler bakımından eleştirel olsa da, seçilen kat sayısının ve rastgeleliğin sonuçlara etkisi göz ardı edilemez. Son olarak, bu çalışmanın gözlemsel bir araştırma olması, nedensel ilişkilerin tespiti açısından deneysel çalışmalara kıyasla sınırlı kalmasına neden olur. Elde edilen ilişkilerin yalnızca korelatif olduğu ve nedensel olarak yorumlanmaması gerektiği unutulmamalıdır. Bu kısıtlılıklar, bulguların daha geniş bağlamlarda değerlendirilmesi ve gelecek araştırmalarda metodolojik iyileştirmeler

yapılması gerektiğinin altını çizmektedir. Bu nedenle, araştırmanın sonuçlarının ışığında gelecekteki çalışmalarda bu kısıtlamaların üstesinden gelecek yolların araştırılması önerilmektedir. Gelecek çalışmalarda, mevcut araştırmanın sonuçlarını daha geniş bir perspektiften ele almak ve bulguların kapsamını genişletmek amaçlamalıdır. Bu bağlamda, veri setinin boyutu ve çeşitliliği önemli bir başlangıç noktasıdır. Daha geniş ve demografik açıdan çeşitli bir katılımcı havuzu kullanarak, elde edilen modellerin genel nüfus üzerindeki uygulanabilirliğinin ve güvenilirliğinin artırılması hedeflenmelidir. Bu, farklı etnik kökenlerden, yaş gruplarından ve sosyoekonomik statülerden bireylerin dahil edilmesini gerektirir. Özgül olarak, uyku bozukluğu araştırmaları, biyolojik göstergeler gibi objektif uyku kalitesi ölçümlerini de içermelidir. Bu sayede, modelin öznel raporlama yanlılıklarından etkilenmeden, uyku bozukluklarını daha doğru bir şekilde sınıflandırması sağlanabilir. Ayrıca, uzunlamasına çalışmalar, uyku bozukluğunun ilerleyişini ve tedaviye yanıtını izleme fırsatı sunarak, zaman içindeki değişiklikleri ve tedavi sonuçlarını değerlendirme imkânı verebilir. Bunun yanı sıra, makine öğrenimi modellerinin geliştirilmesi ve doğrulanması da, farklı algoritmaların ve hiperparametre ayarlarının sistematik bir şekilde test edilmesini gerektirir. Model karar süreçlerinin açıklanabilirliğini artırmak için, modelin nasıl karar verdiğini ve hangi özelliklerin önemli olduğunu açıklayan teknikler geliştirilmeli ve uygulanmalıdır. Ayrıca, bulguların klinik uygulamaya dönüşümünü kolaylaştırmak için, gerçek dünya verileri kullanılarak modellerin uygulanabilirliği ve etkililiği değerlendirilmelidir. Klinik ortamlarda gerçekleştirilecek pilot çalışmalar, modellerin uyku bozukluklarının tanı ve tedavi süreçlerine entegrasyonunu test edebilir. Bu tür çalışmalar, makine öğrenimi tabanlı araçların klinik karar verme süreçlerine entegre edilmesi ve hasta bakımının iyileştirilmesi yönündeki pratik adımları belirleyebilir.

Çıkar Çatışması

Yazar(lar) bu makaleyle ilgili herhangi bir çıkar çatışması olmadığını bildirir.

Araştırma ve Yayın Etiği Beyanı

Yazar(lar) bu çalışmanın araştırma ve yayın etiğine uygun olduğunu beyan eder.

Kaynakça

- [1] St-Onge, M., Grandner, M. A., Brown, D. L., Conroy, M. B., Jean-Louis, G., Coons, M. J., ... & Bhatt, D. L. (2016). Sleep duration and quality: impact on lifestyle behaviors and cardiometabolic health: a scientific statement from the american heart association. *Circulation*, 134(18), 367-386.
- [2] Bruce, E., Lunt, L., & McDonagh, J. E. (2017). Sleep in adolescents and young adults. *Clinical Medicine*, 17(5), 424-428.
- [3] Chattu, V. K., Manzar, M. D., Kumary, S., Burman, D., Spence, D. W., & Pandi-Perumal, S. R. (2018). The global problem of insufficient sleep and its serious public health implications. *Healthcare*, 7(1), 1-16.
- [4] Allen, S., Akram, U., & Ellis, J. (2020). Examination of sleep health dimensions and their associations with perceived stress and health in a uk sample. *Journal of Public Health*. 28, 42(1), 34-41.
- [5] Urtnasan, E., Joo, E. Y., & Lee, K. (2021). Ai-enabled algorithm for automatic classification of sleep disorders based on single-lead electrocardiogram. *Diagnostics*, 11(11), 2054.
- [6] Kwon, K., Kwon, S., & Yeo, W. (2022). Automatic and accurate sleep stage classification via a convolutional deep neural network and nanomembrane electrodes. *Biosensors*, 12(3), 155.

- [7] Rakhonde, M. A., Wagh, K., & Mante, R. V. (2020). Sleep stage classification for prediction of human sleep disorders by using machine learning approach. *International Journal of Innovative Science and Research Technology*, 5(7), 1268-1272.
- [8] Pradeepkumar, J., Anandakumar, M., Vinith, K., Suntharalingham, D., Kappel, S. L., Silva, A. C. D., ... & Edussooriya, C. U. S. (2022). Towards interpretable sleep stage classification using cross-modal transformers.
- [9] Sundararajan, K., Georgievska, S., Lindert, B. H. W. T., Gehrman, P., Ramautar, J., Mazzotti, D. R., ... & Hees, V. T. v. (2021). Sleep classification from wrist-worn accelerometer data using random forests. *Scientific Reports*, 11(1), 24.
- [10] Delimayanti, M. K., Purnama, B., Nguyen, N. M., Faisal, M. R., Mahmudah, K. R., Indriani, F., ... & Satou, K. (2020). Classification of brainwaves for sleep stages by high-dimensional fft features from eeg signals. *Applied Sciences*, 10(5), 1797.
- [11] Yulita, I. N., Fanany, M. I., & Arymurthy, A. M. (2018). Fast convolutional method for automatic sleep stage classification. *Healthcare Informatics Research*, 24(3), 170.
- [12] Cho, T., Sunarya, U., Yeo, M. S., Hwang, B. K., Koo, Y. S., & Park, C. (2019). Deep-actinet: end-to-end deep learning architecture for automatic sleep-wake detection using wrist actigraphy. *Electronics*, 8(12), 1461.
- [13] KaggleVeriseti, <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset>
- [14] Saxena, R., Sharma, S. K., Gupta, M., & Sampada, G. C. (2022). A novel approach for feature selection and classification of diabetes mellitus: machine learning methods. *Computational Intelligence and Neuroscience*, 2022, 1-11.
- [15] Battineni, G., Chintalapudi, N., Amenta, F., & Traini, E. (2020). A comprehensive machine-learning model applied to magnetic resonance imaging (mri) to predict alzheimer's disease (ad) in older subjects. *Journal of Clinical Medicine*, 9(7), 2146.
- [16] Yang, Z., Chen, C., Li, H., Yao, L., & Zhao, X. (2020). Unsupervised classifications of depression levels based on machine learning algorithms perform well as compared to traditional norm-based classifications. *Frontiers in Psychiatry*, 11.
- [17] Sharma, A. (2021). Guided parallelized stochastic gradient descent for delay compensation. *Applied Soft Computing*, 102, 107084.
- [18] Karasek, R. (1979). Job demands, job decision latitude and mental strain: Implications for job redesign. *Administrative Science Quarterly*, 24, 285-306.
- [19] Siegrist, J. (1996). Adverse health effects of high-effort/low-reward conditions. *Journal of Occupational Health Psychology*, 1, 27-41.
- [20] Yperen, N. W. V. (2000). A multilevel analysis of the demands-control model: is stress at work determined by factors at the group level or the individual level?. *Journal of Occupational Health Psychology*, 5(1), 182-190.
- [21] Pelfrene, E., Vlerick, P., Kittel, F., Mak, R., Kornitzer, M., & Backer, G. D. (2002). Psychosocial work environment and psychological well-being: assessment of the buffering effects in the job demand-control (-support) model in belstress. *Stress and Health*, 18(1), 43-56.

- [22] Rubino, C., Perry, S. J., Milam, A., Spitzmüller, C., & Zapf, D. (2012). Demand–control–person: integrating the demand–control and conservation of resources models to test an expanded stressor–strain model. *Journal of Occupational Health Psychology*, 17(4), 456-472.
- [23] Spiegelaere, S. D., Ramioul, M., & Gyes, G. V. (2017). Good employees through good jobs. *Employee Relations*, 39(4), 503-522.
- [24] Gemert, J. C. v., Snoek, C. G. M., Veenman, C. J., & Smeulders, A. (2006). The influence of cross-validation on video classification performance. *Proceedings of the 14th ACM International Conference on Multimedia*.
- [25] Schaffer, C. (1993). Selecting a classification method by cross-validation. *Machine Learning*, 13(1), 135-143.
- [26] Wichard, J., Cammann, H., Stephan, C., & Tolxdorff, T. (2008). Classification models for early detection of prostate cancer. *Journal of Biomedicine and Biotechnology*, 2008, 1-7.
- [27] LeDell, E., Petersen, M., & Laan, M. J. v. d. (2015). Computationally efficient confidence intervals for cross-validated area under the roc curve estimates. *Electronic Journal of Statistics*, 9(1).
- [28] Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., Velzen, E. J. J. v., & Dorsten, F. A. v. (2008). Assessment of plsda cross validation. *Metabolomics*, 4(1), 81-89.