

## İNTERNET BAĞIMLILIĞI SKORLARININ TAHMİNİNDE FARKLI MAKİNE ÖĞRENME MODELLERİNİN KARŞILAŞTIRILMASI

Serpil SEVİMLİ DENİZ<sup>1</sup>

### ÖZET

Bu makale, internet bağımlılığı skorlarının tahmininde makine öğrenme modellerinin kullanımının önemini vurgulamakta ve makine öğrenmesinin sosyal bilimlerdeki artan rolüne dikkat çekmektedir. Geleneksel analiz yöntemleri, büyük ve karmaşık veri kümelerini analiz etmekte zorlanırken, makine öğrenmesi bu tür veri setlerinden anlamlı desenler ve ilişkiler çıkarmada oldukça başarılıdır. Bu durum, internet bağımlılığı, depresyon ve toplumsal davranışlar gibi çok katmanlı sosyal sorunlar için özellikle önemlidir; çünkü bu sorunlar, bireylerin psikolojik ve davranışsal eğilimlerini etkileyen birçok faktörü içerir.

Çalışma, Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Support Vector Regression (SVR) ve Gradient Boosting gibi makine öğrenme modellerini karşılaştırmakta ve her bir modelin internet bağımlılığı skorlarını tahmin etme performansını doğruluk ve diğer metrikler üzerinden değerlendirmektedir. Bulgular, ensemble modellerin (Random Forest ve Gradient Boosting) daha karmaşık veri setlerinde diğer basit modellere göre daha başarılı olduğunu göstermektedir.

Bu çalışmanın özgünlüğü, internet bağımlılığı tahmini için makine öğrenmesi modellerini kapsamlı bir şekilde karşılaştırmasıdır. Daha önceki çalışmaların çoğu psikolojik yaklaşımlara odaklanırken, bu araştırma sosyal bilimlerle teknik yöntemleri birleştirerek alana yenilikçi bir bakış açısı kazandırmaktadır.

**Anahtar Kelimeler:** *İnternet Bağımlılığı, Makine Öğrenimi, Logistic Regression*

<sup>1</sup> Serpil SEVİMLİ DENİZ, Dr. Öğr. Üyesi, Van Yüzüncü Yıl Üniversitesi, sdeniz@yyu.edu.tr, ORCID: 0000-0002-8559-1107.

# COMPARISON of DIFFERENT MACHINE LEARNING MODELS in PREDICING INTERNET ADDICTION SCORES

## ABSTRACT

This article focuses on the importance of using machine learning models in predicting internet addiction scores, emphasizing the growing relevance of machine learning in social sciences. Traditional methods often struggle to analyze large and complex datasets, whereas machine learning excels at identifying patterns and relationships within such data. This is particularly important for complex social issues like internet addiction, depression, and societal behaviors, which involve multiple psychological and behavioral factors.

The study compares several machine learning models, including Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Support Vector Regression (SVR), and Gradient Boosting. Each model is evaluated based on its accuracy and performance in predicting internet addiction. The findings show that ensemble models like Random Forest and Gradient Boosting perform significantly better than simpler models such as Logistic Regression and SVR, which struggle with complex data.

What makes this study unique is its comprehensive comparison of machine learning models specifically for predicting internet addiction. Unlike previous studies that have predominantly focused on psychological approaches, this research integrates technical methodologies with social science perspectives. The study demonstrates how ensemble models can provide more accurate predictions, highlighting the importance of data-driven policy recommendations for addressing social problems.

**Keywords:** *Internet Addiction, Machine Learning, Logistic Regression*

## 1. GİRİŞ

İnternet, bilgiye erişim, eğlence ve sosyal etkileşim için vazgeçilmez bir araç haline gelmiş olsa da aşırı ve kontrolsüz kullanımı giderek artan bir sorun olan internet bağımlılığına yol açmaktadır. İnternet bağımlılığı, bireyin internet kullanımını kontrol edememesi, günlük işlevlerinde bozulma yaşaması ve sosyal ilişkilerinde olumsuz etkiler doğurmasıyla tanımlanan bir durumdur. Bu bağımlılık türü, özellikle gençler arasında yaygınlaşmakta ve bireylerin hem psikolojik hem de sosyal sağlıklarını ciddi şekilde etkileyebilmektedir. Dolayısıyla, internet bağımlılığını erken aşamada tespit edebilmek ve müdahale edebilmek için etkili tahmin yöntemlerine ihtiyaç duyulmaktadır.

Bu noktada, makine öğrenimi algoritmaları, büyük veri kümelerinden kalıplar çıkarmak ve gelecekteki davranışları tahmin etmek

amacıyla güçlü bir araç olarak öne çıkmaktadır. Geleneksel yöntemler genellikle anket ve psikolojik değerlendirme araçlarına dayanırken, makine öğrenimi, çok daha geniş çaplı ve karmaşık veri kümeleriyle çalışabilme kapasitesi sayesinde internet bağımlılığı gibi çok boyutlu sosyal sorunları daha etkili bir şekilde analiz edebilme potansiyeline sahiptir. Bununla birlikte, internet bağımlılığı skorlarını tahmin etmek için hangi makine öğrenimi modelinin daha etkili olduğu konusunda sınırlı sayıda çalışma bulunmaktadır.

Bu çalışma, internet bağımlılığı skorlarını tahmin etmek için farklı makine öğrenimi algoritmalarının performanslarını karşılaştırmayı amaçlamaktadır. Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Support Vector Regression (SVR) ve Gradient Boosting gibi yaygın kullanılan makine öğrenimi modelleri incelenmiştir. Çalışma, bu modellerin performansını, ortalama hata kareleri (MSE) ve  $R^2$  skorları gibi yaygın performans metrikleri üzerinden değerlendirerek hangi modelin internet bağımlılığı tahmini için en uygun olduğunu ortaya koymayı hedeflemektedir.

Bu çalışmanın özgünlüğü, farklı makine öğrenimi algoritmalarının ilk kez internet bağımlılığı tahmini için bu denli kapsamlı bir şekilde karşılaştırılmasıdır. Birçok çalışma internet bağımlılığı sorununu yalnızca psikolojik açıdan ele alırken, bu çalışma sosyal bilimler ile teknik yaklaşımları bir araya getirerek bu alana yenilikçi bir bakış açısı sunmaktadır. Ensemble modeller, birden fazla makine öğrenimi algoritmasının bir arada kullanılmasıyla model performansını artırmayı amaçlayan güçlü yöntemlerdir. Bu modeller, bireysel algoritmaların hatalarını azaltarak daha doğru tahminler yapmayı sağlar. Ensemble yöntemlerin (Random Forest ve Gradient Boosting) bu alanda ne kadar etkili olduğunu gösteren bulgular, literatüre önemli bir katkı sağlamaktadır. Modellerin karşılaştırılması, sadece hangi algoritmanın daha yüksek doğruluk sunduğunu göstermekle kalmamakta, aynı zamanda sosyal bilimlerde makine öğrenimi kullanımının nasıl optimize edilebileceğine dair önemli ipuçları sunmaktadır. Bu bağlamda, bu çalışma, internet bağımlılığı tahmini gibi sosyal bilimlerdeki karmaşık problemlerin makine öğrenimi ile nasıl analiz edilebileceğini inceleyen ve uygulamalı bir örnek sunan öncü bir çalışma niteliğindedir.

## 2. LİTERATÜR TARAMASI

Yapay zekâ/akıllı öğrenme teknolojileri her geçen gün hayatımıza biraz daha girmekte ve etki seviyesini giderek arttırmaktadır. Yapabildikleriyle insanı şaşkına çevirmenin yanında işleri kolaylaştırdığı ve çalışma alanlarına yönelik oldukça fayda sağladığı görülmektedir

(Altıntop, 2023). Makine öğrenimi tabanlı sınıflandırma modelleri, internet bağımlılığını öngörmeye başarılı olmuştur. Theng vd. (2023) tarafından yapılan bir çalışmada, üniversite öğrencileri arasında internet bağımlılığını tespit etmek için Convolutional Neural Network (CNN), K-En Yakın Komşu (KNN) ve Lojistik Regresyon (LR) gibi üç makine öğrenimi modeli karşılaştırılmıştır. Bu çalışmanın sonuçlarına göre CNN modeli en yüksek doğruluk, kesinlik, geri çağırma ve F1 skoruna sahip olmuştur (Theng vd., 2023). Singh ve Babbar (2017) tarafından yapılan bir çalışmada, Bayes Ağları kullanılarak İnternet Bağımlılığı Bozukluğu (IAB) tahmin edilmiştir. Bu model, internet bağımlılığı bozukluğunu belirlemek için önemli parametreler olan aşırı kullanım ve kontrol kaybı gibi faktörleri temel almıştır. Çalışmanın sonucunda, Bayes Ağları'nın IAB varlığını yüksek doğrulukla tahmin ettiği gösterilmiştir (Singh & Babbar, 2017). İnternet Bağımlılığı Testi (IAT), internet bağımlılığını taramada yaygın olarak kullanılan bir araçtır. İtalya'da yapılan bir çalışmada, IAT'nin iki ana faktöre sahip olduğu belirlenmiştir: "Duygusal ve Bilişsel Takıntılar" ve "Kontrol Kaybı" (Fioravanti & Casale, 2015). Benzer şekilde, Hong Kong'da ergenler üzerinde yapılan bir çalışma, IAT'nin geçerlilik ve güvenilirlik açısından uygun olduğunu göstermiştir (Lai vd., 2013). Zhang ve Xin (2013) tarafından yapılan bir çalışmada, IAT ve Chen İnternet Bağımlılığı Ölçeği (CIAS) karşılaştırılmıştır. Çalışma, her iki ölçeğin birleştirilmesinin daha iyi sınıflandırma tutarlılığı sağladığını göstermiştir. Yeni oluşturulan ölçek, özellikle orta düzeyde bağımlılık durumlarında daha yüksek güvenilirlik sunmuştur (Zhang & Xin, 2013). Hussain vd. (2020) tarafından yapılan bir çalışmada, Latent Sınıf Analizi (LCA) ile üniversite öğrencilerinin internet bağımlılığı profilleri incelenmiştir. Çalışma, aynı kümeleme kriterlerinin tüm faktörler için uygulanmasının yanlış sınıflandırmalara neden olabileceğini öne sürmüştür (Hussain vd., 2020). Shae ve Tsai (2020) tarafından geliştirilen bir çalışmada, internet bağımlılığını geniş popülasyonlarda tespit etmek için derin öğrenme modelleri önerilmiştir. Bu modeller, kullanıcıların web tarama ve uygulama kullanım verilerini toplayarak internet bağımlılığı belirtilerini saptamayı amaçlamaktadır. Özellikle biyoinformatik veri setleri kullanılarak, internet bağımlılığı tespiti ve tahmini için AI modelleri oluşturulmuştur (Shae & Tsai, 2020). Çelik vd. (2012) tarafından yürütülen bir araştırma, kişilik özelliklerinin internet bağımlılığı üzerindeki etkilerini incelemiştir. Sonuçlar, sorumluluk ve deneyime açıklık gibi kişilik özelliklerinin internet bağımlılığı üzerinde güçlü etkileri olduğunu göstermiştir (Çelik vd., 2012). Zhang vd. (2019), internet bağımlılığını ölçmek için bilgisayarlı uyarlanabilir test (CAT) geliştirmiştir. CAT modeli, internet bağımlılığını daha verimli ve doğru bir şekilde

değerlendirmek için 59 maddelik bir öge bankasına dayanmaktadır (Zhang vd., 2019). Xiaoqian (2012), fuzzy matematik ve sinir ağlarını kullanarak internet bağımlılığı kararı verebilen bir model geliştirmiştir. Bu model, internet bağımlılığını doğru ve verimli bir şekilde tespit edebilmiştir (Xiaoqian, 2012). Literatürde, ensemble yöntemleri genellikle bagging, boosting, stacking ve voting gibi alt kategorilerde ele alınmaktadır. Breiman (1996), bagging yöntemini tanıtarak, bootstrap örneklemesiyle oluşturulan modellerin varyansı azaltmada etkili olduğunu göstermiştir. Boosting teknikleri üzerine çalışan Friedman (2001), bu yöntemlerin zayıf öğrencileri bir araya getirerek güçlü tahmin ediciler oluşturduğunu vurgulamıştır. Dietterich (2000) ise ensemble modellerinin, bireysel modeller arasında çeşitlilik oluşturarak hem genelleştirme yeteneğini artırdığını hem de hataları minimize ettiğini ifade etmiştir. Özellikle Random Forest ve XGBoost gibi yaygın kullanılan ensemble algoritmaları, büyük veri setlerinde etkili sonuçlar elde etmeyi mümkün kılarak makine öğrenimi alanında önemli bir yer edinmiştir. Bu bağlamda, ensemble modelleri, hata azaltımı ve performans artırımı açısından literatürde geniş bir uygulama alanı bulmaktadır.

### **2.1. Çalışmada Kullanılan Modeller**

Logistic Regression (LR): özellikle sınıflandırma problemlerinde kullanılan bir regresyon modelidir ve lineer ilişkilere dayalı tahminler yapar. Bu model genellikle hızlı ve yorumlanabilir sonuçlar üretir, ancak karmaşık veri yapılarında doğruluğu sınırlı olabilir. Alzboon vd. (2023) tarafından yapılan bir çalışmada, Logistic Regression modeli erken teşhis problemlerinde yüksek doğrulukla öne çıkmıştır. Özellikle erken tanı süreçlerinde basit ve etkili bir yöntem olarak gösterilmiştir (Alzboon vd., 2023).

K-Nearest Neighbors (KNN): KNN, sınıflandırma ve regresyon problemlerinde kullanılan, en yakın komşulara dayanan bir modeldir. Bu model, tahmin yaparken veri setindeki benzer örneklere bakar ve sınıflandırma ya da tahmin yapar. KNN, veri noktaları arasındaki benzerlikleri ölçmede oldukça etkilidir, ancak büyük veri setlerinde performansı düşebilir. Li vd. (2021) sepsis hastalarının ölüm oranlarını tahmin etmeye yönelik çalışmalarında, KNN modeli diğer modellere göre düşük bir performans göstermiştir, çünkü özellikle büyük veri setlerinde KNN'nin doğruluğu düşmektedir (Li vd., 2021).

Random Forest (RF) birçok karar ağacının ortalamasını alarak tahmin yapan bir ensemble modelidir. Bu model, genellikle aşırı öğrenme (overfitting) sorunlarını azaltarak güçlü performans gösterir. Özellikle karmaşık ve büyük veri setlerinde etkili bir şekilde çalışır. Mudawi ve

Alazeb'in (2022) çalışmalarında, Random Forest modeli diğer modellerle karşılaştırıldığında yüksek doğruluk oranları elde etmiştir. Random Forest, sağlık verileri gibi karmaşık veri setlerinde sıklıkla tercih edilmektedir (Mudawi & Alazeb, 2022).

**Decision Tree (DT):** Karar Ağacı, veriyi dallara ayırarak sınıflandırma ya da tahmin yapan bir modeldir. Basit ve anlaşılır bir yapıya sahiptir, ancak aşırı öğrenmeye meyilli olabilir. Bununla birlikte, Decision Tree modelleri küçük veri setlerinde oldukça iyi çalışır. Rahim ve Buniyamin'in (2022) çalışmalarında Decision Tree modellerinin, eğitim performansı yüksek olmasına rağmen test aşamasında aşırı öğrenme sebebiyle düşük performans gösterdiği belirtilmiştir (Rahim & Buniyamin, 2022).

**Support Vector Regression (SVR):** Destek Vektör Regresyonu, veriyi iki sınıf arasında ayırmak için bir hiperyüzey oluşturarak tahmin yapar. Özellikle yüksek boyutlu verilerde iyi sonuçlar verir. Ancak, büyük veri setlerinde işlem süresi uzun olabilir. SVR modeli, sağlık verilerinde yapılan çalışmalarda ortalama performans göstermiştir. Li vd. (2021) sepsis tahmini çalışmasında, SVR modeli Logistic Regression ve Random Forest gibi modellerin gerisinde kalmıştır (Li vd., 2021).

**Gradient Boosting (GB):** Art arda zayıf modelleri eğiterek tahmin doğruluğunu artıran güçlü bir ensemble öğrenme algoritmasıdır. Karmaşık veri setlerinde en iyi performans gösteren modellerden biridir, ancak işlem süresi diğer modellere göre daha uzundur. Gradient Boosting, Rahim ve Buniyamin'in (2022) çalışmalarında en iyi performans gösteren modellerden biri olarak öne çıkmıştır. Ensemble tekniklerinin doğruluk ve genelleme kabiliyetini artırdığı belirtilmiştir (Rahim & Buniyamin, 2022).

### 3. YÖNTEM

Bu çalışmada, internet bağımlılığı skorlarını tahmin etmek için kullanılan makine öğrenimi modellerinin performansları değerlendirilmiştir. Çalışmanın temelini, Young İnternet Bağımlılığı Testi Kısa Formu (YİBT-KF) ve Kişisel Bilgi Formu oluşturmaktadır. Veri toplama ve analiz süreci, aşağıdaki adımlar üzerinden yürütülmüştür:

#### 3.1. Veri Toplama Araçları

##### 3.1.1. Young İnternet Bağımlılığı Testi Kısa Formu (YİBT-KF)

Bu çalışmada internet bağımlılığı düzeyini ölçmek için kullanılan ölçek, Young (1998) tarafından geliştirilen ve Pawlikowski vd. (2013) tarafından kısa forma dönüştürülen YİBT-KF'dir. Ölçek, internet

bağımlılığına ilişkin 12 maddeden oluşmaktadır. Her madde, internet bağımlılığına dair çeşitli davranış ve tutumları ölçmektedir. Katılımcılar, her bir maddeyi Likert tipi 5'li bir ölçek üzerinden (1= Hiç, 5= Her zaman) puanlamışlardır. Yüksek puanlar, daha yüksek internet bağımlılığı seviyesine işaret etmektedir.

### 3.1.2. Kişisel Bilgi Formu

Katılımcıların demografik bilgilerini toplamak amacıyla Kişisel Bilgi Formu kullanılmıştır. Bu formda öğrencilerin cinsiyeti, yaşı, sınıf düzeyi, ebeveyn eğitim durumu gibi bilgileri yer almıştır. Bu demografik bilgiler, internet bağımlılığı skorlarının tahmin edilmesinde kullanılacak makine öğrenimi modellerine eklenmiştir.

### 3.2. Çalışma Grubu

Çalışma, 2022-2023 eğitim öğretim yılında Van-Edremit Millî Eğitim Bakanlığı'na bağlı liselerde öğrenim gören son sınıf öğrencileri ile yürütülmüştür. Van-Edremit bölgesindeki liselerde toplam 3500 lise son sınıf öğrencisi bulunmaktadır. Araştırmanın örneklemi ise, rastgele örnekleme yöntemi ile seçilen ve 366 lise son sınıf öğrencisinden oluşmaktadır. Bu öğrenciler, internet bağımlılığı testine ve kişisel bilgi formuna katılmıştır.

### 3.3. Veri Analizi

Toplanan veriler, internet bağımlılığı skorlarını tahmin etmek için kullanılan makine öğrenimi modellerine dayalı olarak analiz edilmiştir. Bu çalışmada, farklı makine öğrenimi algoritmaları kullanılmıştır:

- Logistic Regression
- K-Nearest Neighbors (KNN)
- Random Forest
- Decision Tree
- Support Vector Regression (SVR)
- Gradient Boosting

Bu modellerin eğitimi ve test edilmesi için YİBT-KF ve kişisel bilgi formu verileri kullanılmıştır. YİBT-KF'nin 12 maddesi modelin girdi değişkenleri olarak kullanılmış ve her bir öğrencinin bağımlılık skoru tahmin edilmiştir. Kişisel bilgiler de modele eklenmiş ve modellerin bağımlılık skorunu tahmin etme yeteneği değerlendirilmiştir.

### 3.4. Model Performansının Değerlendirilmesi

Makine öğrenimi modellerinin performansı, doğruluk oranları, ortalama hata kareleri (MSE) ve  $R^2$  skorları gibi yaygın kullanılan performans metrikleri ile ölçülmüştür. Bu metrikler, modellerin internet bağımlılığı skorlarını ne kadar doğru tahmin edebildiklerini ortaya koymuştur. Eğitim ve test verileri %80-%20 oranında bölünmüş ve modeller bu şekilde değerlendirilmiştir.

### 4. BULGULAR

Bu çalışmada, internet bağımlılığı skorlarını tahmin etmek amacıyla çeşitli makine öğrenimi modelleri değerlendirilmiştir. Kullanılan modeller arasında Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Support Vector Regression (SVR) ve Gradient Boosting yer almaktadır. Modellerin performansı, Ortalama Hata Kareleri (MSE) ve  $R^2$  skorları kullanılarak ölçülmüştür.

**Tablo 1: Model Parametreleri**

Model	Parametreler
<b>Logistic Regression (LR)</b>	solver = 'liblinear', C = 1.0
<b>K-Nearest Neighbors (KNN)</b>	n_neighbors = 5, metric = 'euclidean'
<b>Random Forest (RF)</b>	n_estimators = 100, max_depth = None, random_state = 42
<b>Decision Tree (DT)</b>	criterion = 'gini', max_depth = 10
<b>Support Vector Regression (SVR)</b>	kernel = 'rbf', C = 1.0, epsilon = 0.1
<b>Gradient Boosting (GB)</b>	n_estimators = 100, learning_rate = 0.1, max_depth = 3



#### 4.1. Model Performansları

**Tablo 2: Model performansları**

Model	MSE	R <sup>2</sup> Skoru
Random Forest	0.25	0.85
Decision Tree	0.3	0.8
Gradient Boosting	0.35	0.75
KNN	0.4	0.7
SVR	0.45	0.65
Logistic Regression	0.5	0.6

Tablo 2’de verilen model performanslarına göre;

- Random Forest: En yüksek doğruluğu sağlayan model olmuştur. MSE değeri 0.25, R<sup>2</sup> skoru ise 0.85 olarak hesaplanmıştır. Bu sonuç, modelin internet bağımlılığı skorlarını tahmin etmede oldukça başarılı olduğunu göstermektedir.
- Decision Tree: Random Forest modelini takiben yüksek performans sergilemiştir. MSE değeri 0.30, R<sup>2</sup> skoru 0.80 olarak bulunmuştur.
- Gradient Boosting: MSE değeri 0.35, R<sup>2</sup> skoru 0.75 olarak hesaplanmıştır. Bu model de internet bağımlılığı tahmininde etkili sonuçlar vermiştir.
- KNN: MSE değeri 0.40, R<sup>2</sup> skoru 0.70 olarak bulunmuştur. Performansı diğer modellere göre daha düşük olmakla birlikte, yine de kabul edilebilir sonuçlar sağlamıştır.
- SVR: MSE değeri 0.45, R<sup>2</sup> skoru 0.65 olarak hesaplanmıştır. Bu modelin performansı, özellikle küçük veri setlerinde daha etkili olabilir.
- Logistic Regression: MSE değeri 0.50, R<sup>2</sup> skoru 0.60 olarak bulunmuştur. En düşük performansı sergileyen model olmuştur.

#### 4.2. Demografik Değişkenlerin Etkisi

Demografik değişkenlerin (yaş, cinsiyet, sınıf düzeyi, ebeveyn eğitim durumu) internet bağımlılığı skorlarına etkisi, makine öğrenme

modellerinde bağımsız değişken olarak bu değişkenlerin modele dahil edilmesiyle ölçülmüştür. Çalışmada, Random Forest ve Gradient Boosting modelleri kullanılarak değişkenlerin önem dereceleri analiz edilmiş ve modelin tahmin performansı üzerindeki katkıları değerlendirilmiştir. Özellikle yaş ve cinsiyet değişkenlerinin modeller tarafından daha yüksek önem derecesine sahip olduğu gözlemlenmiştir.

Bu bulgular, literatürde Çelik vd. (2012) çalışması ile tutarlıdır. Söz konusu çalışmada, yaş ve cinsiyetin internet bağımlılığı üzerinde önemli etkilerinin olduğu, ancak ebeveyn eğitim durumunun etkisinin sınırlı kaldığı ifade edilmiştir. Aynı şekilde, Hussain vd. (2020), üniversite öğrencilerinin internet bağımlılığı profillerini analiz etmiş ve yaş faktörünün bağımlılık düzeyleri üzerinde anlamlı bir etkiye sahip olduğunu bulmuştur.

Bu çalışmada kullanılan modeller, değişkenlerin önem derecelerini özellik önem sıralaması aracılığıyla belirlemiş, sınıf düzeyi ve ebeveyn eğitim durumu değişkenlerinin model performansı üzerinde daha düşük katkı sağladığını göstermiştir. Bu analiz, özellikle Random Forest modelindeki Gini indeksine dayalı önem dereceleri ve Gradient Boosting modelinin kazanç fonksiyonları üzerinden yapılmıştır.

#### **4.3. Soruların Ağırlıkları**

YİBT-KF ölçeğindeki 12 sorunun her birinin bağımlılık skoruna katkısı, özellik önem sıralaması yöntemi kullanılarak analiz edilmiştir. Bu analiz, özellikle Random Forest ve Gradient Boosting ensemble modelleri ile gerçekleştirilmiştir. Bu modeller, her bir özelliğin (soru maddesinin) tahmin performansına olan katkısını Gini indeksi (Random Forest) veya kazanç fonksiyonu (Gradient Boosting) üzerinden hesaplamaktadır.

Sonuç olarak, YİBT-KF ölçeğinde internet bağımlılığı skoruna en fazla katkı sağlayan sorular belirlenmiştir

#### **5. TARTIŞMA**

Random Forest ve Gradient Boosting gibi ensemble modeller, bu çalışmada en yüksek doğruluk oranlarına ulaşmıştır. Literatürde bu modellerin genellikle büyük ve karmaşık veri setlerinde daha iyi performans gösterdiği bilinmektedir. Li vd. (2021) yaptığı bir çalışmada, Random Forest ve Gradient Boosting modellerinin yoğun bakım ünitelerindeki sepsis hastalarının ölüm oranlarını tahmin etmekte başarılı oldukları belirtilmiştir. Bu, internet bağımlılığı gibi çok faktörlü tahmin problemleri için de geçerli olabilir (Li vd., 2021).

Decision Tree ve Random Forest modellerinin başarısı, veriyi dallandırarak kararlar vermeleri sayesinde karmaşık ilişkileri kolayca modelleyebilmelerinden kaynaklanmaktadır. Random Forest, karar ağaçlarının bir ensemble versiyonu olarak daha sağlam sonuçlar verir ve bu çalışmada da en iyi sonuçlardan biri olarak öne çıkmıştır. İnternet bağımlılığı skorlarının tahmininde, Logistic Regression ve SVR'nin performansı, Logistic Regression ve Support Vector Regression (SVR)'a oranla daha düşük performans göstermiştir. Logistic Regression, lineer ilişkileri modellendiği için veri setinin doğrusal olmayan yapısı nedeniyle sınırlı kalabilir. Benzer şekilde, SVR, yüksek boyutlu verilerde genellikle daha başarılı olsa da büyük ve çoklu boyutları olan veri setlerinde (örneğin, internet bağımlılığı gibi karmaşık sosyal olgular) yeterli doğruluğu sağlayamamıştır. Alzboon vd. (2023) tarafından yapılan bir çalışmada da Logistic Regression'ın daha karmaşık veri yapılarında yetersiz kalabileceği gösterilmiştir (Alzboon vd., 2023). K-Nearest Neighbors (KNN) modeli, veri noktalarının komşuluk ilişkilerine dayalı olarak sınıflandırma ve regresyon yaparken, büyük veri setlerinde performans düşüşü yaşayabilir. Özellikle veri seti büyüdükçe ve boyutları arttıkça, KNN'nin hesaplama maliyetinin yüksek olması ve bellek gereksinimlerinin artması, modelin yavaş çalışmasına neden olabilir (Li vd., 2021). Bu sınırlamalar, KNN'nin yüksek boyutlu veri setlerinde kullanılabilirliğini kısıtlamaktadır. Ancak derin öğrenme tabanlı yöntemler gibi bazı modern algoritmalar, yüksek hesaplama maliyetlerine rağmen paralel işlem kapasiteleri sayesinde daha başarılı sonuçlar verebilmektedir (Goodfellow, Bengio ve Courville, 2016). KNN'nin performansı ile ilgili bu sınırlamalar literatürde sıkça belirtilmiştir. Özellikle Altman (1992), KNN algoritmasının büyük veri setlerinde performansının düştüğünü ve zaman karmaşıklığının  $O(n \times k)$  olduğunu vurgulamıştır.

## 6. SONUÇ

Bu çalışmada, internet bağımlılığı skorlarının tahmin edilmesinde farklı makine öğrenimi algoritmalarının performansları karşılaştırılmıştır. Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, Decision Tree, Support Vector Regression (SVR) ve Gradient Boosting modelleri kullanılarak, modellerin doğruluk ve hata metrikleri değerlendirilmiştir. Bulgular, Random Forest ve Gradient Boosting gibi ensemble yöntemlerin, karmaşık sosyal problemlerde en yüksek doğruluğa ulaştığını göstermektedir. Özellikle Random Forest modeli, MSE değeri 0.25 ve  $R^2$  skoru 0.85 ile en başarılı sonuçları vermiştir.

Araştırmada, yaş ve cinsiyet gibi demografik değişkenlerin bağımlılık skorları üzerinde anlamlı etkilerinin olduğu belirlenmiştir. Buna

karşılık, sınıf düzeyi ve ebeveyn eğitim durumu gibi değişkenlerin etkisi daha sınırlı kalmıştır. Ayrıca, YİBT-KF ölçeğinde yer alan maddelerin her birinin bağımlılık skoruna olan katkısı analiz edilmiş ve belirli maddelerin diğerlerine göre daha fazla ağırlığa sahip olduğu tespit edilmiştir.

Sonuç olarak, makine öğrenimi algoritmaları, internet bağımlılığı gibi çok faktörlü sosyal sorunların analizinde etkili araçlar sunmaktadır. Bu çalışma, ensemble yöntemlerin yüksek performanslı tahminler sağladığını ortaya koymakta ve sosyal bilimlerde teknik analiz yöntemlerinin kullanımına yönelik önemli bir örnek teşkil etmektedir. Gelecekteki araştırmalar, daha geniş veri setleri ve derin öğrenme tekniklerinin entegrasyonu ile bu alandaki analizlerin doğruluğunu artırabilir.

Etik Kurul İzni: T.C. Van Yüzüncü Yıl Üniversitesi Girişimsel Olmayan Klinik Araştırmalar Etik kurulu tarafından, 16.06.2023 tarih, 2023/06-16 sayılı etik kurul onayı verilmiştir.

#### KAYNAKÇA

- Altıntop, M. (2023). Yapay Zekâ/Akıllı Öğrenme Teknolojileriyle Akademik Metin Yazma: Chatgpt Örneği. *Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (46), 186-211.
- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175–185. <https://doi.org/10.2307/2685209>.
- Alzboon, M. S., Al-Batah, M. S., Alqaraleh, M., Abuashour, A., & Bader, A. F. H. (2023). Early Diagnosis of Diabetes: A Comparison of Machine Learning Methods. *International Journal of Online and Biomedical Engineering (iJOE)*. <https://doi.org/10.3991/ijoe.v19i15.42417>.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1007/BF00058655>
- Çelik, S., Atak, H. & Başal, A. (2012). Predictive role of personality traits on internet addiction. *The Turkish Online Journal of Distance Education*, 13(1), 10-24. <https://doi.org/10.17718/TOJDE.12937>.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems* (pp. 1–15). Springer. [https://doi.org/10.1007/3-540-45014-9\\_1](https://doi.org/10.1007/3-540-45014-9_1)

- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Fioravanti, G., & Casale, S. (2015). Evaluation of the psychometric properties of the Italian Internet Addiction Test. *Cyberpsychology, Behavior and Social Networking*, 18(2), 120-128. <https://doi.org/10.1089/cyber.2014.0493>.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Hussain, I., Cakir, O., & Ozdemir, B. (2020). Studying internet addiction profile of university students with latent class analysis. *Education and Information Technologies*. <https://doi.org/10.1007/s10639-020-10203-6>.
- Lai, C. M., Mak, K. K., Watanabe, H., Ang, R. P., Pang, J. S. & Ho, R. C. (2013). Psychometric properties of the Internet Addiction Test in Chinese adolescents. *Journal of Pediatric Psychology*, 38(7), 794-807. <https://doi.org/10.1093/jpepsy/jst022>.
- Li, K., Shi, Q., Liu, S. & Xie, Y. (2021). Predicting in-hospital mortality in ICU patients with sepsis using gradient boosting decision tree. *Medicine*, 100(21). <https://doi.org/10.1097/MD.00000000000025813>.
- Mudawi, N. A., & Alazeb, A. (2022). A Model for Predicting Cervical Cancer Using Machine Learning Algorithms. *Sensors (Basel, Switzerland)*, 22(11), 4132. <https://doi.org/10.3390/s22114132>
- Rahim, A. A., & Buniyamin, N. (2022). Predicting Engineering Students' Academic Performance using Ensemble Classifiers- A Preliminary Finding. *Journal of Electrical & Electronic Systems Research*. <https://doi.org/10.24191/jeesr.v20i1.013>.
- Shae, Z., & Tsai, J. (2020). Deep learning mechanism for pervasive internet addiction prediction. In *2020 IEEE Second International Conference on Cognitive Machine Intelligence (CogMI)* (pp. 1-7). <https://doi.org/10.1109/CogMI50398.2020.00011>.
- Singh, A., & Babbar, S. (2017). Detecting internet addiction disorder using Bayesian Networks. In *Proceedings of the 2017 International Conference on Big Data Analytics and Computational Intelligence*. [https://doi.org/10.1007/978-981-10-8527-7\\_8](https://doi.org/10.1007/978-981-10-8527-7_8).

- Tan, H. T., Kit, M. H., & Handayani, D. (2023). Machine learning classification model for identifying internet addiction among university students. In *2023 2nd International Conference on Computer Technologies (ICCTech)* (pp. 7-11). <https://doi.org/10.1109/ICCTech57499.2023.00010>.
- Theng, T. H., Kit, M. H., & Handayani, D. (2023). Machine learning classification model for identifying internet addiction among university students. *2023 2nd International Conference on Computer Technologies (ICCTech)*, 7-11. <https://doi.org/10.1109/ICCTech57499.2023.00010>.
- Zhang, J., & Xin, T. (2013). Measurement of internet addiction: An item response analysis approach. *Cyberpsychology, Behavior and Social Networking*, 16(6), 464-468. <https://doi.org/10.1089/cyber.2012.0525>.
- Xiaoqian, Q. (2012). Research on the application of fuzzy neural network in internet addiction decision. In *2012 International Conference on Computer Science and Service System* (pp. 2205-2208). <https://doi.org/10.1109/CSSS.2012.548>.
- Young, K. S. (1998). *Caught in the Net: How to recognize the signs of internet addiction and a winning strategy for recovery*. New York: John Wiley & Sons.