

**T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**ELİT VE ELİT OLMAYA ADAY SPORCULARIN
SINIFLANDIRILMASINDA KULLANILAN VERİ
MADENCİLİĞİ YÖNTEMLERİNİN PERFORMANSLARININ
İNCELENMESİ**

Adeviye Gamze AKIN

**Biyoistatistik Programı
YÜKSEK LİSANS TEZİ**

**ANKARA
2025**

**T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ**

**ELİT VE ELİT OLMAYA ADAY SPORCULARIN
SINIFLANDIRILMASINDA KULLANILAN VERİ
MADENCİLİĞİ YÖNTEMLERİNİN PERFORMANSLARININ
İNCELENMESİ**

Adeviye Gamze AKIN

**Biyoistatistik Programı
YÜKSEK LİSANS TEZİ**

**TEZ DANIŞMANI
Prof. Dr. Erdem KARABULUT**

**ANKARA
2025**

TEZ ONAY SAYFASI

**ELİT VE ELİT OLMAYA ADAY SPORCULARIN SINIFLANDIRILMASINDA KULLANILAN VERİ
MADENCİLİĞİ YÖNTEMLERİNİN PERFORMANSLARININ İNCELENMESİ**

Adeviye Gamze AKIN

Danışman: Prof. Dr. Erdem KARABULUT

Bu tez çalışması 26/05/2025 tarihinde jürimiz tarafından "Biyoistatistik Programı" nda yüksek lisans tezi olarak kabul edilmiştir.

Jüri Başkanı:	Prof. Dr. Gökhan DELİCEOĞLU <i>Gazi Üniversitesi</i>
Tez Danışmanı:	Prof. Dr. Erdem KARABULUT <i>Hacettepe Üniversitesi</i>
Üye:	Prof. Dr. Atilla Halil ELHAN <i>Ankara Üniversitesi</i>
Üye:	Dr. Öğr. Üyesi H. Yağmur ZENGİN <i>Hacettepe Üniversitesi</i>
Üye:	Dr. Öğr. Üyesi Ebru ÖZTÜRK <i>Hacettepe Üniversitesi</i>

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin ilgili maddeleri uyarınca yukarıdaki jüri tarafından uygun bulunmuştur.

11 8 Haziran 2025

Prof. Dr. Müge YEMİŞÇİ ÖZKAN

Enstitü Müdürü

YAYIMLAMA ve FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan **“Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”** kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir. ⁽¹⁾
- Enstitü / Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren... ay ertelenmiştir. ⁽²⁾
- Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

18.06./2025

Adeviye Gamze AKIN

¹ “Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge”

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez **danışmanın** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulu** iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6. 2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez **danışmanın** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulunun** gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, **tezin yapıldığı kurum** tarafından verilir *. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, **ilgili kurum ve kuruluşun önerisi** ile **enstitü** veya **fakültenin** uygun görüşü üzerine **üniversite yönetim kurulu** tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.
Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

* Tez **danışmanın** önerisi ve **enstitü anabilim dalının** uygun görüşü üzerine **enstitü** veya **fakülte yönetim kurulu tarafından karar verilir.**

ETİK BEYANI

Bu çalışmadaki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi, görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu, kullandığım verilerde herhangi bir tahrifat yapmadığımı, yararlandığım kaynaklara bilimsel normlara uygun olarak atıfta bulunduğumu, tezimin kaynak gösterilen durumlar dışında özgün olduğunu, Prof. Dr. Erdem KARABULUT danışmanlığında tarafımdan üretildiğini ve Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazım Yönergesine göre yazıldığımı beyan ederim.

Adeviye Gamze AKIN

TEŞEKKÜR

Bu tez çalışmasında değerli bilgileri, görüş ve önerileriyle bana yol gösteren kıymetli danışman hocam Prof. Dr. Erdem KARABULUT'a,

Biyoistatistik Ana Bilim Dalı'nda ve tez savunma jürimde bulunan tüm hocalarıma, çalışmamda destek veren kıymetli arkadaşım Hüseyin Erkin SÜLEKLİ'ye ve bölüm sekreterimiz kıymetli ablam Menekşe TARLA'ya,

Bilgisini, tecrübesini, rehberliğini ve yardımını esirgemeyen kıymetli hocam Prof. Dr. Gökhan DELİCEOĞLU'na, Sporcu Sağlığı ve Performansı ekibi çalışma arkadaşlarıma ve hocalarıma,

Her zaman desteği ile yanımda olan kıymetli eşim Ahmet Sinan AKIN'a, canım oğlum Ertuğrul'a, canlarım Annem ve Babam'a teşekkürlerimi sunarım.

Adeviye Gamze AKIN

ÖZET

Akın, A. G., Elit ve Elit Olmaya Aday Sporcuların Sınıflandırılmasında Kullanılan Veri Madenciliği Yöntemlerinin Performanslarının İncelenmesi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Programı Yüksek Lisans Tezi, Ankara, 2025. Bu tez çalışmasında Elit ve Elit olmaya aday sporcuların sınıflandırılması amacıyla veri madenciliği sınıflama modelleri kullanılmıştır. Sınıflama amacıyla 66 elit, 122 elit olmaya aday toplam 188 sporcunun performans test verisi kullanılarak farklı atama yöntemleri ile veri tamamlama işlemleri gerçekleştirilmiştir. SMOTE dengeleme algoritması ile sınıf oranları dengelenmiştir. Rastgele Orman, XGBoost, AdaBag, KNN, Lojistik Regresyon Modeli, Karar Ağaçları (CART, CART Bagging, C4.5), Naive Bayes, Destek Vektör Makineleri olmak üzere 10 farklı sınıflandırma modeli; aritmetik ortalama, Rastgele Orman algoritması ve KNN algoritması ile eksik gözlem atanan 4 farklı farklı senaryoda çalışılmıştır. Tüm senaryolarda en iyi tahmin performansı gösteren model olarak Rastgele Orman ve XGBoost modeli bulunmuştur. Ayrıca, en iyi tahmin performansı gösteren sınıflama modelinde değişken önemliliklerine göre elit sporcu olmaya etki eden en önemli performans bileşenleri belirlenmiştir. Bu bileşenler, sporcunun yaşı, sporcunun ölçüm tarihi itibarıyla aktif olarak spor yaptığı yıl sayısı, çok yönlü reaktif çeviklik testi bitirme süresi, izometrik orta uyluk çekiş testi ve alt ekstremite wingate anaerobik güç testi yorgunluk yüzdesi, üst ekstremite anaerobik güç testi ortalama güç, aerobik güç ve kapasite testinde anaerobik eşikteki hızı ve maksimum hız elitlik kategorisi tahminine öne çıkan en önemli kriterler olarak belirlenmiştir. Sporcunun yaşı ve aktif olarak spor yaptığı yıl sayısı tüm senaryolarda en önemli kriter olarak öne çıkması sebebiyle, son aşamada bu iki değişken analiz dışında bırakılıp tahmin modelleri uygulanmış ve yaş ve spor yılından sonra gelen değişkenlerin önem derecelerinin aynı sırayla olmasa da ilk 5'te yer aldığı görülmüştür.

Anahtar Kelimeler: Sportif performans, sınıflandırma, optimizasyon

ABSTRACT

Akın, A. G., Examining the Performance of Data Mining Methods in Classifying Elite and Elite Candidate Athletes, Master's Thesis, Hacettepe University, Graduate School of Health Sciences, Biostatistics Program, Ankara, 2025. In this thesis, data mining classification models were used to classify elite and candidate elite athletes. For classification purposes, data completion processes were performed with different assignment methods using the performance test data of a total of 188 athletes, 66 elite and 122 candidate elite athletes. Class ratios were balanced with SMOTE balancing algorithm. Ten different classification models including Random Forest, XGBoost, AdaBag, KNN, Logistic Regression Model, Decision Trees (CART, CART Bagging, C4.5), Naive Bayes, Support Vector Machines; arithmetic mean, Random forest algorithm and KNN algorithm were used in 4 different scenarios with missing observations. In all scenarios, Random Forest and XGBoost models were found to have the best prediction performance. In addition, in the classification model with the best prediction performance, the most important performance components that affect being an elite athlete were determined according to variable importance. These components, the age of the athlete, the number of years the athlete has been active in sports as of the measurement date, the completion time of the multidirectional reactive agility test, isometric mid-thigh traction strength, fatigue percentage in the lower extremity wingate anaerobic power test, average power in the upper extremity anaerobic power test, speed at anaerobic threshold and maximum speed in the aerobic power and capacity test were determined as the most important criteria for predicting the elite category. As the age of the athlete and the number of years of active sport were the most important criteria in all scenarios, these two variables were excluded from the analysis in the last stage and prediction models were applied and it was seen that the importance of the variables following age and years of sport were in the top 5, although not in the same order.

Key Words: sports performance, classification, optimization

İÇİNDEKİLER

TEZ ONAY SAYFASI	iii
YAYIMLAMA ve FİKRİ MÜLKİYET HAKLARI BEYANI	iv
ETİK BEYANI	v
TEŞEKKÜR	vi
ÖZET	vii
ABSTRACT	viii
İÇİNDEKİLER	ix
SİMGELER ve KISALTMALAR	xii
ŞEKİLLER	xiv
TABLolar	xv
1. GİRİŞ	1
2. GENEL BİLGİLER	3
2.1. Sportif Performans	3
2.2. Sportif Performans Parametreleri	3
2.3. Sportif Performansı Etkileyen Faktörler	4
2.3.1. İçsel Faktörler	4
2.3.2. Dışsal Faktörler	5
2.4. Performansın Değerlendirilmesinde Kullanılan Yöntemler	6
2.4.1. Ölçümler	7
2.4.2. Performans Testleri	8
2.4.3. Temel Verimlilik Testleri	9
2.4.4. Metabolik Testler	9
2.4.5. Psikolojik Durum Değerlendirme Testleri	10
2.5. Çalışmada Kullanılan Performans Testleri	10
2.5.1. İzometrik Orta Uyluk Çekiş Testi (IMTP)	10
2.5.2. Wingate Anaerobik Güç ve Kapasite Testi	12
2.5.3. Aerobik Güç ve Kapasite Testi	15
2.5.4. Çok Yönlü Reaktif Çeviklik Testi	18
2.6. Elit Sporcu Kavramı	20

2.6.1. Judo ve Güreş Sporcularında Elit Sporcu Kavramı	20
2.6.2. Elit Olma Süreci	21
2.7. Veri Madenciliği	22
2.8. Makine Öğrenmesi	24
2.8.1. Danışmansız Öğrenme	24
2.8.2. Danışmanlı Öğrenme	24
2.9. Veri Madenciliği Yöntemleri	25
2.9.1. Tanımlayıcı Yöntemler	25
2.9.2. Kestirici Yöntemler	25
2.10. Sınıflama/Regresyon Yöntemleri	25
2.11. Karar Ağacı Yöntemleri	26
2.11.1. Bölünme Kriterleri	27
2.12. C4.5 Modeli	29
2.13. CART Modeli	29
2.14. Topluluk (Ensemble) Yöntemleri	30
2.15. XGBoost	31
2.16. AdaBag (Adaptive Bagging)	32
2.17. Rastgele Orman Modeli (RO)	32
2.18. Lojistik Regresyon Modeli (LRM)	33
2.19. Destek Vektör Makineleri (DVM)	34
2.20. Naive Bayes Modeli (NB)	35
2.21. K-En Yakın Komşu (KNN)	36
2.22. Sınıf Dengesizliği	36
2.22.1. SMOTE (Synthetic Minority Over-sampling Technique)	37
2.23. Model Performans Değerlendirme Ölçütleri	37
2.23.1. Sınıflama Modelleri Performans Değerlendirme Ölçütleri	37
2.24. Spor Bilimlerinde Veri Madenciliği Uygulamaları	41
3. GEREÇ ve YÖNTEM	45
4. BULGULAR	51
5. TARTIŞMA	67
6. SONUÇ ve ÖNERİLER	75
7. KAYNAKLAR	77

8. EKLER	86
EK-1: Hacettepe Üniversitesi Etik Komisyon İzni	
EK-2: Turnitin Orijinallik Raporu Ekran Görüntüsü	
EK-3: Dijital Makbuz	
9. ÖZGEÇMİŞ	89



SİMGELER ve KISALTMALAR

IRM	Maksimum Kuvvet Testleri
ACE	Anjiyotensin Dönüştürücü Enzim
ACTN3	Alfa-Aktinin-3
AP	Ortalama Güç
APR	Kilogram Başına Ortalama Güç
AT HIZ	Anaerobik Eşiğe Ulaşıldığında Ölçülen Hız
CART	Sınıflandırma Ve Regresyon Ağaçları
DVM	Destek Vektör Makineleri
EAA	Eğri Altında Kalan Alan
IJF	Judo Branşında Sporcuların Uluslararası Sıralaması
IMTP	İzometrik Orta Uyluk Çekiş Testi
KNN	K En Yakın Komşuluk
LRM	Lojistik Regresyon Modeli
MAX KAH	Maksimum Kalp Atım Hızı
MAXVO₂	Maksimum Oksijen Tüketimi
MCAR	Tamamen Rastgele Kayıp
ML	Makine Öğrenme
MLP	Çok Katmanlı Algılayıcı
MP	Test Boyunca Ulaşılan En Düşük Güç
MRI	Manyetik Rezonans Görüntüleme
MSE	Hata Kareler Ortalaması
NB	Naive Bayes
NBA	Amerika Merkezli Profesyonel Lig
NN	Yapay Sinir Ağları
PD(%)	Güç Kaybı Oranı Yüzdesi.
PP	Zirve Güç
PPR	Kilogram Başına Zirve Güç
RO	Rastgele Orman
ROC	Receiver Operating Characteristic Curve (Alıcı İşletim Özelliği Eğrisi)

RQ	Karbondioksit Üretim Miktarının Oksijen Tüketimine Oranı
SMOTE	Sentetik Azınlık Aşırı Örnekleme Tekniđi
TGF	Türkiye Güreş Federasyonu
TJF	Türkiye Judo Federasyonu
tPP	Zirve Güce Ulaşma Süresi
U15	15 Yaş ve Altı
U17	17 Yaş ve Altı
U18	18 Yaş ve Altı
U20	20 Yaş ve Altı
U21	21 Yaş ve Altı
VCO2	Dakikada Üretilen Karbondioksit Miktarı
VE	Ventilasyon
VO2	Dakikada Tüketilen Oksijen Miktarı
WAnT	Wingate Anaerobik Güç Testi

ŞEKİLLER

Şekil	Sayfa
2.1. Sportif performans ölçümleri, sağlık taramaları ve rehabilitasyon hizmetleri.	7
2.2. IMTP testinin uygulanaşına ait görsel.	11
2.3. Üst ekstremite Wingate anaerobik güç ve kapasite testine ait görsel.	13
2.4. Alt ekstremite Wingate anaerobik güç ve kapasite testine ait görsel.	14
2.5. Aerobik güç ve kapasite testi koşu ergometresine ait bir görsel.	16
2.6. Aerobik güç ve kapasite testi veri akışına ait bir görsel.	17
2.7. Aerobik güç ve kapasite testi bisiklet ergometresine ait bir görsel.	17
2.8. Çok yönlü reaktif çeviklik testine ait bir görsel.	19
2.9. Örnek bir karar ağacı yapısı.	26
2.10. Veri setlerinin doğrusal olma durumlarına göre çizilen karar sınırları.	34
2.11. Veri madenciliğinin sporda kullanım alanları.	42
4.1. Rastgele Orman modeline göre değışkenlerin önem ağırlıkları (eksik gözlemlere aritmetik ortalama ile atama yapılan veri setinde).	56
4.2. XGBoost modeline göre değışkenlerin önem ağırlıkları (eksik gözlemlere RO algoritması ile atama yapılan veri setinde).	58
4.3. RO modeline göre değışkenlerin önem dereceleri (RO+SMOTE uygulanan veri setinde).	61
4.4. RO modeline göre yaş ve aktif spor yaptığı yıl sayısı değışkenleri çıkartılarak elde edilen faktörlerin önem dereceleri (RO+SMOTE uygulanan veri setinde).	62
4.5. Uygulanan senaryolarda dengeli doğruluk ölçütüne göre modellerin karşılaştırılması.	63
4.6. Uygulanan senaryolarda F1 ölçütüne göre modellerin karşılaştırılması.	65

TABLOLAR

Tablo	Sayfa
2.1. DVM çekirdek fonksiyonları (88).	35
2.2. İki sınıf içeren veri setine ait 2x2 çapraz tablo	38
3.1. Araştırmada kullanılan değişkenler ve açıklamaları.	46
3.2. Analizden dışlanan/dahil edilen değişkenler ve korelasyon katsayıları	47
4.1. Sporcu verilerine ait tanımlayıcı istatistikler.	51
4.2. Branşlara göre sporcu verilerine ait tanımlayıcı istatistikler.	52
4.3. Eksik gözlemlere aritmetik ortalama ile atama yapılan veri setinden elde edilen sınıflama modellerine ait çapraz tablo.	53
4.4. Elit olma sınıflama modellerine ait test performans ölçütleri (eksik gözlemlere aritmetik ortalama ile atama yapılan veri setinde).	54
4.5. Elit olma sınıflama modellerine ait test performans ölçütleri (eksik gözlemlere RO algoritması ile atama yapılan veri setinde).	57
4.6. Elit olma sınıflama modellerine ait test performans ölçütleri (eksik gözlemlere KNN algoritması ile atama yapılan veri setinde).	59
4.7. Eksik gözlemlere RO algoritması ile atama yapılan veri setinde SMOTE ile dengeleme yapıldıktan sonra elde edilen sınıflama modellerine ait test performans ölçütleri.	60
4.8. Modellerin genel performans karşılaştırması.	63
4.9. Modellerin uygulanan senaryoya bağlı performans karşılaştırması.	63
4.10. Uygulanan senaryolarda dengeli doğruluk ölçütüne göre modellerin karşılaştırılması.	64
4.11. Uygulanan senaryolara göre modellerin dengeli F1 ölçülerinin karşılaştırılması.	65

1. GİRİŞ

Sporcuların atletik performanslarını belirlemeye yönelik yapılan performans test verileri, sporculara, antrenörlere ve araştırmacılara yol gösterecek değerli bilgiler sunmaktadır. Performans bileşenlerinden; hız, kuvvet, çeviklik, denge, güç ve bununla ilişkilendirilen motor becerileri gibi birçok faktör performans sergilemede önem arz etmekte ve sporcuların performans sonuçlarının değerlendirilmesinde birçok analiz yöntemi kullanılmaktadır. Sporcu verilerinden veri madenciliği yöntemleri ile anlamlı bilgiler elde edilerek özellikle sporcuların seçimi, sınıflandırılması, desteklenmesi ve sporcuların performans gelişimi takibinde yer alan karar vericilerin öznellikten uzaklaşarak daha etkin kararlar almasını sağlayacaktır.

Son yıllarda spor alanında veri madenciliği yöntemlerinin uygulanması yaygın hale gelmiştir. Spor bilimlerinde veri madenciliği yöntemleri, sakatlık risk analizi (1), bireysel ve takım performans izlemi ve ileriye yönelik performans tahmini (2,3), skor tahmini (4), sporcu seçimi (5), kariyer yörüngesi tahmini (6) ve taktik oluşturma (7) konularında karar verme süreçlerinde yardımcı olmaktadır. Özellikle yurt dışındaki profesyonel spor kulüpleri; potansiyel yetenekleri keşfetmek, mevcut sporcuların performanslarındaki eksiklikleri belirlemek ve rakip takımların analizini gerçekleştirmek amacıyla istatistikçi ve performans analistleri istihdam etmektedir (8). Diğer yandan spora başlayacak bireylerin, öznel yöntemler yerine, yetenek taraması testlerinden edinilen verilerinin veri madenciliği ile anlamlandırılması ve bireylerin uygun oldukları branşlara yönlendirilmesi mümkündür. Bu kapsamda çalışmamızın amacı judo ve güreş branşında elit ve elit olma potansiyeli taşıyan sporcuların sınıflandırılmasında veri madenciliği yöntemlerinin uygulanması hedeflenmekte ve veri madenciliği yöntemleri kullanılarak, elit ve elit olmaya aday sporcuların performans değişkenlerine göre bir sınıflandırma algoritması oluşturmaktır.

Bu amaçla öncelikle, performans testlerine katılan sporcuların performans test ölçüm sonuçlarına bir sınıflandırma problemi açısından bakılmıştır. Veri madenciliğinde kullanılan sınıflandırma algoritmalarından; Rastgele Orman (RO), Extereme Gradient Boosting (XGBoost), Adaptive Bagging (AdaBag), K-Nearest Neighbors (KNN), Lojistik Regresyon Modeli (LR), Karar Ağaçları (CART, CART

Bagging, C4.5), Naive Bayes (NB), Destek Vektör Makineleri (DVM) yöntemleri ile 10 farklı model oluşturulmuştur. Eksik gözlem ataması için grup ortalaması, Rastgele Orman ve KNN algoritmalarından faydalanılmıştır. Sınıf dengesizliği analizi yapılarak en yüksek sınıflama performansını gösteren modeller ile sporcuları ayırt etmede uygulanan performans test sonuçlarının önem ağırlıkları sıralanmıştır.

Bu bölümde, tez çalışmasının amacı ve kapsamı ele alınmaktadır. Genel Bilgiler bölümünde, ilgili literatür örnekleriyle birlikte sporcu performans ölçüm yöntemleri, yaygın olarak kullanılan testler ve çalışmada uygulanan veri madenciliği teknikleri hakkında bilgi verilmiştir. Gereç ve Yöntem bölümünde, araştırmada kullanılan değişkenler ayrıntılı biçimde tanımlanmış ve analiz sürecine ilişkin yöntemsel açıklamalara yer verilmiştir. Bulgular bölümünde, performans testlerine katılan sporculara yönelik sınıflandırma analizleri, geliştirilen modellerin performansları ve test verilerine ilişkin değişken önem düzeyleri sunulmuştur. Tartışma bölümünde, elde edilen bulgular benzer çalışmalar ve elit sporcu seçim kriterleri ile karşılaştırılarak değerlendirilmiş; Sonuç ve Öneriler bölümünde ise çalışmanın ortaya koyduğu temel sonuçlar ve ileriye dönük öneriler aktarılmıştır.

2. GENEL BİLGİLER

2.1. Sportif Performans

Sportif performans, yapılan bir görevin yerine getirilmesinde maksimum başarıyı hedefleyen tüm çabaların birleşimi olarak tanımlanır. Bu performans, bireyin sosyolojik, psikolojik, fizyolojik, motorsal ve fiziksel becerileri gibi çeşitli değişkenlerden etkilenir. Her bir faktör, bireyin genel sportif başarısına farklı yönlerden katkı sağlar, bu yüzden çok yönlü bir antrenman programı ve her bir parametrenin uyumlu bir şekilde geliştirilmesi gereklidir (9).

2.2. Sportif Performans Parametreleri

Sportif performans parametreleri dört ana başlık altında incelenebilir:

- **Sosyolojik Faktörler:** Sosyal destek, aile ve arkadaş çevresi gibi dış sistemlerle birlikte, sporcunun sosyal çevresi de performansı etkiler. Özellikle rekabet ve sosyal çevre desteği, sporcunun stresle başa çıkma yeteneğini güçlendirebilir (10).

- **Psikolojik Faktörler:** Spor performansını belirleyen önemli psikolojik etkenler arasında motivasyon, özgüven, kaygı ve stres gibi unsurlar bulunur. Özellikle özgüvenin yüksek olması, sporcunun başarıya ulaşmasında kritik bir rol oynamaktadır (11). Psikolojik faktörlerin bireyin spor performansı üzerindeki etkisi büyük ölçüde çalışılmıştır.

- Örneğin, sporcular arasındaki psikolojik dayanıklılık farklılıkları, yüksek başarıya ulaşmada belirleyici bir etken olarak öne çıkmaktadır (12).

- **Fizyolojik Faktörler:** Aerobik kapasite, anaerobik güç, kardiyovasküler sağlık gibi fizyolojik özellikler sportif performansın temel bileşenlerindedir. Örneğin genetik varyasyonların bireylerin dayanıklılık ve güç kapasiteleri üzerindeki etkileri, yüksek performanslı sporcular arasında farklılık yaratır (13).

- **Motorsal ve Fiziksel Parametreler:** Çeviklik, dayanıklılık, kuvvet, denge ve esneklik gibi motor beceriler, sporcunun fiziksel başarısını etkileyen başlıca unsurlardır. Örneğin, denge yeteneğinin çeşitli spor dallarında performans üzerindeki etkisi ve farklı sporlar arasında denge yeteneği farklılıkları belirlenmiştir (14).

Sportif performans, sosyolojik, psikolojik, fizyolojik ve motor becerileri kapsayan kompleks bir bütündür. Bu nedenle, performansın en yüksek düzeye çıkartılabilmesi için bu faktörlerin tümünün koordineli bir şekilde geliştirilmesi önemlidir.

2.3. Sportif Performansı Etkileyen Faktörler

Sportif performans, sporcunun bireysel özellikleri, çevresel etmenler ve psikolojik durum gibi birçok faktörün bir araya gelerek şekillendirdiği bir olgudur. Performansın yükseltilmesi için bu faktörlerin detaylı bir biçimde incelenmesi ve sporcunun özelliklerine uygun bir şekilde optimize edilmesi gerekmektedir. Bu bağlamda, performansı etkileyen faktörler genelde içsel ve dışsal faktörler olarak iki ana başlık altında sınıflandırılmaktadır.

2.3.1. İçsel Faktörler

İçsel faktörler, sporcunun biyolojik, fizyolojik ve psikolojik özelliklerini içermekte olup, bireyin doğuştan getirdiği veya zamanla kazandığı özellikleri kapsar. İçsel faktörlerin temel unsurları şunlardır:

- **Yaş ve Cinsiyet:** Yaş ve cinsiyet, fiziksel kapasite üzerinde doğrudan etkilidir ve bireylerin fizyolojik performansını belirleyen başlıca biyolojik değişkenlerdir (12).

- **Genetik ve Fiziksel Uygunluk:** Genetik yapı, bireyin dayanıklılık, hız ve kuvvet gibi fiziksel kapasite unsurlarını etkileyerek performans üzerinde önemli rol oynar. Özellikle bazı genetik varyasyonların (ACE ve ACTN3 gibi) bireylerin fizyolojik kapasitelerini artırdığı belirlenmiştir (15).

- **Psikolojik Durum:** Sporcunun motivasyon düzeyi, özgüveni, stres yönetimi gibi psikolojik faktörler performans üzerinde belirleyici bir etkiye sahiptir. Psikolojik faktörlerin olumlu yönde desteklenmesi, sporcuların sahada daha başarılı olmasını sağlamaktadır (10).

- **Fizyolojik Yapı:** Kardiyovasküler sistemin verimliliği, metabolizma hızı ve enerji üretim mekanizmalarının sağlıklı olması, bireyin fiziksel performansını doğrudan etkiler. Yüksek kardiyovasküler kapasite ve hızlı enerji üretimi, uzun süreli dayanıklılık gerektiren spor dallarında büyük avantaj sağlar (16).

2.3.2. Dışsal Faktörler

Dışsal faktörler, bireyin yaşadığı çevre, antrenman koşulları ve sosyal destek gibi etmenleri kapsar. Çevresel şartlar ve toplumsal destek, sporcuların performanslarını etkileme potansiyeline sahiptir. Başlıca dışsal faktörler şunlardır:

- **İklim ve Çevre Koşulları:** Spor yapılan alanın iklimsel özellikleri (sıcaklık, nem vb.), sporcunun performansını etkileyen önemli dışsal faktörlerdendir. Özellikle açık hava sporlarında iklim koşulları performansı belirgin şekilde etkileyebilir (17).

- **Antrenman Teknikleri ve Periyotları:** Antrenman metotları, sporcunun performans gelişimini doğrudan etkileyen önemli bir faktördür. Uygulanan antrenman teknikleri ve periyotlamının doğru şekilde planlanması, performansın sürdürülebilirliğini sağlamada kritik rol oynamaktadır (18).

- **Sosyal Destek ve Ekonomik Durum:** Sosyal çevrenin desteği, özellikle aile ve arkadaşların sporcunun gelişim sürecinde sağladığı moral desteği performansı olumlu yönde etkileyebilir. Ekonomik durum ise, kaliteli antrenman ekipmanlarına erişim gibi performansı arttıracak faktörlere olanak sağlar (10).

- **Beslenme ve Takviyeler:** Sporcuların düzenli ve dengeli beslenmesi, kas gelişimini ve dayanıklılığı arttırmada önemli bir rol oynar. Ayrıca, besin takviyeleri de performans artırıcı olarak kullanılabilir (19).

- Sportif performans, içsel ve dışsal faktörlerin etkileşimi sonucunda ortaya çıkan bir olgudur. Performansın artırılabilmesi için bu faktörlerin detaylı bir şekilde analiz edilmesi ve bireye özgü antrenman ve destek planları oluşturulması gerekmektedir.

2.4. Performansın Değerlendirilmesinde Kullanılan Yöntemler

Sportif performansın ölçülmesi, sporcunun fiziksel kapasitesinin, antrenman süreçlerinin ve müsabaka sırasında gösterdiği performansın objektif verilerle takip edilmesini sağlar. Egzersiz ile birlikte vücudun verdiği fizyolojik yanıtları anlamak, performansın takibi ve bu bilgilerin sayısal verilere dökülmesi için çeşitli değerlendirme yöntemleri geliştirilmiştir. Bu değerlendirme yöntemleri, sporcuların fiziksel, zihinsel ve teknik kapasitelerini ölçmek ve antrenman programlarını optimize etmek amacıyla kullanılmaktadır (20). Performans değerlendirme yöntemleri genellikle dört ana başlık altında toplanır:

- **Ölçümler:** Fiziksel ölçümler, sporcunun fiziksel durumunu değerlendiren objektif veriler sağlar. Bu ölçümler, ağırlık, boy, yağ oranı gibi antropometrik verilerden, kardiyovasküler dayanıklılığı ve güç kapasitesini ölçen fizyolojik testlere kadar geniş bir yelpazede uygulanmaktadır (21).

- **Performans Testleri:** Atletik performansı doğrudan ölçen testler, sporcunun kuvvet, hız, dayanıklılık ve esneklik gibi özelliklerini değerlendirir. Örneğin, maksimum kuvvet testleri (1RM), anaerobik güç ölçümleri, squat ve bench press gibi kuvvet testleri, performans ölçümünde yaygın olarak kullanılmaktadır (22).

- **Sportif Hareket Analizi:** Hareket analizi, sporcunun teknik becerilerini inceleyerek performansı optimize etmeye yönelik veriler sağlar. Hareket analizi, kuvvet plakaları, yüksek hızlı video analizi gibi teknolojilerle yapılabilir ve özellikle koşu, atlama ve dönüş gibi teknik detayların analiz edilmesinde kullanılır (23).

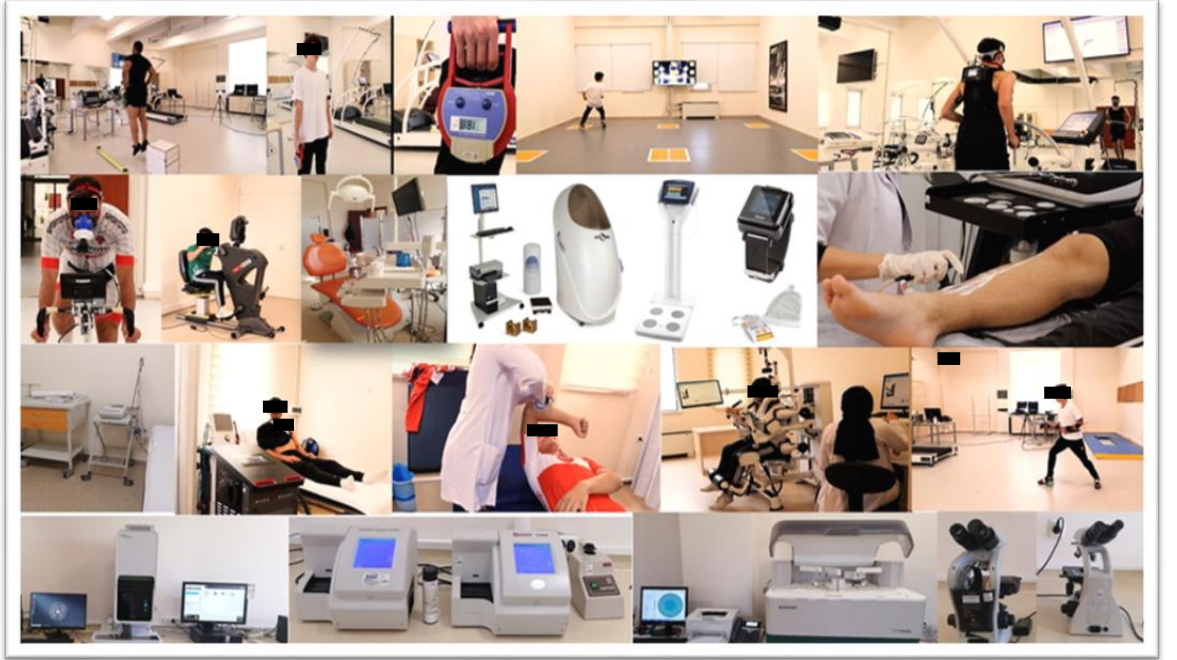
- **Müsabaka Analizi:** Sporcunun müsabaka sırasında gösterdiği performansı analiz eden bu yöntem, oyun sırasında yapılan hareketlerin ve stratejilerin veriye

dökülmesini içerir. Müsabaka analizi, oyun içinde yapılan hareketlerin etkinliğini ölçmek için oyun verilerini kullanarak takım sporlarında sıkça uygulanmaktadır (24).

Spor performansını değerlendirme yöntemleri, sporcunun yetenek ve gelişim seviyesini objektif bir şekilde anlamayı ve antrenmanları buna göre düzenlemeyi sağlar. Performans analizlerinin doğru uygulanması, sporcunun hem kişisel gelişimini hem de müsabaka sonuçlarını olumlu etkileyebilir.

2.4.1. Ölçümler

Sportif performansın geliştirilmesi ve sakatlık risklerinin yönetilmesi için yapılan ölçümler, sporcuların fiziksel kapasiteleri ve performans profilleri hakkında değerli bilgiler sağlar. Bu ölçümler, antropometrik ve fizyolojik olmak üzere iki ana kategoriye ayrılır. Sportif performans ölçümleri, sağlık taramaları ve rehabilitasyon hizmetlerine ait görsel Şekil 2.1’de verilmiştir.



Şekil 2.1. Sportif performans ölçümleri, sağlık taramaları ve rehabilitasyon hizmetleri.

- **Antropometrik Ölçümler**

Antropometrik ölçümler, vücut kompozisyonu, kas kütlesi ve yağ oranı gibi parametreleri içerir. Sporcuların fiziksel yapısına yönelik bu ölçümler, branşlarına göre uygunluklarını ve sakatlık risklerini belirlemeye yardımcı olur (25). Örneğin, boy, kilo, deri kıvrımı kalınlığı, kas çevresi gibi ölçümler sporcuların performans kapasitesini belirlemek ve sakatlık risklerini öngörmek için kullanılır (26). Futbol ve basketbol gibi spor dallarında bu antropometrik veriler sporcunun dayanıklılığı ve hızlı toparlanma yeteneği ile doğrudan ilişkilidir (27).

- **Fizyolojik Ölçümler**

Fizyolojik ölçümler, sporcuların aerobik kapasite, maksimal oksijen tüketimi (VO₂ max), anaerobik güç gibi performans faktörlerini değerlendirir. Bu ölçümler sporcunun dayanıklılığını, hızını ve genel performans kapasitesini belirlemekte kritik rol oynar (28). Örneğin, dayanıklılık sporlarında (maraton koşuları, bisiklet vb.) VO₂ max değeri sporcuların performansını değerlendirmede temel bir gösterge olarak kabul edilir (29). Bu iki ölçüm kategorisi, sporcunun bireysel performans profilini oluşturmak, sakatlık risklerini en aza indirmek ve antrenman programlarının etkinliğini artırmak için kullanılmaktadır.

2.4.2. Performans Testleri

Sporcuların fiziksel, fizyolojik, psikolojik ve teknik becerilerini değerlendirmek amacıyla kullanılan performans testleri, spor bilimi alanında objektif veri elde etme açısından büyük önem taşır. Bu testler, sporcunun branşına uygun performans seviyesinin değerlendirilmesi ve izlenmesi için gereklidir. Performans testlerinin geçerli, güvenilir ve bilimsel temellere dayalı olması zorunludur. Ayrıca, test sonuçlarının karşılaştırılabileceği standart ve normal değerlere sahip olması gerekir (30).

Performans değerlendirmelerinde kullanılan testler saha veya laboratuvar ortamında yapılabilir. Saha testleri genellikle sporcunun antrenman ortamına daha

uygun, gerçek koşullara benzer bir yapıdadır. Laboratuvar testleri ise kontrollü ortamlarda yapılır ve daha hassas ölçüm imkânı sunar (31). Performans testleri; genel olarak, temel verimlilik, metabolik ve psikolojik durum değerlendirme testleri olarak üç ana başlık altında sınıflandırılabilir:

2.4.3. Temel Verimlilik Testleri

Temel verimlilik testleri, sporcunun güç, kuvvet, hız, dayanıklılık, esneklik ve reaksiyon zamanı gibi fiziksel özelliklerini ölçmeyi hedefler. Bu tür testler sporcunun branşına özel gereksinimlere uygun olarak seçilmelidir. Örneğin:

- **Güç Testleri:** Dikey sıçrama ve derinlik sıçrama testleri sporcunun patlayıcı güç kapasitesini ölçmek için kullanılır.
- **Kuvvet Testleri:** Maksimal kuvvet ve izometrik kuvvet testleri, sporcunun kas gücünü belirlemede önemli rol oynar (32).
- **Dayanıklılık Testleri:** Aerobik ve anaerobik dayanıklılık, sporcunun uzun süreli ve yüksek yoğunluklu aktivitelerde performansını ölçer.
- **Esneklik Testleri:** Y-Stabilite testi ve uzan eriş testi gibi yöntemlerle sporcunun esnekliği ölçülür (33).

Bu testlerin bilimsel geçerliliği, sporcuya uygun protokollerle yapılması ve sonuçların karşılaştırılabilir standart değerlere sahip olması, performans değerlendirme süreçlerinde başarı sağlar.

2.4.4. Metabolik Testler

Metabolik testler, sporcunun enerji üretim mekanizmalarını değerlendirmeye yönelik testlerdir ve genellikle laboratuvar ortamında yapılır. Laktik asit takibi ve gaz değişim analizi gibi testler metabolik kapasitenin belirlenmesinde kullanılır (34). Bu testlerle elde edilen veriler, sporcunun dayanıklılık seviyesini ve enerji verimliliğini değerlendirmek için önemlidir. Örneğin, laktik asit birikimi ile dayanıklılık arasındaki

ilişki, sporcunun anaerobik kapasitesini anlamamızı sağlar. Gaz değişim analizleri ise sporcunun oksijen tüketimi (VO₂ max) gibi aerobik kapasite değerlerini ölçer. Bu testler, spor branşına uygun olarak farklı yoğunluk ve sürelerde yapılabilir ve dayanıklılık sporlarında sporcunun performans kapasitesini ortaya koymada kritik öneme sahiptir (28).

2.4.5. Psikolojik Durum Değerlendirme Testleri

Psikolojik durum değerlendirme testleri, sporcunun mental durumunu, stresle başa çıkma kapasitesini, kaygı düzeyini ve kişilik özelliklerini analiz eder. Bu testler, sporcuların zihinsel dayanıklılığını ve müsabaka esnasında performanslarını etkileyen psikolojik faktörleri değerlendirmede önemlidir (35).

Psikolojik değerlendirme testlerinde kişilik, stres yönetimi ve anksiyete gibi unsurlar ele alınır. Bu testlerin sonucunda elde edilen veriler, sporcunun müsabaka sırasında karşılaştığı mental zorluklarla nasıl başa çıktığını anlamada önemli bir rol oynar. Örneğin, kendine güven, motivasyon ve duygu kontrolü gibi psikolojik özellikler performans üzerinde doğrudan etkiye sahiptir (36).

2.5. Çalışmada Kullanılan Performans Testleri

Bu çalışmada, sporcuların branşlarına uygun olarak seçilen performans testlerinden izometrik orta uyluk çekiş testi, üst ve alt ekstremiteler anaerobik güç testi, çok yönlü reaktif çeviklik testi ve aerobik güç ve kapasite testi uygulanmıştır. Bu testlerin sonuçları veri madenciliği yöntemleriyle analiz edilerek sporcuların sınıflandırılması sağlanmıştır. Böylece sporcuların performans özellikleri daha ayrıntılı şekilde değerlendirilebilmektedir.

2.5.1. İzometrik Orta Uyluk Çekiş Testi (IMTP)

İzometrik Orta Uyluk Çekiş Testi (IMTP), sporcuların tüm vücut kuvvetini ölçmeye yönelik bir test olup, maksimum izometrik kuvvet üretme kapasitesini belirler. Bu test, sporcuların kuvvet ve dayanıklılık gelişimini analiz etmek ve sakatlık riskini değerlendirmek için yaygın olarak kullanılmaktadır (37).

IMTP testinde, sporcunun diz ve kalça açıları ayarlanarak, belirli bir pozisyonda sabit kalması sağlanır. Sporcu, ayakları kuvvet platformuna doğru iterek ve elleriyle çubuğu çekerek maksimum kuvveti üretmeye çalışır. Test sırasında kullanılan kuvvet platformları, bilgisayar ara yüzleri aracılığıyla kuvvet-zaman eğrisini doğrudan ölçmeye olanak tanır (38). Şekil 2.2’de IMTP testinin uygulanışına ait görsel verilmiştir.



Şekil 2.2. IMTP testinin uygulanışına ait görsel.

IMTP'nin Avantajları ve Kullanım Alanları:

IMTP, antrenörler ve araştırmacılar tarafından sporcunun fiziksel kapasitesini değerlendirmek ve sporcular arasında karşılaştırma yapabilmek amacıyla tercih edilen bir yöntemdir. Test sonuçları, sporcuların rölatif çekiş değerleriyle performanslarını karşılaştırma imkânı sunar ve sağ-sol ekstremitte kuvvet dengesizlikleri sakatlık riski açısından değerlendirilir. IMTP, özellikle:

- **Kuvvet ve Patlayıcı Güç Ölçümü:** IMTP testinde üretilen tepe kuvvet değerleri, sporcunun patlayıcı güç kapasitesini değerlendirir. Bu özellik, özellikle basketbol gibi kuvvet gerektiren branşlarda performansın öngörülmesi açısından önemlidir (39).

- **Dayanıklılık ve Sürat İlişkisi:** IMTP'nin dayanıklılık ve hız performansı ile olan ilişkisi yapılan araştırmalarda gösterilmiştir. Özellikle, kuvvet-zaman eğrisi üzerinden elde edilen veriler, sporcunun hızlanma kapasitesi gibi dinamik performans göstergeleriyle ilişkilendirilmiştir (40).

Uygulama ve Yöntem:

IMTP, sabit ya da taşınabilir platformlarda yapılabilir. Platformdaki kuvvet-zaman eğrisi analizi ile sporcunun maksimum kuvvet değeri hesaplanır. Bu test, maksimum kuvvet seviyesini objektif olarak değerlendirmek amacıyla yapılır ve düşük hata oranları ile yüksek doğrulukta sonuçlar vermektedir (41).

2.5.2. Wingate Anaerobik Güç ve Kapasite Testi

Wingate Anaerobik Güç Testi (WAnT), sporcuların anaerobik güç ve kapasitesini ölçmek amacıyla geliştirilmiş kısa süreli, yüksek yoğunluklu bir testtir. Bu test, yüksek yoğunluklu egzersizlerde kas metabolizmasını değerlendirir ve patlayıcı kuvvet, kuvvette devamlılık ile kassal yorgunluk gibi parametreleri analiz eder. WAnT, özellikle sprint, halter ve voleybol gibi anaerobik kapasitenin yüksek önem taşıdığı branşlarda sıkça kullanılır (42).



Şekil 2.3. Üst ekstremitte Wingate anaerobik güç ve kapasite testine ait görsel.

Testin Uygulanışı:

Wingate testinde sporcu, alt ve üst bisiklet ergometresi üzerinde sabit bir direnç altında maksimum pedal hızına ulaşmaya çalışır ve 30 saniye boyunca bu yüksek hızını korumaya çabalar. Testin ilk birkaç saniyesinde ölçülen maksimum güç değeri, sporcunun anaerobik gücünü temsil ederken; dayanıklılık ve yorgunluk oranı gibi değişkenler de bu süreçte analiz edilir (43). Şekil 2.3'te anaerobik güç ve kapasite testinin üst ekstremitte uygulamasına, Şekil 2.4'te ise alt ekstremitte uygulamasına ait görseller verilmiştir.



Şekil 2.4. Alt ekstremitte Wingate anaerobik güç ve kapasite testine ait görsel.

Test Sonuçlarında Ölçülen Ana Kavramlar:

Wingate Anaerobik Güç Testi sonucu elde edilen veriler şu kavramlar çerçevesinde değerlendirilebilir:

- **PP (W):** Test boyunca ulaşılan zirve güç (Watt).
- **AP (W):** Test boyunca uygulanan ortalama güç (Watt).

- **MP (W)**: Test boyunca ulařılan en düşük güç (Watt).
- **tPP (s)**: Zirve güce ulařma süresi (saniye).
- **PPR (W/kg)**: Kilogram başına zirve güç (Watt/kilogram).
- **APR (W/kg)**: Kilogram başına ortalama güç (Watt/kilogram).
- **PD (%)**: Güç kaybı oranı yüzdesi.

Bu deęişkenler sporcuların anaerobik kapasitesini belirlemede önemlidir ve spor dalları arasında performans karşılařtırmaları yapmayı sağlar (44).

Anaerobik Kapasitenin Spor Performansına Etkisi:

Wingate Testi, özellikle patlayıcı güç gerektiren spor dallarında sporcuların performans potansiyelini deęerlendirmek için kullanılır. Test sonuçları, sporcuların anaerobik kapasitelerini ve bu kapasitenin antrenman programlarına uygunluęunu belirlemede yardımcı olur (45).

2.5.3. Aerobik Güç ve Kapasite Testi

Aerobik güç ve kapasite testi, sporcunun maksimum oksijen tüketimini (MaxVO₂) yani büyük kas gruplarının aktif olduęu egzersiz esnasında vücudun oksijen taşıma ve kullanma kapasitesini ölçer. Bu test, kardiyovasküler dayanıklılıęın geçerli bir ölçüsü olarak kabul edilir ve sporcunun yoğun egzersiz koşullarında oksijen kullanabilme yetisini deęerlendirir (46). MaxVO₂ deęeri yüksek olan sporcuların atletik performanslarının artması beklenir, çünkü daha fazla oksijen taşıyabilme kapasitesi uzun süreli dayanıklılıęı destekler.



Şekil 2.5. Aerobik güç ve kapasite testi koşu ergometresine ait bir görsel.

Aerobik Güç ve Kapasite Testinin Uygulanışı ve Yöntemleri:

Aerobik güç ve kapasite testleri, bisiklet, koşu veya kürek ergometresinde uygulanabilir. Ölçüm sırasında "breath-by-breath" otomatik gaz analiz sistemi ile sporcunun oksijen tüketimi ve karbondioksit üretimi gibi solunum parametreleri gerçek zamanlı olarak kaydedilir. Bu test, sporcuların dayanıklılık düzeyini belirlemede kullanılır ve atletlerin aerobik güç kapasitelerini ölçmek için güvenilir bir yöntemdir (47). Şekil 2.5'te aerobik güç ve kapasite testinin koşu ergometrisine, Şekil 2.6'da aerobik güç ve kapasite testinin test esnasındaki veri akış şemasına, Şekil 2.7'de ise bisiklet ergometresi uygulamasına ait görseller verilmiştir.

MaxVO₂ ölçümünde, aşağıdaki kriterlerden üçünün gözlemlenmesi, maksimal oksijen kullanım kapasitesine ulaşıldığını gösterir:

- İş yükü artmasına rağmen VO₂ değerindeki artışın 1,5 ml/dk/kg'den az olması,
- Algılanan yorgunluk düzeyinin yüksek olması (Borg skalasında ≥ 17),
- Karbondioksit-oksijen tüketim oranının (RQ) 1,10 veya daha yüksek olması (48).

Test Sonuçlarında Ölçülen Ana Kavramlar

Aerobik güç ve kapasite testi sonucunda elde edilen temel parametreler, sporcunun oksijen tüketim kapasitesini ve dayanıklılığını değerlendirmek için kullanılır. Bu parametreler:

- **VO₂ (ml/dk)**: Dakikada tüketilen oksijen miktarı,
- **VCO₂ (ml/dk)**: Dakikada üretilen karbondioksit miktarı,
- **RQ**: Karbondioksit üretim miktarının oksijen tüketimine oranı,
- **Max KAH (bpm)**: Maksimum kalp atım hızı
- **AT Hız (km/s)**: Anaerobik eşiğe ulaşıldığında ölçülen hız.

Bu parametreler, sporcunun dayanıklılık kapasitesini detaylı bir şekilde değerlendirme imkânı sunar ve özellikle dayanıklılık gerektiren spor branşlarında performansın öngörülmesine katkı sağlar (49).

2.5.4. Çok Yönlü Reaktif Çeviklik Testi

Çok yönlü reaktif çeviklik testi, sporcuların ani yön değiştirme ve çeviklik becerilerini ölçmek için kullanılan, görsel ve karar verme bileşenlerini de içeren bir performans testidir. Bu test, hız, koordinasyon, reaksiyon ve patlayıcı kuvvet gibi yön değiştirme yeteneğine katkı sağlayan unsurları değerlendirir (50). Testin görsel uyaranlara yanıt verme yeteneğini ölçmesi, sporcuların yarışma sırasında rakip veya top gibi dış faktörlere karşı verdikleri reaksiyonları simüle eder (51). Şekil 2.8'de çok yönlü reaktif çeviklik testine ait görsel verilmiştir.



Şekil 2.8. Çok yönlü reaktif çeviklik testine ait bir görsel.

Testin Uygulanışı:

Sporcu, belirli bir alanda ekranda gösterilen işlemlere veya rakip olarak simüle edilen görsel uyarılara en kısa sürede tepki verir. Bu test, farklı yönlerde yapılan dönüşlerde sporcunun reaksiyon süresini ölçmek için kullanılır. Elde edilen verilere göre:

- **Ortalama Dönüş Süresi (sn):** Her yön değiştirme süresinin ortalaması,
- **Ortalama Soldan ve Sağdan Dönüş Süresi (sn):** Sol ve sağ dönüşlerdeki ortalama süre farkları,
- **Sol-Sağ Farkı (sn):** Her iki yöne yapılan dönüşlerdeki fark, sporcunun dengesizliğini gösterir (52).

Test Sonuçlarının Yorumlanması:

Sporcuların sağ ve sol yönlerdeki dönüş sürelerindeki fark, sakatlık öyküsü veya kas dengesizliği gibi faktörlerden etkilenebilir. Test sonuçları, sporcuların sağ ve sol bacak kuvveti arasındaki farklılıkları dengelemeye yönelik antrenman ihtiyaçlarını belirler. Yön değiştirme çevikliği ile karar verme becerilerini aynı anda geliştirmek, müsabaka performansını artırmada önem taşır (53).

2.6. Elit Sporcu Kavramı

Elit sporcular, genellikle fiziksel, teknik ve zihinsel açıdan üst düzey yeteneklere sahip olup, ulusal veya uluslararası düzeyde önemli başarılar elde etmiş bireyler olarak tanımlanmaktadır (54). Ancak, elit sporcuları elit olmayanlardan kesin ve net bir şekilde ayırt edebilecek kriterler tam anlamıyla belirlenmemiştir. Geleneksel olarak, daha üst turlarda seçilen (draft edilen) veya daha üst düzey liglerde oynayan sporcular, elit sporcu olarak kabul edilmektedir (55). Elit sporcuların tanımlanmasına yönelik çalışmalarda, çeşitli antropometrik ve fizyolojik özellikler (56), psikolojik faktörler (57), denge (14), sporcunun takımdaki rolü (58), antrenman süresi, antrenman türü (59) yetenek gelişimi ve olgunlaşma (60) ve fiziksel performans (61) gibi çeşitli değişkenler incelenmiştir.

Araştırmalara göre, elit sporcuların diğer sporculara kıyasla daha yüksek fiziksel kapasiteye, stratejik düşünme becerisine ve stresle başa çıkma yeteneğine sahip olduğu ortaya konmuştur. Ayrıca, biyomekanik ve fizyolojik ölçütler açısından da belirgin üstünlükler sergiledikleri vurgulanmaktadır.

2.6.1. Judo ve Güreş Sporcularında Elit Sporcu Kavramı

Mücadele sporları arasında yer alan judo ve güreşte elit sporcular, fiziksel güç, teknik beceri, stratejik düşünme ve zihinsel dayanıklılık açısından en yüksek performansı sergileyen bireyler olarak tanımlanmaktadır. Bu sporcular, genellikle ulusal veya uluslararası düzeyde madalyalar kazanmış, spor federasyonları tarafından tanınıp desteklenen kişilerdir (62).

Judo Branşında Elit Sporcu Özellikleri:

Judo, güç, çeviklik ve dayanıklılığın yanı sıra üst düzey teknik ve taktik beceriler gerektiren bir spor dalıdır. Elit düzeydeki judocular, müsabaka sırasında rakibin hareketlerini hızla analiz ederek uygun teknikleri uygulama yeteneğine sahiptir. Yapılan araştırmalar, elit judocuların anaerobik kapasite, el-göz koordinasyonu ve hız açısından diğer düzeydeki sporculara göre üstün olduğunu ortaya koymaktadır (63). Ayrıca, judo branşında başarının psikolojik faktörlerle de yakından ilişkili olduğu vurgulanmaktadır. Stres yönetimi, müsabaka öncesi odaklanma ve mücadele sırasında zihinsel dayanıklılık elit düzeydeki judocuların performanslarını destekleyen kritik unsurlardır (64).

Güreş Branşında Elit Sporcu Özellikleri:

Güreş, kuvvet, dayanıklılık ve esneklik gibi fiziksel özelliklerin ön planda olduğu bir spor dalıdır. Elit güreşçiler, rakiplerini kısa sürede etkisiz hale getirmek için patlayıcı güç ve üstün teknik hakimiyet sergilerler. Araştırmalar, güreşte başarılı olmanın hem anaerobik hem de aerobik enerji sistemlerinin etkili kullanımına bağlı olduğunu göstermektedir (64). Psikolojik açıdan, güreşçilerin motivasyonu, baskı altında karar verme becerileri ve yenilgiyi yönetebilme yetenekleri başarılarında önemli bir rol oynamaktadır. Özellikle olimpik düzeyde yarışan güreşçiler, yoğun antrenman programlarına uyum sağlama ve yüksek rekabet ortamında performanslarını sürdürebilme yetenekleriyle öne çıkmaktadır (65).

2.6.2. Elit Olma Süreci

Judo ve güreş gibi mücadele sporlarında elit sporcu olma süreci genellikle şu aşamalardan oluşmaktadır:

- **Erken Yaşta Başlama:** Çocukluk döneminde spora yönlendirilme ve temel tekniklerin öğrenilmesi.

- **Özel Antrenmanlar:** Teknik, fiziksel ve taktik becerilerin geliştirilmesine yönelik yoğun ve düzenli antrenmanlar.

- **Ulusal ve Uluslararası Müsabakalar:** Federasyonlar tarafından düzenlenen turnuvalarda başarılı olarak üst düzey rekabet tecrübesi edinme.

- **Psikolojik Hazırlık:** Zihinsel dayanıklılık ve müsabaka stresine karşı etkili stratejiler geliştirme.

- **Etik Kurallara Uygunluk:** Disiplin, etik değerler ve yönetmelikte belirtilen engel haller (doping ihlali, şike, kanuna aykırı bahis oyunları oynatma vs.)

Sonuç olarak, judo ve güreş gibi mücadele sporlarında elit sporcular, yalnızca fiziksel üstünlükleriyle değil, aynı zamanda teknik becerileri, psikolojik sağlamlıkları ve disiplinli yaşam tarzlarıyla dikkat çekmektedir. Bu özellikler, onları kendi branşlarında zirveye taşıyan temel unsurlardır.

2.7. Veri Madenciliği

Veri Madenciliği, büyük hacimli verilerden anlamlı ve değerli bilginin keşfedilme süreci olarak tanımlanabilir (66). Bilgi sistemleri teknolojilerinin gelişmesiyle birlikte veri madenciliği uygulamaları yaygınlaşmış, eğitim, sağlık, finans, spor gibi çeşitli alanlarda karşılaşılan problemlere pratik çözümler üretebilmek amacıyla veri madenciliği algoritmaları kullanılmaya başlamıştır. Veri madenciliği bilgi keşfi sürecinde modelin kurulması ve değerlendirilmesi aşamalarını kapsamaktadır (67). Bilgi keşfi süreci ise; verilerin toplanması, temizlenmesi, birleştirilmesi, seçilmesi, dönüştürülmesi, veri madenciliği modeli oluşturulması ve örüntü değerlendirmesi ve sunum aşamalarından oluşmaktadır (68).

Veri madenciliği aşamaları:

- **Problemin tanımlanması:** Problemi tanımlamak veri madenciliği aşamalarının başında gelmektedir. Öncelikle problem çözümüne uygun hangi analiz

yönteminin kullanılacağına karar vermek gerekmektedir. Böylece süreçte karşılaşılabilecek problemleri en aza indirip yanıltıcı sonuçlar elde etmekten kaçınılmış olunacaktır.

- Veriyi anlama: Verilerin toplanması, özetlenmesi ve görselleştirilmesini içerir. Verideki değişkenleri anlamlandırmayı barındırmaktadır.
- Veri hazırlama: Toplanan veriler içerisinde değişken seçimi, veri temizliği, veriyi dönüştürme, normalleştirme/standartlaştırma, temizleme, boyut indirgeme gibi işlemlerden oluşmaktadır.
- Verilerin Modellenmesi: Bu aşamada veri madenciliği algoritmalarının (sınıflandırma/regresyon, kümeleme, birliktelik kuralları vb.) uygulanması ile model geliştirilir. Model parametreleri ayarlanarak eğitim işlemi gerçekleştirilir.
- Değerlendirme: Modelin performansı ve doğruluğu test verileri üzerinden değerlendirilir.

Modelin değerlendirilmesi (Model Performans Ölçütleri)

- ✓ Doğru sınıflama başarısı (Accuracy)
- ✓ Kappa uyum istatistiği
- ✓ Duyarlılık (Sensitivity)
- ✓ Seçicilik (Specificity)
- ✓ Ortalama mutlak hata (MAE)
- ✓ ROC AUC (ROC eğrisi altında kalan alan)
- ✓ Göreli mutlak hata (RAE)
- ✓ Pozitif Kestirim Değeri (PPV)
- ✓ Negatif Kestirim Değeri (NPV)
- ✓ F1 ölçüsü (F1-Measure) vb.

- Uygulama: Bu aşamada kurulan ve geçerliliği kabul görülen modelin izlenmesi ve zamanla karşılaşılan değişiklikler karşısında güncellenme yapılması aşamalarından oluşmaktadır.

- Sonuçların sunulması ve kullanılması: Elde edilen model ve sonuçlar, iş süreçlerine uygulanır. Bu, raporlar hazırlamak, karar destek sistemlerine entegre etmek veya otomatikleştirilmiş sistemlerde kullanmak şeklinde olabilir (69).

2.8. Makine Öğrenmesi

Makine öğrenmesi (ML) yöntemleri veri ve deneyimlerden öğrenerek belirli görevleri yerine getirmesini sağlayan ve öğrenilen bilgilerin gelecekteki benzer olaylarda karar verilmesinde yardımcı olmak amacıyla kullanılan bir yapay bir zekâ dalıdır. Makine öğrenmesi türleri “Danışmanlı Öğrenme” ve “Danışmansız Öğrenme” olarak adlandırılmaktadır.

2.8.1. Danışmansız Öğrenme

Danışmansız öğrenme yöntemlerinde, etiketlenmemiş veri setlerini giriş verileri olarak kullanılarak öğrenme gerçekleşir. Birliktelik analizi, kümeleme analizi gibi yöntemler danışmansız öğrenme yöntemlerindedir (68).

2.8.2. Danışmanlı Öğrenme

Danışmanlı öğrenme yöntemlerinde, veri setindeki gözlemler etiketlidir. Giriş yapılan yeni verinin sınıfı, eğitim seti ile kurulan model yardımı ile belirlenir. Bu başlık altında bulunan sınıflama yöntemleri, gözlem sınıflarını tahmin etmeye çalışırken, regresyon yöntemleri ise sayısal değişkenlere yönelik kestirim yapılması için kullanılır. Karar ağaçları, rastgele orman, yapay sinir ağları, destek vektör makineleri, Naive Bayes, k-en yakın komşu gibi algoritmalar danışmanlı öğrenme yöntemlerindedir (70).

2.9. Veri Madenciliği Yöntemleri

2.9.1. Tanımlayıcı Yöntemler

Tanımlayıcı modellerin amacı veriyi analiz ederek anlamaktır. Bunu veri setindeki örüntü kurallarını tanımlayarak değişkenler arasındaki ilişkileri ve anlamlı bilgilerin çıkarımını yaparak sağlar. Etiketsiz veriler kullanılır. Örnek olarak; birliktelik kuralları, kümeleme, aykırı değer analizleri verilebilir (71).

2.9.2. Kestirici Yöntemler

Kestirici modeller, mevcut veriler ve bu verilerden oluşturulan modeller yardımıyla kestirimler yapmayı amaçlamaktadır. Genellikle etiketli veriler kullanılmaktadır. Örnek olarak; sınıflandırma (karar ağaçları vb.) ve regresyon algoritmaları verilebilir (71).

2.10. Sınıflama/Regresyon Yöntemleri

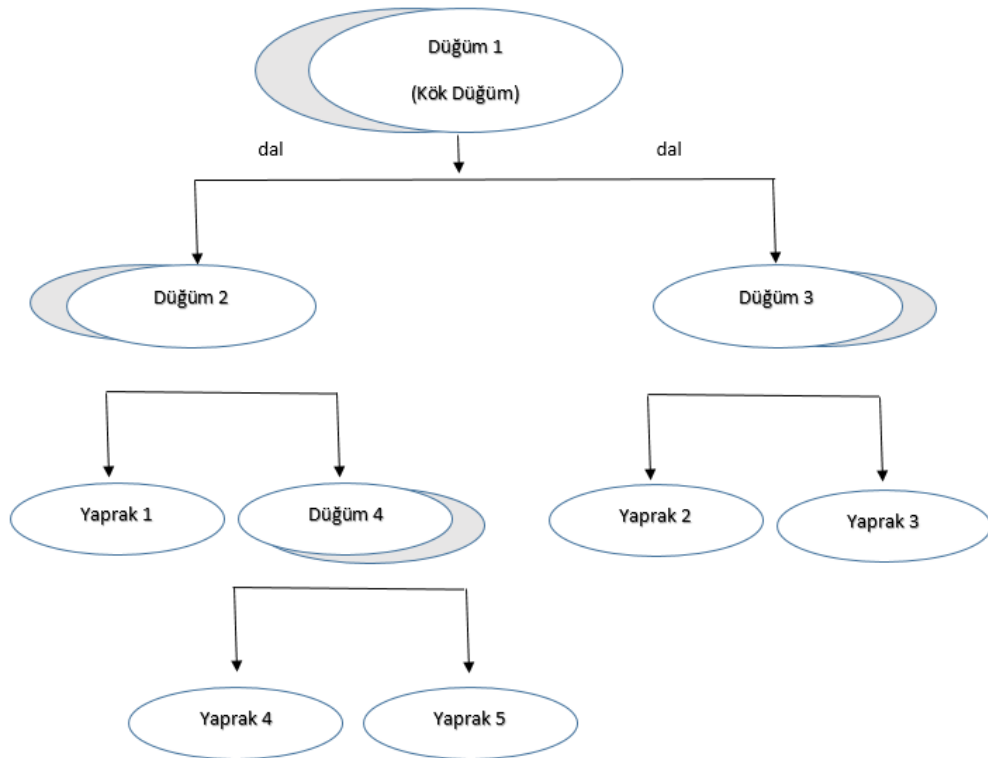
Sınıflandırma, veri sınıflarını tanımlayan ve ayıran bir model oluşturma sürecidir. Modeller, sınıf etiketi bilinen bir dizi eğitim verisinin analizine dayanarak türetilir ve sınıf etiketi bilinmeyen verilerin hangi sınıfa ait olduğunu tahmin etmek için kullanılır (68). Sınıflama yöntemleri kategorik, regresyon yöntemleri ise sayısal değişken kestirimi amacıyla kullanılmaktadır (70).

Sınıflama/Regresyon yöntemleri öğrenme-doğrulama-uygulama olmak üzere 3 aşamadan oluşmaktadır. İlk aşama olan öğrenme aşaması veri setinde yüksek performans sergileyen modelin seçimini kapsar. Doğrulama aşamasında, ilk aşamada seçilen model test veri seti ile çalıştırılır. Üçüncü aşama olan uygulama aşamasında ise yeni ve farklı bir veri setinde model çalıştırılarak tahminde bulunulur.

Bu tez kapsamında kullanılan sınıflama/regresyon yöntemleri ve bu yöntemlerden; Rastgele Orman, XGBoost, AdaBag, KNN, Lojistik Regresyon Modeli, Karar Ağaçları (CART, CART Bagging, C5), Naive Bayes ve Destek Vektör Makinaları yöntemleri üzerinde durulacaktır.

2.11. Karar Ağacı Yöntemleri

Karar ağaçları, sınıflama yöntemleri için sıklıkla kullanılan ve iyi performans gösteren algoritmalarıdır. Bağımsız değişkenin düzeyine bağlı olarak bağımlı değişkeni parçalara ayıran bir algoritma olarak da tanımlanabilir. Karmaşık veri yapılarında, verileri dallara ayırarak anlaşılması ve yorumlanması kolay çıktılar veren, aşırı değerlere karşı dayanıklı olan etkili bir model olarak öne çıkmaktadır. Ağacın başlangıç düğümüne kök düğüm, veriyi sınıflandırmak için belirli özelliklere göre belirlenmiş bölünme noktalarına terminal düğüm, sınıfları veya kararları temsil eden kısımlara ise yaprak adı verilir. Şekil 2.9'da örnek bir karar ağacı yapısı verilmiştir. Karar ağaçları algoritmasında; sonlandırma kriteri, yapraklardaki minimum gözlem sayısı ve yapraklarda yer almayan gözlemler için seçim yöntemi olmak üzere temel koşullar barındırmaktadır (72).



Şekil 2.9. Örnek bir karar ağacı yapısı.

Aşırı uyum (overfitting) problemi ile birçok sınıflandırma probleminde olduğu gibi karar ağacı yöntemlerinde de karşılaşılabilmektedir. Aşırı uyum, eğitim veri seti yardımıyla kurulan modelin farklı veri seti üzerinde çok düşük bir performans

göstermesi durumu olarak tanımlanmaktadır. Ağaç budama (pruning) yöntemleri bu aşırı uyum problemini iki farklı yöntem ile ele alır. Ön budama, bölünmeye önceden müdahale edilerek; sonradan budama ise, karar ağacı tamamen oluşturulduktan sonra modele katkı sağlamayan alt ağaçları belirleyerek modelden çıkarılması ile gerçekleştirilir. Karar ağacı algoritmaları oluşturulurken farklı ölçütler yardımı ile bölünmenin başlayacağı değişken seçimi yapılmaktadır. Karar ağaçlarında, ID3, C4.5, C5.0, CART, CHAID ve QUEST kullanılan çeşitli algoritmalar (68).

2.11.1. Bölünme Kriterleri

Bilgi Kazancı:

ID3, nitelik seçimi ölçüsü olarak bilgi kazanımını kullanır. D sınıf etiketli veri seti, C_i sınıf değişkeni, $|D|$ D veri setinde bulunan gözlem sayısı ve D veri setindeki sınıf değişkenin sayısı $|C_i D|$ olmak üzere;

$$Bilgi(D) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (2.1)$$

$Bilgi(D)$ D değişkeninin entropi değeridir. Eşitlik 2.1'de; p_i , veri setindeki rastgele bir gözlemin, $|C_i D|/|D|$ ile hesaplanan, C_i sınıfında bulunma olasılığıdır. Entropi rastgele bir değişkenle ilgili belirsizliğin (belirli bir bilginin beklenen içeriğinin) bir ölçüsü olarak tanımlanır (73). Entropi sistemdeki düzensizliğin ve belirsizliğin bir ölçüsü olarak tanımlanmış ve daha yüksek entropi değerinin daha yüksek karmaşıklık olarak ortaya çıktığını belirtmiştir. Eğer değişkenin tüm değerleri birbirine eşit ise belirsizlik yok yani entropi değeri 0 olur. Değişken düzeyleri birbirinden farklı ise belirsizlik artar ve entropi maksimum olan 1 değerine ulaşır. Yüksek entropi değerine sahip değişken daha çok bilgi içerir (74).

D veri seti, $A: \{a_1, a_2, a_3, \dots, a_v\}$ v düzeyli A değişkenine göre v kısma bölünebilir. Bu kısımlar A düğümünden çıkan dallara karşılık gelmektedir. Bölünmeden sonra elde edilen bilgi Eşitlik 2.2 ile elde edilir.

$$Bilgi_A(D) = \sum_{j=1}^m \frac{|D_j|}{|D|} \times Bilgi(D_j) \quad (2.2)$$

Dallanmanın hangi deęişkenden başlayarak yapılacağı bilgi kazancı hesabı ile belirlenir. Hesaplanan en yüksek kazancın ait olduęu deęişkenden başlayarak bölünme yapılır.

$$\text{Bilgi Kazancı}(A) = \text{Bilgi}(D) - \text{Bilgi}_A(D) \quad (2.3)$$

Her bir deęişken için hesaplanan bilgi kazancı deęeri hangi deęişken için en yüksek ise düęüm dallanmaya bu deęişkenden başlayarak yineleyerek devam eder (68).

Kazanç Oranı:

Bilgi kazancı çok deęişkenli veri setlerinde verilerin ölçü birimlerinden kaynaklı yanlı sonuçların ortaya çıkmasına sebep olabilir. Bu durumda bazı algoritmalarda kazanç oranı kullanılmaktadır. Örnek olarak C5.0 algoritması verilebilir. Büyük deęerli ölçüm verilerinden kaynaklı yanlı sonuçlardan kaçınabilmeyi sağlar. D_j , deęişkendeki deęerlerin tekrarlanma adedi; D belirlenen olayın gerçekleşme sayısı olmak üzere bölünme bilgisi Eşitlik 2.4 ile hesaplanmaktadır.

$$\text{Bölünme Bilgisi}_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right) \quad (2.4)$$

Bölünme, Eşitlik 2.5 ile hesaplanan bölünme bilgisi deęerinden elde edilecek en yüksek kazanç oranına sahip deęişkenden başlar (68).

$$\text{Kazanç Oranı} = \frac{\text{Bilgi}(A)}{\text{Bölünme Bilgisi}(A)} \quad (2.5)$$

Gini İndeksi:

Gini indeksi, bir veri setindeki saflığın (*purity*) veya homojenliğin ölçüsüdür. Bir düęümdeki örneklerin tek bir sınıfa ait olma olasılığının ölçüsüdür. p_i , n sınıf içeren

D veri setindeki herhangi bir gözlemin C_i sınıfına ait olma olasılığıdır ve $(|C_i|/|D|)$ olmak üzere gini indeksi Eşitlik 2.6 ile hesaplanır (68).

$$Gini(D) = 1 - \sum_{i=1}^n p_i^2 \quad (2.6)$$

D veri setinde A değişkeninin iki düzeyine göre D_1 ve D_2 bölünmesi için gini indeksi hesaplaması yapılır.

$$Gini_A(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2) \quad (2.7)$$

A değişkeni üzerinden yapılan ikili bölünme belirsizliğindeki indirgeme;

$$\Delta Gini(A) = Gini(D) - Gini_A(D) \quad (2.8)$$

Eşitliği ile hesaplanır. Bölünme, en büyük gini indeksini veren değişkenden başlar. CART algoritması gini indeksini bölünme kriteri olarak kullanmaktadır (68).

2.12. C4.5 Modeli

Ross Quinlan tarafından 1993 yılında oluşturululan C4.5 algoritması, ID3 modelinden farklı olarak kategorik ve sürekli sayısal veri setlerinde de çalışan, karar ağaçlarının anlamlı ve duyarlı kurallar çıkarabilmesi amacıyla geliştirilmiş modellerdir. Kayıp gözlemleri hesaplama dışında tutarak hesaplanan kazanç oranlarının en yüksek değerinden başlayarak tekrarlı olarak bölünme yapılır (75).

2.13. CART Modeli

Breiman ve ark. tarafından 1984 yılında geliştirilmiştir. CART (Classification and Regression Trees) modeli hem sayısal hem de kategorik değişkenler ile kestirim yapmak amacıyla kullanılır. Gini indeksini bölünme kriteri olarak kullanmakta bu sebeple üretilen ağaçlar ikili bölünme yapısı ile alt düğümlere ayrılabilir. Bu ise çok kategorili değişkenlerin bulunduğu veri setlerinde çalışma konusunda dezavantaja sahip olmasına sebep olur (76).

Avantajları:

- Hem sınıflama hem de regresyon için uygun bir modeldir.
- Parametrik olmayan bir modeldir. Herhangi bir dağılım varsayımı bulunmamaktadır.
- Kolay yorumlanabilir modeller oluşturur.

2.14. Topluluk (Ensemble) Yöntemleri

Veri madenciliğinde, model doğruluğunu iyileştirmek, hataları azaltmak, daha nihai bir modele ulaşmak için birden fazla model tahminlerini birleştiren topluluk yöntemleridir. Bu alt başlıkta en çok kullanılan ve araştırmamızdaki modellerde de kullanılan yöntemlerden Bagging (Bootstrap Aggregating) ve Boosting yöntemleri ile bu yöntemlerde yaygın olarak kullanılan veri örnekleme ve ağırlıklandırma stratejisi olan Bootstrap yöntemine değinilecektir.

Bootstrap:

Bootstrap 1979'da Efron tarafından önerilen yeniden örnekleme yöntemlerinden biridir. Küçük genişlikteki örneklemler için kullanılan yöntemlerden bir tanesidir (77). Bu yöntem ile n gözleme sahip bir veri setinde, yerine konarak rastgele seçilen n örneklemden eğitim için veri seti oluşturulur. Her örneğin seçilme olasılığı eşit ve bu olasılık $\frac{1}{n}$ ile örnekleme seçilmemiş olan gözlemlerin ise seçilmeme olasılığı $(1 - \frac{1}{n})$ şeklinde ifade edilir. Eğitim setlerinde bir örnekten hiç veya birden daha çok bulunabilir. Bu işlen N kez tekrarlandığında; seçilmemişlerin seçilmesi olasılığı $(1 - \frac{1}{n})^N = e^{-1} \approx 0.368$ 'dir. Bu da test setini verilerin %36.8'inin, eğitim setini ise %63.2'sinin oluşturduğunu gösterir.

Bagging:

Bagging (*Bootstrap Aggregating*) Leo Breinman tarafından 1996 yılında önerilen bir topluluk (ensemble) öğrenme yöntemidir (78). Doğru sınıflama oranını

arttırarak varyansı düşürmekte kullanılmaktadır. Bagging yöntemi veri setinden yerine koyarak rastgele örnekler seçerek, sınıf yapısını bozmadan n tane eğitim setini oluşturur. Bu aşamada (*Bootstrap*) her bir gözlemin seçilme olasılıkları eşittir. Elde edilen eğitim setlerinden n farklı model kurulur ve modeller ile sınıflar belirlenir. Sonuç sınıfını belirlerken sınıf kestirimleri oylanmasına göre çoğunluk oya sahip olan sınıf belirlenir.

Boosting:

Boosting, 1990 yılında Freund tarafından bir topluluk (ensemble) öğrenme yöntemi olarak önerilmiştir (79). Öncelikle eğitim veri setinden n tane eğitim seti yerine koyularak seçilir. Oluşturulan eğitim veri setindeki her bir gözlem için eşit ağırlıklar verilerek model oluşturulur ve sınıflar belirlenir. Sonrasında eğitim yineleme adımlarında yanlış sınıflandırılan örneklere daha yüksek, doğru sınıflandırılan örneklere ise daha düşük ağırlıklar verilir. Bu şekilde n tane temel sınıflayıcı oluşturulur ve sınıflardan çoğunluk oya sahip olan sınıf kestirilen sınıf şeklinde belirlenir.

AdaBoost (Adaptive Boosting), Gradient Boosting ve XGBoost (Extreme Gradient Boosting) algoritmaları Boosting algoritmalarının başlıcalarıdır (68).

2.15. XGBoost

Chen ve Guestrin tarafından 2016 yılında Washington üniversitesinde önerilen XGBoost modeli, ağaç yapılarını kullanır. XGBoost (Extreme Gradient Boosting), yapılandırılmış veriler üzerinde çalışan denetimli öğrenme (regresyon ve sınıflandırma) problemlerinde sıklıkla kullanılan, gradient boosting yönteminin optimize edilmiş bir uygulamasıdır (80). XGBoost ağaç modellerinin her birini önceki modelin hatasını düzeltmeye yönelik eğitir. L1 (Lasso) ve L2 (Ridge) düzenleme terimleri, modelin aşırı öğrenmesini önler. Eksik verilerle çalışabilir ve özel hata fonksiyonları tanımlanmasına izin verir. Zayıf gözlemler üzerinde daha fazla ağırlık vererek bu tür örneklerin öğrenilmesini iyileştirir.

İlk model tüm veri seti üzerinden bir tahmin yapar. Tahmin edilen ve gerçek değerler arasındaki hatayı (loss) hesaplar. Hataların gradyanı (eğimi) belirlenerek her adımda, önceki modelin hatalarını düzeltecek karar ağaçları yinelemeli olarak, belirli sayıdaki iterasyon tamamlanana veya belirlenen performans kriterine ulaşıncaya kadar devam eder (81).

2.16. AdaBag (Adaptive Bagging)

AdaBag, bagging yönteminin bir türevidir. Zayıf öğrencileri bir araya getirerek tahmin gücü yüksek modeller oluşturmayı hedefler. Klasik bagging yöntemine ek olarak, modellerin ağırlıklarını adaptif olarak güncelleyerek güçlü sınıflandırıcılar oluşturur (78). Bootstrap ile rastgele veri alt kümeleri oluşturur. Genellikle CART tipi karar ağaçları oluşturur ve her ağaca belirli ağırlıklar vererek modelin daha esnek hale gelmesini sağlar (82). Modellerin tahminleri çoğunluk oylaması veya ortalama ile birleştirilir (78).

2.17. Rastgele Orman Modeli (RO)

RO modeli eğitim verilerinden *bootstrap* (yeniden örnekleme) yöntemiyle örneklem oluşturulmak suretiyle bu örneklemden üretilen çeşitli karar ağacı sınıflandırıcılarını kullanarak bir karar ormanı meydana getirir. Sınıflandırma performansını iyileştirmeyi amaçlayan bir topluluk (ensemble) modelidir. Bu algoritma 2001 yılında Breiman tarafından geliştirilmiştir (83). Bunlar, büyük bir ağaç tabanlı model kümesi (sınıflama ve regresyon ağacı-CART) tarafından oluşturulur. Değişken yapısına göre sınıflama veya regresyon amacıyla kullanılması algoritmanın bir avantajıdır. Eğitim için ayrılan verilerin üçte ikisi (*inbag*) ile karar ağaçları oluşturulur, geri kalan üçte birlik kısmı (*out of bag*) ile kestirim hataları hesaplanır. Modelin hata oranını kestirim hatalarının ortalaması oluşturur. Sonrasında tüm karar ağaçlarının tahmin sınıfları oylanır çoğunluk oyu alan sınıf finalde model tahmin sınıfını oluşturur. Sonradan budama işlemi yapılmaz ve ağaç oluşturma sürecinin her adımında, her düğüm için en iyi bölünme, niteliklerin rastgele bir alt kümesinden seçilir. Karar ağaçlarının her biri için rastgele seçilen değişkenlerin gini indeksi hesaplanarak değişkenlerin önem ağırlıkları belirlenir (71).

2.18. Lojistik Regresyon Modeli (LRM)

LRM, tahmin değişkeni nitel bir değişken olduğu durumlarda doğrusal regresyon modeline alternatif kullanılan bir modeldir. Bağımlı değişkenin durumuna göre 3 farklı şekilde isimlendirilir. İki durumlu lojistik regresyon (*binary*); nominal lojistik regresyon, çok durumlu (*polytomous*) ve sıralı lojistik regresyon (*ordinal*) şeklinde isimlendirilir (67,84).

LRM, bağımlı değişken (cevap değişkeni) ve bağımsız değişkenler arasındaki ilişkiyi tanımlar. İki durumlu lojistik regresyon modeli Eşitlik 2.9'daki gibi ifade edilebilir.

$$P(y = 1|x) = \frac{e^{\beta_0 + \beta x}}{1 + e^{\beta_0 + \beta x}} \quad (2.9)$$

Denklemden, x bağımsız değişkenler vektörü, β_0 sabit regresyon katsayısını, β , bağımsız değişken regresyon katsayılarını verir. Eşitlik 0 ya da 1 değerini alan bağımsız değişkenlerinin kestirim değerlerinin ortaya çıkma olasılığını verir ve $0 \leq P(y = 1|x) \leq 1$ arasında değer alabilir. $P(y = 1|x)$ pozitif sınıfın ortaya çıkma olasılığı şeklinde yorumlanır.

Logit fonksiyonu ise Eşitlik 2.10'daki logit dönüşümü ile elde edilir.

$$\text{logit}(x) = \ln \left[\frac{P(y=1|x)}{1-P(y=1|x)} \right] = \beta x \quad (2.10)$$

Logit fonksiyonunun doğrusallık, süreklilik ve $[-\infty, +\infty]$ aralığında olması özellikleri doğrusal regresyon ile benzerlik göstermektedir (67).

Modelin varsayımları şunlardır:

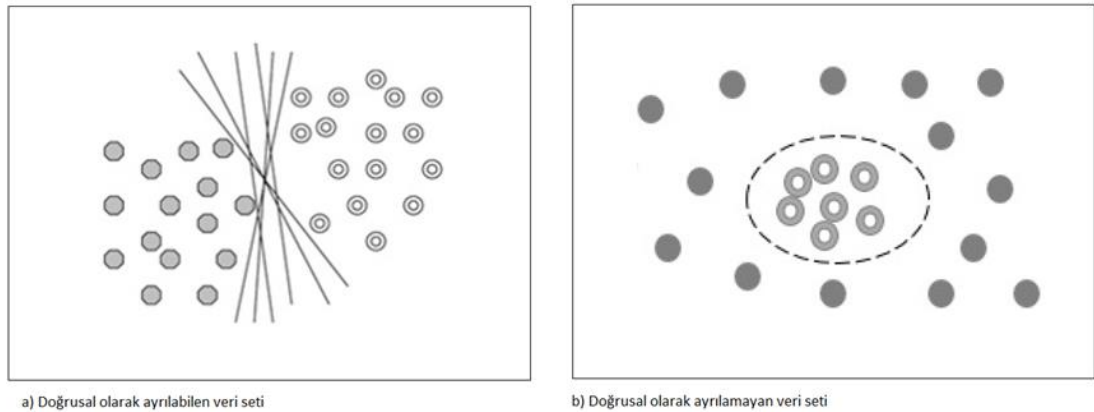
1. Hatalar (artıklar) binom dağılımı gösterir.
2. Artıkların ortalaması 0, varyans ise $P(y = 1|x)[1 - P(y = 1|x)]$ değerine eşittir.

3. Bağımsız değişkenler için dağılım ve ölçek koşulu yoktur fakat logit fonksiyonu doğrusal olmak zorundadır (84).

2.19. Destek Vektör Makineleri (DVM)

DVM, doğrusal olan ve olmayan veri setlerinin her birinde kullanılabilen bir sınıflama algoritmasıdır. Aynı zamanda regresyon problemlerinde de kullanılabilir. Özellikle doğrusal olmayan verilerde karar sınırlarını çizmede yüksek derecede doğru sonuçlar elde etmeye yardımcı olur. Güçlü kuramsal temele sahip ve çekirdek fonksiyonlar ile esnek bir yapıda olması, büyük veri setleri üzerinde çalışabilmesi ve sonuçların yüksek doğrulukta olması bakımından çokça tercih edilen bir yöntemdir.

Sınıf değişkenlerini ayıran marjini maksimum, doğrusal olan ve olmayan bir karar fonksiyonu bulmayı amaçlar. İlk olarak iki sınıflı bağımlı değişken içeren veri setlerinde doğrusal bir sınıflama için kullanılan bu yöntem daha sonra ikiden fazla sınıf içeren ve doğrusal olmayan problemlerin çözülebilmesi amacıyla da kullanılmaya başlanmıştır (85,86). Veri setlerinin doğrusal olma durumlarına göre çizilen karar sınırları Şekil 2.10'da verilmiştir.



Şekil 2.10. Veri setlerinin doğrusal olma durumlarına göre çizilen karar sınırları (87).

İki sınıfı birbirinden ayırabilmek için sonsuz sayıda doğru çizilebilir. İki sınıfı birbirinden ayıran sınıflama hatasını en küçük yapacak hiper düzlemi seçmek DVM yönteminin amacıdır. Bu işlemler çekirdek fonksiyonları $K(x_i, x_j)$ (kernel function)

kullanılarak yapılmaktadır. DVM de kullanılan çekirdek fonksiyonları Tablo 2.1’de verilmiştir.

Tablo 2.1. DVM çekirdek fonksiyonları (88).

Çekirdek Fonksiyonu	Formül
Doğrusal Fonksiyon	$K(x_i, x_j) = x_i^t x_j$
Polinomial Fonksiyon	$K(x_i, x_j) = (\alpha x_i^t x_j + b)^d$
Sigmoid Fonksiyon	$K(x_i, x_j) = \tanh(\sigma x_i^t x_j + r)$
Radyal Tabanlı Fonksiyon (RTF)	$K(x_i, x_j) = e^{(-\frac{\ x_i - x_j\ ^2}{\sigma^2})}$

2.20. Naive Bayes Modeli (NB)

NB, bir veri setinde belli bir sınıfta olma olasılığı hesabı ile sınıflandırma yapılan bir yöntemdir. En yüksek olasılığa sahip olan sınıfı tahmin olarak seçer. Bayes teoremine dayanan bu yöntem büyük hacimli verilerde hızlı çalışan ve doğruluğu yüksek olan sonuçlar üretebilen bir yöntemdir. NB yönteminde bağımsız değişkenlerin sınıf değişkeni üzerinde etkisinin olmadığı kabul edilmektedir ve bu kabul sınıf koşullu bağımsızlık (*class conditional independence*) şeklinde adlandırılır (68,88).

$X_1, X_2 \dots \dots X_n$ bağımsız rassal değişkenler vektörü ve C rassal sınıf değişkeni olmak üzere, x’in bilindiği durumlarda c’nin ortaya çıkma olasılığı Eşitlik 2.11 ile elde edilir.

$$P(C = c | X = x) = \frac{P(X=x | C=c)P(C=c)}{P(X=x)} \quad (2.11)$$

$x_1, x_2 \dots \dots x_n$ koşullu bağımsız olduğu varsayıldığı durumda;

$$P(X = x | C = c) = \prod_i^n P(X_i = x_i | C = c) \quad (2.12)$$

Eşitlik 2.12 elde edilir. En yüksek olasılığa sahip olan sınıfı tahmin edilen sınıf olarak atanır (89).

2.21. K-En Yakın Komşu (KNN)

KNN algoritması, hem sınıflama hem regresyon problemlerinde kullanılan temel örnek tabanlı algoritmaların başında gelmektedir. Örnek tabanlı algoritmalarda, öğrenme eğitim setindeki verilere dayalı olarak gerçekleşir. Yeni karşılaşılan bir örneğin sınıfını eğitim veri setindeki en yakın k komşuya bakarak belirlemektedir (90). Temel KNN algoritmasında, sınıf etiketlerinin çoğunluğa göre belirlenmesi, yeni örneklerin sınıflarının belirlenmesinde simetrik olmayan veri setlerindeki çoğunluğu oluşturan sınıfların rolünün baskın olmasına sebep olmaktadır (91). Bu sebeple uzaklık ölçütünün etki değerinin ağırlıklandırılması için farklı yöntemler kullanılmaktadır (92).

KNN algoritması basit yapıda olmasına rağmen hesaplama maliyeti yüksektir. Sınıf etiketi belirlenmek istenen örneklerin arasındaki mesafelerin belirlenmesi, büyük hacimdeki veri kümeleri için oldukça zor bir işlemdir. Bu işlemi kolaylaştırabilmek için, KNN ile birlikte boyut azaltma veya arama ağaçları yöntemlerine başvurulabilir (93).

KNN algoritmasında kullanılan uzaklık ölçütü (*Minkowski*, *Öklid*, *Manhattan*, *Chebyshev* ve *Dilca* vb.), komşu sayısı (k) ve ağırlıklandırma yöntemleri performansı belirleyen faktörlerdendir. Sınıflandırmada veriler, komşu sayısının 1 e eşit olduğu durumlarda sadece en yakın komşu sınıfına atanırken, komşu sayısı örnek sayısına (N) yaklaştıkça veri setinin hepsi hesaplama dahil edilerek oylama ile seçim yapılmaktadır. Komşular için ağırlık değerleri atanarak, örneğe en yakın olan komşuların, çoğunluk oylamasındaki katkısının daha fazla olması sağlanır. Komşunun ağırlıkları, komşuların birbirine olan uzaklıkları d olması koşuluyla, $1/d$ ya da $1/d^2$ şeklinde alınır (94).

2.22. Sınıf Dengesizliği

Bir sınıflandırma probleminde bazı sınıfların diğerlerine göre çok daha fazla örneğe sahip olması durumudur. Özellikle azınlık sınıfın doğru şekilde tahmin edilmesini güçleştirmektedir.

2.22.1. SMOTE (Synthetic Minority Over-sampling Technique)

Sınıf dengesizliği problemiyle başa çıkmak amacıyla geliştirilmiş en yaygın aşırı örnekleme (oversampling) tekniklerinden biri SMOTE (Synthetic Minority Over-sampling Technique) yöntemidir. Bu yöntem, azınlık sınıfına ait örneklerin sayısını artırmak için mevcut verilerden doğrudan kopyalar üretmek yerine, yeni ve sentetik veri örnekleri oluşturur. Böylece modelin azınlık sınıfı üzerinde aşırı uyum (overfitting) yapması önlenir ve sınıflar arasındaki dağılım daha dengeli hale getirilir (95).

Her azınlık sınıfı örneği için, veri kümesinde bulunan aynı sınıftan k adet en yakın komşu belirlenir (genellikle $k=5$ kullanılır). Bu işlem genellikle Öklid uzaklığı temel alınarak gerçekleştirilir. Rastgele bir komşu seçilir ve seçilen komşu ile mevcut örnek arasındaki fark alınarak, bu farkın rastgele bir çarpanı kadar ilerlenir. Matematiksel olarak bu işlem;

$$x_{\text{sentetik}} = x_{\text{orijinal}} + \lambda \times (x_{\text{komşu}} - x_{\text{orijinal}})$$

şekilde ifade edilir.

Belirlenen miktarda sentetik örnek üretildikten sonra, bu yeni sentetik örnekler orijinal veri setine eklenir. Böylece azınlık sınıfı örnek sayısı artırılmış olur ve sınıf dağılımı daha dengeli hale gelir. Veri çoğaltma yerine çeşitlilik sağlayan, aşırı uyum riskini azaltan basit ve etkili bir yöntem olmasının yanı sıra sınıf sıralarında gürültü oluşumuna ve aykırı değerlerin etkisi ile performansın düşmesine sebep olabilir (96).

2.23. Model Performans Değerlendirme Ölçütleri

Veri madenciliği yöntemleri kullanılarak oluşturulan kestirim ya da sınıflama modellerinin doğru sonuç verme performanslarının değerlendirilmesi amacıyla modellerin performansları karşılaştırılarak en iyi ve en güvenilir tahmin modeli seçilir.

2.23.1. Sınıflama Modelleri Performans Değerlendirme Ölçütleri

Sınıflama modellerinde performans ölçütleri olarak,

- Doğruluk (Accuracy),
- Duyarlılık (Sensitivity),
- Seçicilik (Specificity),
- Pozitif Tahmin Değeri (Positive Predictive Value),
- Negatif Tahmin Değeri (Negative Predictive Value),
- F1 Ölçüsü
- ROC Eğrisi Altında Kalan Alan (Area Under The ROC Curve)
- Kappa Katsayısı

kullanılmaktadır.

Yukarıda bahsedilen ve eğri altında kalan alan dışındaki model performans ölçütleri, gerçek ve tahmin sınıflarının çapraz tablosu şeklinde hazırlanan bir tablo (confusion matrix) yardımıyla hesaplanır (68,97). İki sınıflı veri setine ait 2x2 lik çapraz tablo Tablo 2.2’de verilmiştir.

Tablo 2.2. İki sınıf içeren veri setine ait 2x2 çapraz tablo

Gerçek Durum		
Tahmin Edilen Sınıf	Pozitif	Negatif
Pozitif	DP	YP
Negatif	YN	DN

Doğru Pozitif (DP) : Gerçek ve tahmin edilen değerleri pozitif olanların sayısı

Doğru Negatif (DN) : Gerçek ve tahmin edilen değerleri negatif olanların sayısı

Yanlış Pozitif (YP) : Gerçek negatif iken tahmin edilen değeri pozitif olanların sayısı

Yanlış Negatif (YN) : Gerçekte pozitif iken tahmin edilen değeri negatif olanların sayısı

Doğruluk (Accuracy): Doğru pozitif ve doğru negatif sonuçların toplamının çalışmadaki tüm gözlemlerin toplamına oranıdır. [0,1] aralığında değişen doğruluk değeri 1'e yaklaştıkça sınıflandırma performansının yükseldiği anlamına gelmektedir (68,98).

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (2.13)$$

Duyarlılık (Sensitivity): Gerçekte pozitif olan bir gözlemin tahmin değerinin de pozitif çıkması olasılığıdır. [0,1] aralığında değişen duyarlılık değeri 1'e yaklaştıkça sınıflandırma performansının yükseldiği anlamına gelmektedir (68,98).

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (2.14)$$

Seçicilik (Specificity): Gerçekte negatif olan bir gözlemin tahmin değerinin de negatif çıkması olasılığıdır. [0,1] aralığında değişen seçicilik değeri 1'e yaklaştıkça sınıflandırma performansının yükseldiği anlamına gelmektedir (68,98).

$$\text{Seçicilik} = \frac{DN}{YP+DN} \quad (2.15)$$

Pozitif Kestirim Değeri (Kesinlik/Precision): Pozitif olarak tahmin edilen bir değer gerçekte de pozitif olma olasılığıdır. [0,1] aralığında değişen kesinlik değeri 1'e yaklaştıkça sınıflandırma performansının yükseldiği anlamına gelmektedir (98).

$$\text{Pozitif Kestirim Değeri} = \frac{DP}{DP+YP} \quad (2.16)$$

Negatif Kestirim Değeri: Negatif olarak tahmin edilen bir değer gerçekte de negatif olma olasılığıdır. [0,1] aralığında değişen negatif kestirim değeri 1'e yaklaştıkça sınıflandırma performansının yükseldiği anlamına gelmektedir (98).

$$\text{Negatif Kestirim Değeri} = \frac{DN}{DN+YN} \quad (2.17)$$

F1 Ölçüsü: Kesinlik ve duyarlılık ölçütlerinin harmonik ortalaması F ölçüsünü verir. [0,1] aralığında değişen F ölçüsü 1'e yaklaştıkça sınıflandırma performansının yükseldiği anlamına gelmektedir.

$$F \text{ Ölçüsü} = 2 * \frac{\text{Pozitif Kestirim Değeri} * \text{Duyarlılık}}{\text{Pozitif Kestirim Değeri} + \text{Duyarlılık}} \quad (2.18)$$

ROC Eğrisi Altında Kalan Alan (Area Under The ROC Curve): ROC (Receiver Operating Characteristics) eğrisinin çizilebilmesi için her bir gözlem birimine ait sınıflama olasılıkları büyükten küçüğe doğru sıralanarak bu değerler eşik değerler olarak kullanılır. Her bir eşik değer için x ekseninde yanlış pozitif oranı (1-seçicilik) ve y ekseninde doğru pozitif oranı (duyarlılık) yerleştirilerek eğri (0,0) ve (1,1) noktaları arasında iki boyutlu bir grafik olarak çizilir. Eğri altında kalan alan (EAA) 0 ile 1 arasında değer alır ve sınıflandırıcının yanlış pozitif oran ile doğru pozitif oran arasındaki dengeyi ne kadar iyi yakaladığını ölçer. Bu değer 1'e yaklaştıkça sınıflama modeli başarısının arttığı anlamına gelmektedir (99,100).

$$\text{Eğri Altında Kalan Alan} = \int_0^1 \widehat{ROC}(t) dt \quad (2.19)$$

Kappa Katsayısı:

Kappa katsayısı, gerçek ile tahmin edilen değerler arasındaki uyumu ölçer. [-1, +1] aralığında değer alır. Gerçek sınıfların oranı (P_0) Eşitlik 2.20, beklenen sınıfların oranı (P_e) Eşitlik 2.21, Kappa katsayısı ise Eşitlik 2.22' de verilen formüller yardımı ile hesaplanır.

$$P_0 = \sum_{r=1}^R \frac{n_{rr}}{n}, r = 1, \dots, R (\text{Satır veya Sütun Sayısı}) \quad (2.20)$$

$$P_e = \sum_{r=1}^R \frac{n_r \cdot n_r}{n^2} \quad (2.21)$$

$$K = \frac{P_0 - P_e}{1 - P_e} \quad (2.22)$$

Kappa katsayısı K ;

$0,75 < K \leq 1$ ise uyum çok iyi,

$0,41 < K \leq 0,75$ ise uyum iyi,

$0 < K \leq 0,41$ ise uyum zayıf,

$K = 0$ ise bağımsız, $K < 0$ ise uyumsuz yorumlanmaktadır (101,102).

2.24. Spor Bilimlerinde Veri Madenciliği Uygulamaları

Elit ve rekabetçi sporlarda büyük verinin keşfi, önemli avantajlar sağlamaktadır (103). Geleneksel olarak sezgi ve gözleme dayalı karar verme süreçleri, yerini bilimsel araştırmalar ve kapsamlı analizlere dayanan modern bir yaklaşıma bırakmıştır (104). Sporda büyük veri kullanımının temel amaçları arasında sporcu performansını değerlendirmek, sakatlık risklerini öngörmek (1), yetenekli sporcuları seçmek ve antrenman programlarını bireyselleştirmek yer almaktadır (105). Bununla birlikte, bu büyük veri yığınlarının basit istatistiksel yöntemlerle yorumlanması oldukça zorlaşmıştır. Bu nedenle, son yıllarda spor bilimlerinde veri madenciliği ve makine öğrenmesi teknikleri, büyük veriden anlamlı sonuçlar elde etmek için yaygın bir şekilde kullanılmaya başlanmıştır (106).

Veri madenciliğinin spor alanında kullanımı ülkemizde az sayıda; uluslararası yayınlarda ise oldukça fazla bilimsel çalışmaya rastlanmaktadır. Son yıllarda yapılmış çalışmalarda, müsabaka (oyun) verileri, sporcuların müsabaka içerisindeki performans verileri, tahmin analizleri, sakatlık risk analizleri üzerine odaklanılmıştır (1,107).



Şekil 2.11. Veri madenciliğinin sporda kullanım alanları (108).

Veri madenciliğinin spor alanında kullanımları Şekil 2.11’de verilmiştir. Spor arşivlerindeki müsabaka sonuçları veya sporcu performanslarının değerlendirilmesiyle geleceğe yönelik müsabaka sonuçları tahminlerinde, sporcu seçimi ve sporcu performanslarının değerlendirilmesi, müsabaka stratejilerinin belirlenmesinde, sporcu, araç-gereç, antrenör ve teknik direktör seçimlerinde, sporcuların sakatlanma risk analizlerinde, kısa, orta veya uzun vadeli alt yapı projelerinin hazırlanmasında, bilet satış tahminlerinde, lig sıralamaları tahminleri ve bahis oyunlarında, sporcu ve takım performans değerlendirilmesi ve sporcu seçimi, spor tesislerinin yapısı ve konumunun belirlenmesinde, organizasyonlar ve sponsorluk anlaşmalarında veri madenciliği yöntemlerinden faydalanılmaktadır (108).

DiCesare ve diğerleri, futbol yaralanmalarını tahmin etmeye yönelik 2020 yılında yapmış oldukları bir çalışmada, 22 kadın lise futbolcusu için baş darbesine maruz kalmanın veri odaklı filtrelemesinin maruziyetin ölçümündeki doğruluğunu ve beyin değişiklikleri ile ilişkilerini incelemek için eşik tabanlı, sezgisel filtreleme ve XGBoost sınıflandırıcısını kullanmışlardır (109). Kullanılan yöntemlerin gerçek darbeleri tespit edebilme yeteneğinin doğrulanması için giyilebilir sensör tarafından kaydedilen darbelerin bir alt kümesinin eş zamanlı video kaydı kullanılarak

belirlenmiş ve filtrelenmiş veri seti sporcuların sezon öncesi ve sonrası MRI beyin taramaları ile ilişkilendirilerek uzunlamasına beyaz cevher değişikliklerini ortaya çıkarmıştır. Eşik tabanlı, sezgisel ve ML yaklaşımları sırasıyla %22,0 doğruluk, %44,6 doğruluk ve %83,5 doğruluk elde etmiştir. Başa alınan darbelerin sayısının boylamsal makine öğrenimi yoluyla beyindeki değişiklikleri yüksek doğruluk ile öngörebildiğini ortaya koymuştur.

Futbolda performansı tahmin etmeye yönelik 2019 yılında yapılan bir çalışmada 2017-2018 sezonu İngiltere Premier Ligi'ndeki maç sonuçlarından yola çıkarak; rakip kalitesine bağlı olarak maç sonucunu (galibiyet, mağlubiyet veya beraberlik) önemli ölçüde etkileyen durumsal değişkenler ve performans göstergelerini belirlemeye çalışmışlardır (2). Analizlerde k-ortalamalar kümeleme ve karar ağacı algoritmaları kullanılmış ve her karar ağacında, ilk golü atma maç sonucu üzerinde en etkili faktör olarak bulunurken, uzaklaştırma, şut, kaleyi bulan şut, topa sahip olma yüzdesi ve maç lokasyonunun maç sonucu üzerindeki etkileri, rakibin kalitesine göre değişmekte olduğu sonucuna ulaşmışlardır. Rakibin kalitesine dayalı karar ağaçları, dengeli, daha güçlü ve daha zayıf rakiplere karşı oynanan oyunlarda sırasıyla %67,9, %73,9 ve %78,4'lük sonuçları doğru tahmin ederken, tüm oyunlarda (rakibin kalitesine bakılmaksızın) bu oran sadece %64,8'dir. Bu sonuç ise rakibin kalitesinin dikkate alınmasının önemini ve göstermektedir.

Ćwiklinski ve diğerleri (110); futbolda yetenekleri ve oyuncu transferinin başarısını tahmin etmeye yönelik 2021 yılında yaptıkları bir çalışmada, en popüler 8 lige ait 156 klüpten 4700 oyuncuya ait 29 teknik, 7 psikolojik, ve 4 sonuç değişkeni (oynanan, kazanan maçlar vs.) verileri kullanılarak rastgele orman, Naive Bayes ve adaboost algoritmaları ile yapılan tahminler sonucunda oluşturulan model (doğruluk = 0,82, kesinlik = 0,84, duyarlılık = 0,82 ve F1 puanı = 0,83) yüksek tahmin performansı ile transfer başarısında etkili sonuçlar ortaya konulabilmesi açısından yardımcı olunabileceği sonucunu ortaya koymuştur.

Miljković ve diğerleri (111); 2010 yılında yapmış oldukları bir çalışmada 2009-2010 sezonunda oynanan 778 NBA oyunlarının istatistiksel verileri Yahoo web sitesinden toplanmıştır. Naive Bayes sınıflandırma yöntemi ile maç sonuç tahmini

yapılmıştır ve maçlarının galibi %67 doğruluk oranı ile tahmin edilmiştir. Diğer sistemler ile karşılaştırıldığında çok iyi bir sonuç olduğunu savunmaktadırlar. Yayılma tahmin sonucu ise, sistem 778 maçta 78'i için başarı göstermiştir(~10%). Bu sonucun beklenen bir sonuç olduğunu çünkü basketbol oyununun hangi farkla biteceğini tam olarak tahmin etmenin zor olduğunu çalışmalarının amacının olası fark hakkında yalnızca yaklaşık bilgi sağlamak olduğunu belirtmişlerdir.

Bu tez kapsamında, judo ve güreş branşında elit ve elit olma potansiyeli taşıyan sporcuların sınıflandırılmasında veri madenciliği yöntemlerinin uygulanması hedeflenmektedir.



3. GEREÇ ve YÖNTEM

Bu çalışma; ülkemizde faaliyet gösteren, elit ve elit olmaya aday, en az 5 yıldır müsabaka geçmişi olan sporcuların performans ölçümlerinin yapıldığı bir merkezde, 01.12.2022-31.12.2023 tarihleri arasındaki sporcu performans ölçüm verilerini kapsamaktadır.

Araştırma için “Sağlık Bilimleri Araştırma Etik Kurulu”na başvurulmuş, 09.01.2024 tarih ve 2024/01-11 sayılı yazı ile araştırmanın etik açıdan uygunluğu belirtilmiştir. İlgili yazı Ek-1 olarak paylaşılmıştır.

Verilerin edinildiği laboratuvarın bağlı bulunduğu kurumdan gereken izinler alınmış ve kurumun ismi saklı olacak şekilde “ölçüm laboratuvarı” olarak belirtilmiştir.

Araştırma Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı’nda yürütülmüş olup, analizlerde R 4.2.1 ve RStudio yazılımları ve bu yazılımlarda *plyr*, *caret*, *caretEnsemble*, *party*, *rpart*, *rpart.plot*, *e1071*, *randomForest*, *naivebayes*, *klaR*, *RWeka*, *kernelab*, *Liblinear*, *AdaBag*, *rio*, *missRanger*, *pROC*, *VIM* ve *UBL* paketleri kullanılmıştır.

Araştırma kapsamında; sportif performans ölçümlerine katılan 66 elit ve 122 elit olmaya aday sporcuya ait veriler analiz edilmiştir. Ölçüm laboratuvarında performans ölçüm testlerine giren 188 sporcunun bilgileri, antropometrik ölçümleri ve 5 farklı performans test verileri ve açıklamaları Tablo 3.1’de verilmiştir.

Tablo 3.1. Araştırmada kullanılan değişkenler ve açıklamaları.

Değişken Adı	Açıklama
KATEGORİ	Sporcunun kategorisi, Elit ya da Aday
YAŞ	Sporcunun performans ölçüm günü hesaplanan yaşı
CİNSİYET	Sporcunun cinsiyeti, kadın/erkek
SPOR_YILI	Sporcunun ölçüm tarihi öncesi aktif olarak spor yaptığı yıl sayısı
BOY	Sporcunun performans ölçümüne kabul tarihi itibariyle ölçülen boyu (cm)
KILO	Sporcunun performans ölçümüne kabul tarihi itibariyle ölçülen vücut ağırlığı (kg)
IMTP	İzometrik Uyluk Çekiş Gücü (Newton)
ÜPP	Üst ekstremitte wingate anaerobik güç testi zirve güç değeri (Watt)
ÜAP	Üst ekstremitte wingate anaerobik güç testi ortalama güç değeri (Watt)
ÜPD	Üst ekstremitte wingate anaerobik güç testi yorgunluk yüzdesi değeri
ÜTPP	Üst ekstremitte wingate anaerobik güç testi zirve güce ulaşma süresi (sn)
APP	Alt ekstremitte wingate anaerobik güç testi zirve güç değeri (Watt)
AAP	Alt ekstremitte wingate anaerobik güç testi ortalama güç değeri (Watt)
APD	Alt ekstremitte wingate anaerobik güç testi yorgunluk yüzdesi değeri
ATPP	Alt ekstremitte wingate anaerobik güç testi zirve güce ulaşma süresi (sn)
ÇAL_SÜRE	Çok yönlü reaktif çeviklik testi bitirme süresi (sn)
DÖNÜŞ_SÜRE	Çok yönlü reaktif çeviklik testi ortalama dönüş süresi (sn)
SÜRE	Aerobik güç ve kapasite testi (MAXVO ₂) bitirme süresi (dk)
VE_L_MİN	Aerobik güç ve kapasite testi (MAXVO ₂) Dakika ventilasyonu değeri (litre/dk)
VO ₂	Aerobik güç ve kapasite testi (MAXVO ₂) dakika başı oksijen tüketimi değeri (ml/dk)
RQ	Aerobik güç ve kapasite testi (MAXVO ₂) karbondioksit üretiminin oksijen tüketimine oranı
VO ₂ _KG	Aerobik güç ve kapasite testi (MAXVO ₂) relatif oksijen tüketim miktarı (ml/dk/kg)
BPM	Aerobik güç ve kapasite testi (MAXVO ₂) maksimum kalp atım hızı (bpm)
HIZ	Aerobik güç ve kapasite testi (MAXVO ₂) maksimum ulaşılan hız değeri (km/saat)
AT_HIZ	Aerobik güç ve kapasite testi (MAXVO ₂) anaerobik eşiğe ulaşıldığında ölçülen hız (km/saat)

Analiz aşamasında sporcuların elit olma tahminleri yapılmıştır. Elit olma tahmin çalışmasında 10 ayrı veri madenciliği sınıflama modeli kurulmuştur. Tablo 3.1’de yer alan kategori değişkeni bağımlı, diğer değişkenler ise bağımsız değişken olarak ele alınmıştır. Sınıf kestirimi amacıyla yirmi üç nicelik, iki nitelik değişkeni olmak üzere toplamda yirmi beş değişken kullanılmıştır.

Değişken Seçimi:

Değişken seçimi, korelasyon matrisi incelenerek gerçekleştirilmiştir. Elde edilen özelliklerin korelasyon matrisi incelenmiş ve reaktif çeviklik testine ait soldan ve sağdan dönüş süreleri, ortalama dönüş süresi ile sırasıyla 0,90 ve 0,91 korelasyon değerlerine sahip olduğu gözlemlenmiştir. Performans ölçümlerini yapan uzman ekibe danışılarak sağdan ve soldan dönüş süreleri çoklu bağlantı oluşturma durumuna karşın analiz dışı bırakılmış, ortalama dönüş süresi analize dahil edilmiştir.

Aynı şekilde aerobik güç ve kapasite testine ait dakika ventilasyon hacmi ve oksijen tüketimi değerleri, maksimum oksijen tüketimi ile sırasıyla 0,83 ve 0,94 korelasyon değerlerine sahip olduğu gözlemlenmiştir. Performans ölçümlerini yapan uzman ekibe danışılarak dakika ventilasyon hacmi ve oksijen tüketimi çoklu bağlantı oluşturma durumuna karşın analiz dışı bırakılmış, maksimum oksijen tüketimi analize dahil edilmiştir.

Korelasyon katsayısına göre analiz dışında bırakılan, yüksek derecede ilişki gösteren değişkenler ve ilişki katsayıları Tablo 3.2’de verilmektedir.

Tablo 3.2. Analizden dışlanan/dahil edilen değişkenler ve korelasyon katsayıları

Analiz dışı bırakılan değişkenler	Korelasyon Katsayısı	Analize dahil edilen değişkenler
Soldan Dönüş Süresi	0,9082	Ortalama Dönüş Süresi
Sağdan Dönüş Süresi	0,9123	Ortalama Dönüş Süresi
VE	0,8321	MaxVo2
VO2	0,9438	MaxVo2

Eksik Gözlem Atama:

Gerçek performans test ölçüm sonuçlarından oluşan veri setinde %28 oranında IMTP test verisine ulaşılamamıştır. Eksik gözlemlere, üç farklı yöntem ile veri atama işlemi yapılmıştır. İlk yöntem olarak pratikte çok kullanılan bir yöntem olan grup ortalaması kullanılarak eksik gözlemlere değer atanmıştır. Bulunduğu kategoriye, cinsiyet ve branşına göre grup ortalaması değeri atanarak her bir sporcunun analize dahil edilmesi sağlanmıştır.

İkinci yöntem olarak eksik gözlemlere RO algoritması, üçüncü yöntem olarak ise KNN algoritması kullanılarak değer ataması yapılmış ve üç farklı veri seti ile analizler gerçekleştirilmiştir.

Eksik gözlem atama işlemi, modelin doğruluğunu arttırmak ve eksik veriden kaynaklı oluşabilmesi olası yanlışlıkları azaltmak için yapılan bir işlem adıdır. Eksik gözlemlerin veri setimizdeki oranı IMTP test verisinde %28, Üst Ekstremitte Anaerobik Güç Testine ait Zirve Güç verisinde ise %2 oranında tamamen rastgele (MCAR) bir eksiklik söz konusudur. Arşiv verilerinde bu oranda sporcunun IMTP verilerine ulaşılamamıştır. Bu durumda, Rastgele Orman ve KNN gibi yöntemler ile güvenilir bir şekilde eksik veri ataması yapılabilmektedir.

RO ile eksik gözlem atama yönteminde, eksik olmayan veriler ile model eğitilir ve eksik olan değişken hedef değişkeni kabul edilir. Model eksik değerleri tahmin eder.

KNN ile eksik gözlem atama yönteminde, eksik değer içeren gözlem ile en yakın diğer gözlemler bulunur ve belirlenen bu gözlemlerin değerleri kullanılarak eksik olan değer tahmin edilir.

SMOTE ile Dengeleme Analizi:

Veri setinde Elit gruptan 66, Aday gruptan ise 122 sporcu bulunmaktadır. Sınıf dengesizliğinin model performansı üzerindeki etkisinin giderilebilmesi açısından SMOTE (Synthetic Minority Over-sampling Technique) yöntemi uygulanmıştır. Bu

yöntem ile azınlık sınıfa ait örnekler arasındaki komşuluk ilişkilerine dayalı olarak yeni sentetik örnekler oluşturulmuş ve sınıflardaki gözlem sayıları eşitlenmiştir. R programlama dilinde *UBL* paketi ile sentetik örnek oluşturma işlemi sadece eğitim seti için gerçekleştirilmiş, test veri setine sentetik örnek atanmamıştır. Bu işlem model test performansı diğer yöntemlere kıyasla yüksek çıkan, RO ile eksik gözlemlere atama yapılmış veri setinden rastgele seçilen eğitim setinde uygulanmıştır.

Oluşturulan Modeller ve Hiperparametre Ayarları:

İlk aşamada C4.5 sınıflandırması için *J48* karar ağacı algoritması *RWeka* paketiyle, DVM algoritması ise *e1071* paketi ile gerçekleştirilmiştir. Lojistik regresyon algoritması için yine *RWeka* paketi kullanılmıştır. NB sınıflandırması için ise *klaR* paketi kullanılmıştır.

KNN, RO, CART, CARTBAGGING, XGBOOST ve ADABAG sınıflandırıcıları için *caret* paketi kullanılmıştır.

C4.5 modeli, güven aralığı $c=0,25$ ve her yapraktaki minimum gözlem sayısı=2 olarak kullanılmış ve karar ağaçlarında budama yapılmamıştır.

KNN modeli, 10 katlı çapraz doğrulama ile ($cv=10$) ile eğitilmiştir. $k=5, 7, 9$ değerleri ile eğitilip, doğruluğu en yüksek olan $k=5$ değerinde optimum modele ulaşılmıştır.

LRM modeli, modelleme performansını arttırmak ve aşırı uyumu engellemek için 10 katlı çapraz doğrulama yapılmıştır. Herhangi bir manuel hiperparametre seçimi yapılmamıştır.

CART modeli, karmaşıklık parametresi $cp=0,01$ alınmış ve karar ağaçlarında budama yapılmamıştır.

RO modeli, 10 katlı çapraz doğrulama ile ($cv=10$) ile eğitilmiştir. Varsayılan olarak ağaç sayısı 500 veya üzerinde oluşturulmuştur, rassal olarak seçilen değişken sayısı $mtry=5$ olarak belirlenmiştir.

NB modeli, Laplace düzeltmesi=0, dağılım tipi parametresi $usekernel=F$ olarak alınmıştır.

DVM modeli $cost=64$ ve $gamma=0,016$ olarak alınmıştır.

ADABOOST modeli, 5 katlı çapraz doğrulama ile $mfinal=50$ ve $maxdepth=3$ olarak alınmıştır

XGBOOST modeli, 10 farklı parametre kombinasyonunun denenmesi sağlanmış ve 50 iterasyonla eğitilmiş her iterasyonda kullanılan örnekleme oranı %66.6, maksimum ağaç derinliği 9, öğrenme hızı (eta) 0,4, ağaç bölme sırasında herhangi bir ek düzenleme uygulanmamış, ağaçlar için kullanılan rastgele özellik oranı %60, her yaprak için minimum 1 örnek kullanılmış ve optimizasyonu sağlanmıştır.

İkinci ve üçüncü aşamada modellerin karşılaştırılmasının daha sağlıklı yapılması adına her bir model için *caret* paketi kullanılmıştır. RO ve KNN ile veri atandıktan sonra *TuneGrid* veya *TuneLenght* fonksiyonu ile hiperparametre kombinasyonları denenerek model optimizasyonu gerçekleştirilmiştir. 10 kat 3 tekrarlı çapraz geçerleme yöntemi kullanılmıştır.

Son aşamada RO ile eksik gözlem atanan veri setinde smote sınıf dengeleme analizi yapılmıştır.

Model Seçimi:

Oluşturulan tahmin modelleri arasından en iyi performans gösteren modeller belirlenmiş ve bu modellere göre değişken önem dereceleri dikkate alınarak en önemli faktörler belirlenmiştir.

4. BULGULAR

Sporcu performans testlerine katılan sporcuların elit olma tahminine yönelik sınıflama algoritmalarından C4.5, KNN, LRM (regulaized), CART, CART BAGGING, RO, XGBoost, NB, AdaBag, DVM (Radial) modelleri olmak üzere on ayrı model kullanılmıştır. Elit olma tahmininde kullanılan veri setine ait tanımlayıcı istatistikler sporcu kategorilerine ve branşlarına göre dağılımları sırasıyla Tablo 4.1 ve Tablo 4.2’de gösterilmektedir.

Tablo 4.1. Sporcu verilerine ait tanımlayıcı istatistikler.

Değişkenler	Elit (n=66)				Aday (n=122)			
	Erkek (n=41)		Kadın (n=25)		Erkek (n=91)		Kadın (n=31)	
	Ort.	SS.	Ort.	SS.	Ort.	SS.	Ort.	SS.
Spor yılı (yıl)	13,61	4,5	9,12	4,0	7,93	1,8	7,65	1,5
Yaş (yıl)	24,66	3,6	21,12	3,0	17,28	2,3	18,11	2,0
Boy (cm)	174,69	6,6	162,25	6,7	173,25	7,3	163,00	6,8
Kilo (kg)	82,38	16,3	61,41	8,5	75,57	20,6	62,64	10,0
IMTP (N)	2242,56	234,1	1809,52	160,2	1973,20	220,9	1839,52	199,9
ÜPP (W)	1025,40	281,6	559,00	123,0	830,19	232,1	491,81	69,8
ÜAP (W)	581,75	154,2	323,26	51,0	444,10	89,5	302,92	40,3
ÜPD (%)	73,04	16,6	68,98	14,9	81,28	12,9	70,86	12,1
ÜTPP (sn)	1,63	0,9	2,56	2,6	1,92	1,4	1,98	1,1
APP (W)	1057,30	246,6	664,77	113,5	914,92	209,0	617,01	78,9
ATPP (sn)	2,93	3,5	2,79	1,6	3,70	4,2	2,95	3,8
AAP (W)	748,86	123,2	481,97	65,2	615,58	119,3	436,27	56,3
APD (%)	52,70	16,2	47,44	13,3	59,03	11,5	54,83	10,1
Çalışma Süresi (SN)	590,65	3608,7	28,45	2,1	36,59	12,9	44,73	14,3
Ortalama dönüş süresi (SN)	0,40	0,1	0,44	0,1	0,42	0,1	0,52	0,1
VE_L_MİN	140,82	20,9	105,31	14,8	137,47	27,6	104,95	14,0
VO2	4153,57	672,7	2968,15	339,9	3998,85	726,1	2804,60	602,3
RQ	1,14	0,1	1,13	0,1	1,13	0,1	1,09	0,1
VO2 (ml/dk/kg)	52,15	7,6	48,95	5,1	53,97	7,9	46,75	11,6
MAX-KAH (bpm)	184,12	9,2	193,88	7,7	187,93	9,0	189,61	8,2
MAX-HIZ (km/s)	14,53	1,4	13,64	0,8	14,05	1,1	12,61	1,0
AT-HIZ (km/s)	10,80	1,3	10,14	0,8	10,23	1,3	8,97	1,0

Tablo 4.2. Branşlara göre sporcu verilerine ait tanımlayıcı istatistikler.

Değişkenler	Branşlar		P değerleri
	Güreş	Judo	
Cinsiyet			0,020*
Erkek	62 (%63,9)	72 (%79,1)	
Kadın	35 (%36,1)	19 (%20,9)	
Elit	44	22	
Spor yılı (yıl)	10±4	9±3	0,053
Yaş (yıl)	22±4	18±4	<0,001*
Boy (cm)	168±7	173±9	<0,001*
Vücut ağırlığı (kg)	50,29±8,17	53,21±8,9	0,02*
IMTP (N)	2073,59±278,18	1946,89±311,82	0,003*
ÜPP (W)	764,2±290,4	802,3±283,9	0,36
ÜAP (W)	480,37±67,51	460,00±66,19	0,038*
ÜPD (%)	76,43±12,07	75,79±17,24	0,77
ÜTPP (sn)	1,90±1,61	2,00±1,34	0,64
APP (W)	882,14±230,91	843,80±265,03	0,29
ATPP (sn)	2,52±2,62	4,10±4,48	0,0039*
AAP (W)	601,08±132,96	593,28±166,44	0,72
APD (%)	58,77±11,36	51,83±14,23	<0,001*
Çalışma Süresi (SN)	28,00±2,30	27,23±2,53	0,03*
Ortalama dönüş süresi (SN)	0,464±0,11	0,402±0,08	<0,001*
RQ	1,11±0,093	1,135±0,084	0,054
MAX-KAH (bpm)	186±9	190±9	0,0026*
MAX-HIZ (km/s)	13,85±1,28	13,87±1,23	0,91
AT-HIZ (km/s)	10,06±1,29	10,20±1,37	0,47

Değişken isimleri Tablo 3.1’de ayrıntılı olarak belirtilmiştir.

Araştırma kapsamında, toplam 188 sporcudan 66’sı elit, 122’si ise aday kategorisindedir. Elitlerin 22’si judo, 44’ü güreş branşında; adayların ise 69’u judo, 53’ü güreş branşında yer almaktadır.

Tüm verilerin rastgele seçilen %75’i (141 sporcu) eğitim veri seti ve %25’i (47 sporcu) test veri seti olarak ayrılmıştır. Eğitim veri seti kullanılarak oluşturulan modellerin, test veri seti üzerinde model performans ölçütleri elde edilmiştir.

Modellere ait test performans ölçütleri paylaşılmıştır. Eğitim ve test seti bulguları arasında yüksek bir performans farkına rastlanmamıştır. Sınıflamada kullanılan modellere ait çapraz tablo sonuçları ve modellere ait test verisinden elde edilen performans ölçüleri sırasıyla Tablo 4.3 ve Tablo 4.4'te gösterilmiştir.

Tablo 4.3. Eksik gözlemlere aritmetik ortalama ile atama yapılan veri setinden elde edilen sınıflama modellerine ait çapraz tablo.

Tahmin Modeli		Gerçek Durum	
		Aday	Elit
C4.5	Aday	30	6
	Elit	4	7
KNN	Aday	29	9
	Elit	5	4
LRM	Aday	27	3
	Elit	7	10
CART	Aday	29	7
	Elit	5	6
RO	Aday	33	4
	Elit	1	9
XGBOOST	Aday	32	4
	Elit	2	9
NB	Aday	33	9
	Elit	1	4
ADABAG	Aday	33	5
	Elit	1	8
DVM	Aday	28	2
	Elit	6	11
CART (Bagging)	Aday	31	5
	Elit	3	8
TOPLAM		34	13

Çapraz tablolar gerçek durumla modelin tahmini arasındaki uyumu gösteren bir karşılaştırma tablosudur. Çapraz tablo kullanılarak modellere ait performans göstergeleri ve en iyi performansa sahip modelin sınıflandırma algoritması yorumlanmıştır.

Tablo 4.4. Elit olma sınıflama modellerine ait test performans ölçütleri (eksik gözlemlere aritmetik ortalama ile atama yapılan veri setinde).

Model	C4.5	KNN	CART	CART (Bagging)	RO	LRM	NB	AdaBag	DVM	XGBoost
Doğruluk	0,7872	0,7021	0,7447	0,8298	0,8936	0,7872	0,7872	0,8723	0,8298	0,8723
% 95 GA (Alt-Üst)	(0,643, 0,893)	(0,551, 0,827)	(0,597, 0,861)	(0,6919, 0,9235)	(0,769, 0,965)	(0,643, 0,893)	(0,643, 0,893)	(0,743, 0,952)	(0,692, 0,924)	(0,743, 0,952)
Kappa	0,4418	0,1775	0,3302	0,5534	0,7138	0,5145	0,3436	0,6475	0,6116	0,6651
Duyarlılık	0,5385	0,3077	0,4615	0,6154	0,6923	0,7692	0,3077	0,6154	0,8462	0,6923
Seçicilik	0,8824	0,8529	0,8529	0,9118	0,9706	0,7941	0,9706	0,9706	0,8235	0,9118
PKD	0,6364	0,4444	0,5455	0,7273	0,9000	0,5882	0,8000	0,8889	0,6471	0,8182
NKD	0,8333	0,7631	0,8056	0,8611	0,8919	0,9000	0,7857	0,8684	0,9333	0,8889
Dengeli Doğruluk	0,7104	0,5803	0,6572	0,7636	0,8314	0,7817	0,6391	0,7930	0,8348	0,8167
F1 Ölçüsü	0,5833	0,3636	0,4999	0,6667	0,7826	0,6666	0,4444	0,7273	0,7333	0,7199
EAA	0,629	0,712	0,6651	0,8812	0,926	0,9380	0,6855	0,8812	0,9286	0,9457

Tablo 4.4'te performans ölçütlerine bakıldığında RO, XGBoost ve AdaBag modelleri bir çok metrikte öne çıkmaktadır.

En yüksek doğruluğa sahip modelin RO (0,89), AdaBag (0,87) ve XGBoost (0,87) modellerinden elde edildiği görülmektedir. RO en yüksek doğruluk oranına sahip ve doğruluk güven aralığı oldukça geniştir. AdaBag, XGBoost yine yüksek doğruluk oranlarına ve dar doğruluk güven aralıklarına sahiptir.

Dengeli doğruluk metriğinde ise XGBoost (0,81), DVM (0,83) ve RO (0,83) modelleri yüksek performans göstermiştir.

KNN en düşük doğruluk oranına (0,70) ve geniş bir doğruluk güven aralığına sahiptir. Bu durum, bu modelin daha basit bir yapıya sahip olmasından kaynaklanabilir. Düşük doğruluk oranı ve geniş doğruluk güven aralığı modelin tutarlılığında ve genel performansında zayıf olmasına işaret eder.

Modellerin Kappa değerlerine göre RO (0,71), XGBoost (0,66) ve AdaBag (0,64) modellerinin göreceli en yüksek değere sahip olduğu görülmektedir. RO modeli sınıflandırma açısından en başarılı sonuçları verirken KNN ve CART düşük uyum göstermektedir.

Modellerin EAA ölçütüne bakıldığında ise en yüksek değerin XGBoost (0,94), LRM (0,93), RO (0,92) ve DVM (0,92) modellerinde elde edildiği dikkat çekmektedir.

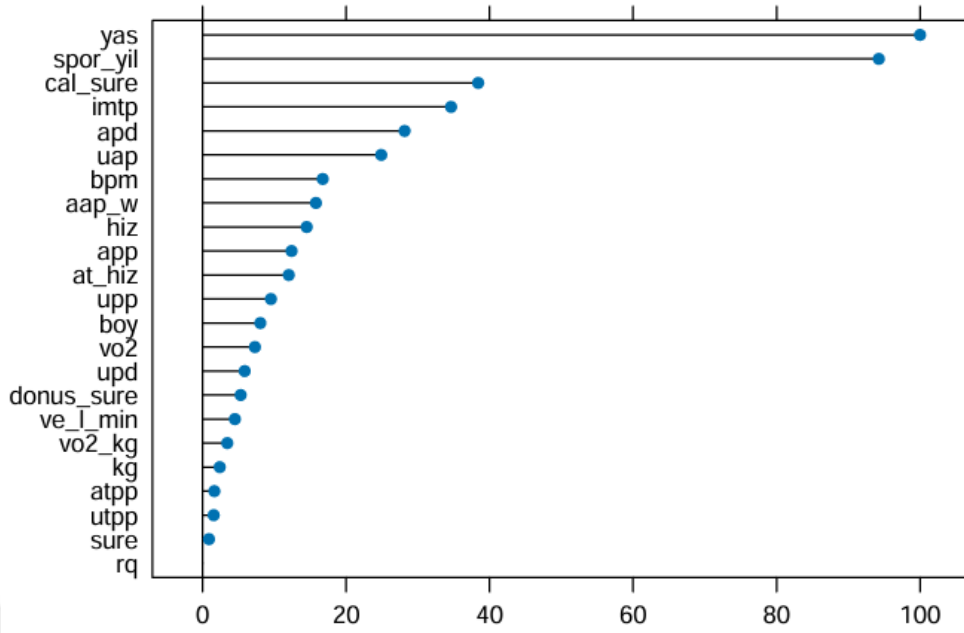
Modellerin duyarlılık oranlarına bakıldığında, DVM (0,84), LRM (0,76), RO (0,69) ve XGBoost (0,69) modellerinin pozitif sınıfı (Elit sınıfı) tespit etmede diğer modellere göre başarılı görünmektedir. KNN ve NB modelleri düşük duyarlılık ile pozitif örnekleri kaçırma eğilimindedirler.

Modellerin seçicilik oranlarına bakıldığında, RO (0,97), NB (0,97) ve AdaBag (0,97) modelleri negatif sınıfı (Aday sınıfı) doğru tahminde başarı gösterdiği görülmektedir. LRM ise en düşük seçicilik oranına sahiptir.

Modelleri F1 ölçütü bakımından karşılaştırdığımız da ise en yüksek değerin RO (0,78), AdaBag (0,72) ve DVM (0,73) modellerinde hesaplandığı görülmektedir.

Model performans ölçütleri birlikte değerlendirildiğinde, sınıflandırma modelleri arasında RO, AdaBag ve XGBoost gibi ağaç tabanlı topluluk modelleri verilerin yapısal ilişkilerini iyi bir şekilde öğrenebilme kapasitesine sahiptir ve yüksek doğruluk oranları ile diğerlerine kıyasla en başarılı modeller olarak öne çıkmıştır. Bu modellerin özellik seçimindeki başarısı ve genelleme yetenekleri yüksektir. RO modelinin yüksek seçicilik ve Kappa değeri modelin yüksek doğrulukla sınıflandırma yaptığını göstermektedir. Fakat aday sınıfı belirleme daha başarılı olduğu şeklinde yorumlanabilir. Bununla birlikte XGBoost ve RO modellerinin yüksek dengeli doğruluk değerleri ve F1 ölçüleri modelin dengeleme kapasitesini arttırmıştır. Bu modellerin yüksek doğruluk oranları ve uyum (Kappa) değerleri, veriler üzerinde önemli bilgi çıkarımı yapabildiklerini göstermektedir. Tüm performans değerlendirme ölçütleri dikkate alındığında, aritmetik ortalama ile eksik gözlem atanmış veri setinde RO ve XGBoost modellerinin genel olarak daha iyi performans göstermekte olduğu sonucuna varılmıştır.

Elit olma durumlarını tahmin etmek için on kat çapraz geçişleme yöntemiyle kurulan RO modeline ait final sınıflandırma ağacının değişken önem dereceleri dikkate alınarak çıkarılan en önemli faktörler Şekil 4.1'te gösterilmiştir.



Şekil 4.1. Rastgele Orman modeline göre değişkenlerin önem ağırlıkları (eksik gözlemlere aritmetik ortalama ile atama yapılan veri setinde).

Şekil 4.1’de görülen, performans testlerine katılan sporcuların elitlik kategorisi tahminine etki eden önem dereceleri yorumlandığında; sporcunun yaşı, sporcunun ölçüm tarihi itibarıyla aktif olarak spor yaptığı yıl sayısı, çok yönlü reaktif çeviklik testi bitirme süresi, izometrik orta uyluk çekiş testi ve alt ekstremite wingate anaerobik güç testi yorgunluk yüzdesi sırası ile elitlik kategorisi tahminine etki eden en önemli kriterlerdir.

Sporcunun yaşı ve aktif olarak spor yaptığı yıl sayısının modelde ön plana çıkmıştır. Bu durumun; sporcunun vücudunun branşına adaptasyon yetisinin, dayanıklılığın, deneyimin, müsabaka tecrübelerinin, performans testlerine katılma oranının yaş ve antrenman yaşı ile doğru orantılı olmasıyla alakalı olduğu şeklinde yorumlanması mümkündür.

Model tahmin performanslarının iyileştirilmesi, elit sınıfın tahmin performanslarının artırılması için eksik veri atama yöntemleri olarak RO ve KNN yöntemleri uygulanmış, karşılaştırmanın daha geçerli hale gelmesi için tüm modeller için *caret* paketi kullanılmış ve *tuneGrid* ve *tuneLength* fonksiyonu ile hiperparametre

kombinasyonları denenerek model optimizasyonları gerçekleştirilmiştir. 10 kat 3 tekrarlı çapraz geçirme yöntemi kullanılmıştır.

Yapılan iyileştirme analizleri sonrası sonuçlar Tablo 4.5 Tablo 4.6 ve Tablo 4.7'de verilmiştir.

Tablo 4.5. Elit olma sınıflama modellerine ait test performans ölçütleri (eksik gözlemlere RO algoritması ile atama yapılan veri setinde).

Model	C4.5	KNN	CART	CART (Bagging)	RO	LRM	NB	AdaBag	DVM	XGBoost
Doğruluk	0,8723	0,5957	0,7660	0,8723	0,8936	0,8085	0,7234	0,8085	0,8936	0,9574
% 95 GA (Alt-Üst)	(0,7428, 0,951)	(0,442, 0,736)	(0,619, 0,877)	(0,742, 0,951)	(0,769, 0,964)	(0,667, 0,908)	(0,573, 0,8438)	(0,667, 0,9085)	(0,769, 0,964)	(0,854, 0,994)
Kappa	0,7441	0,1132	0,5109	0,7369	0,7798	0,6072	0,4327	0,6036	0,7798	0,9131
Duyarlılık	0,9048	0,1429	0,5714	0,7619	0,7619	0,7143	0,6190	0,6667	0,7619	0,9048
Seçicilik	0,8462	0,9615	0,9231	0,9615	1,0000	0,8846	0,8077	0,9231	1,0000	1,0000
PKD	0,8261	0,7500	0,8571	0,9412	1,0000	0,8333	0,7222	0,8750	1,0000	1,0000
NKD	0,9167	0,5814	0,7273	0,8333	0,8387	0,7931	0,7241	0,7742	0,8387	0,9286
Dengeli Doğruluk	0,8755	0,5522	0,7473	0,8617	0,8810	0,7995	0,7134	0,7949	0,8810	0,9524
F1 Ölçüsü	0,8636	0,2400	0,6857	0,8421	0,8649	0,7692	0,6667	0,7568	0,8649	0,9500
EAA	0,8919	0,6694	0,7473	0,9396	0,9725	0,8974	0,7857	0,9048	0,9231	0,9707

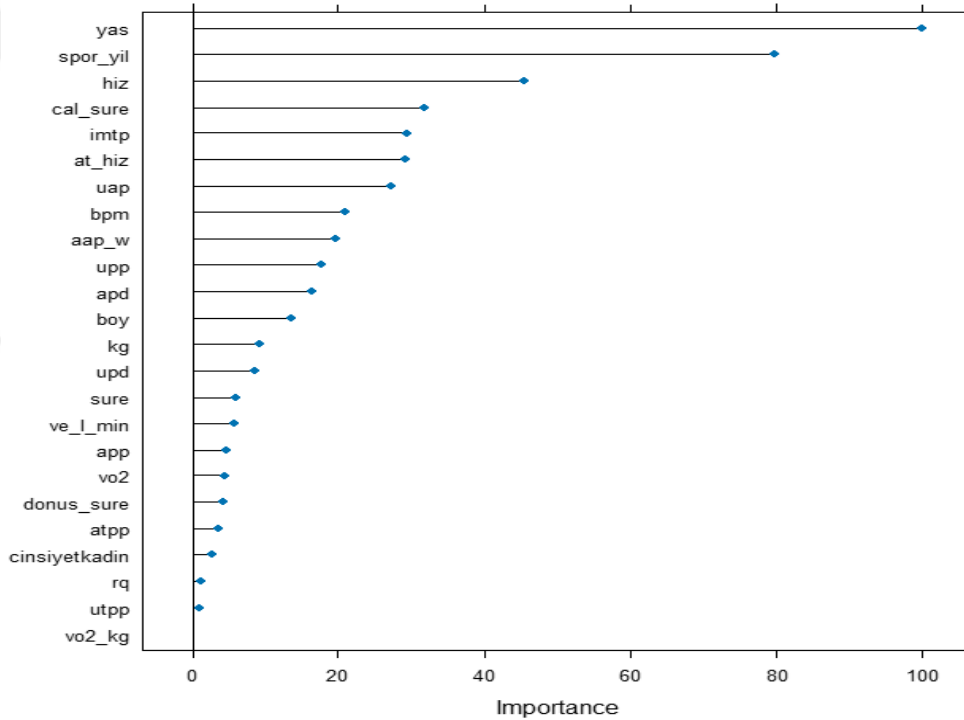
Tablo 4.5'te performans ölçütlerine bakıldığında RO ile eksik veri atandıktan sonra en yüksek doğruluğa sahip modelin XGBoost (0,95), RO (0,89) ve DVM (0,89) olduğu görülmektedir. RO modelinin ortalama ile atanan veri setindeki aynı doğruluk performansını gösterdiği XGBoost ve DVM modellerinin doğruluklarının arttığı gözlemlenmiştir.

Kappa değerleri belirgin bir şekilde artmıştır. XGBoost (0,91) en yüksek Kappa değerine sahip model olmuştur. RO, DVM (0,77) ve ardından C4.5 (0,74) modellerinin Kappa değerleri iyi uyum sınırları içerisindedir.

KNN'nin duyarlılığının hâlâ çok düşük (0,14) olması dengesiz sınıflarda başarısız olduğunu göstermektedir.

Tüm performans ölçütleri birlikte değerlendirildiğinde; doğruluk metriğinde XGBoost, duyarlılık metriğinde C4.5 ve XGBoost, seçicilik metriğinde RO, DVM ve XGBoost; dengeli doğruluk metriğinde XGBoost, EAA metriğinde ise yine XGBoost ve RO modelleri en yüksek performanslara sahip modeller olarak öne çıkmaktadır. XGBoost ve RO modellerinin genel stabiliteyi yüksek olmalarından kaynaklı bu senaryo için önerilebilir.

XGBoost modelinin performans testlerine katılan sporcuların Elit olma durumlarını tahmin etmek için oluşturduğu final sınıflandırma değişken önem dereceleri dikkate alınarak çıkarılan en önemli faktörler Şekil 4.2’de gösterilmiştir.



Şekil 4.2. XGBoost modeline göre değişkenlerin önem ağırlıkları (eksik gözlemlere RO algoritması ile atama yapılan veri setinde).

Şekil 4.2’de görülen, performans testlerine katılan sporcuların elitlik kategorisi tahminine etki eden önem dereceleri yorumlandığında; sporcunun yaşı, sporcunun ölçüm tarihi itibarıyla aktif olarak spor yaptığı yıl sayısı, aerobik güç testinde ulaştığı hızı, reaktif çeviklik testini bitirme süresi ve tüm vücut izometrik kuvveti sırası ile elitlik kategorisi tahminine etki eden en önemli kriter olarak belirlenmiştir.

Tablo 4.6. Elit olma sınıflama modellerine ait test performans ölçütleri (eksik gözlemlere KNN algoritması ile atama yapılan veri setinde).

Model	C4.5	KNN	CART	CART (Bagging)	RO	LRM	NB	AdaBag	DVM	XGBoost
Doğruluk	0,8723	0,6383	0,7660	0,8723	0,8723	0,8511	0,7234	0,8936	0,9149	0,9787
% 95 GA (Alt-Üst)	(0,742, 0,951)	(0,485, 0,773)	(0,6197, 0,877)	(0,742, 0,951)	(0,742, 0,951)	(0,716, 0,938)	(0,573, 0,843)	(0,769, 0,964)	(0,796, 0,976)	(0,887, 0,999)
Kappa	0,7418	0,2066	0,5109	0,7369	0,7345	0,6945	0,4327	0,7798	0,8246	0,9568
Duyarlılık	0,8571	0,1905	0,5714	0,7619	0,7143	0,7619	0,6190	0,7619	0,8095	0,9524
Seçicilik	0,8846	1,0000	0,9231	0,9615	1,0000	0,9231	0,8077	1,0000	1,0000	1,0000
PKD	0,8571	1,0000	0,8571	0,9412	1,0000	0,8889	0,7222	1,0000	1,0000	1,0000
NKD	0,8846	0,6047	0,7273	0,8333	0,8125	0,8276	0,7241	0,8387	0,8667	0,9630
Dengeli Doğruluk	0,8709	0,5952	0,7473	0,8617	0,8571	0,8425	0,7134	0,8810	0,9048	0,9762
F1 Ölçüsü	0,8571	0,3200	0,6857	0,8421	0,8333	0,8205	0,6667	0,8649	0,8947	0,9756
EAA	0,8452	0,6813	0,7473	0,9414	0,9661	0,9139	0,7930	0,9222	0,9322	0,9835

Tablo 4.6’da performans ölçütlerine bakıldığında KNN ile eksik gözlem atanan veri setinde en yüksek doğruluğa sahip modellerin XGBoost (0,98), DVM (0,91) ve AdaBag (0,89) olduğu görülmektedir. XGBoost hem en yüksek doğruluğa hem de dar güven aralığına sahiptir.

XGBoost modeli (0,95) en yüksek duyarlılık metriğine sahip olan modeldir. RO, AdaBag, DVM ve XGBoost (1,00) mükemmel seçicilikle performansı en üst sıradaki modellerdir.

Dengeli doğruluk metriğinde lider konumda olan model XGBoost (0,97) ardından RO (0,96) ve DVM (0,93) olduğu görülmektedir.

Kappa ve F1 ölçüleri de XGBoost’ta çok güçlü (0,96 ve 0,98) olduğu görülmektedir.

KNN ile eksik gözlem atanan veri setinde uygulanan on farklı model arasından XGBoost tüm metriklerde en iyi performansı göstermekte olduğu ve alternatif olarak da DVM, AdaBag ve RO modellerinin kullanılabileceği şeklinde yorum yapılabilir.

Sınıf dengesizliğinin model performansı üzerindeki etkisinin gözlemlenmesi açısından RO ile eksik gözlem atanan veri setinde SMOTE dengeleme analizi yapılmıştır. Dengeleme analizi sadece eğitim veri setine uygulanmış, test veri setine yapay örneklerin sızması engellenmiştir. Veri seti olarak RO ile eksik gözlem atanmış veri setinin seçilmesinin sebebi uygulanan analizler arasından en iyi model test performans ölçütlerini vermesidir.

Tablo 4.7. Eksik gözlemlere RO algoritması ile atama yapılan veri setinde SMOTE ile dengeleme yapıldıktan sonra elde edilen sınıflama modellerine ait test performans ölçütleri.

Model	C4.5	K-NN	CART	CART (Bagging)	RO	LRM (regularized)	NB	AdaBag	DVM (radial)	XGBoost
Doğruluk	0,7600	0,6739	0,7391	0,9130	0,9348	0,8696	0,8043	0,8913	0,9130	0,8913
% 95 GA (Alt-Üst)	(0,6123, 0,8741)	(0,5198, 0,8047)	(0,5887, 0,8573)	(0,7921, 0,9758)	(0,8210, 0,9863)	(0,7374, 0,9506)	(0,6609, 0,9064)	(0,7643, 0,9638)	(0,7921, 0,9758)	(0,7643, 0,9638)
Kappa	0,4315	0,3114	0,4250	0,8083	0,8583	0,7283	0,5981	0,7639	0,8138	0,7639
Duyarlılık	0,5000	0,6250	0,6250	0,8750	0,9375	0,9375	0,8750	0,8750	0,9375	0,8750
Seçicilik	0,9000	0,7000	0,8000	0,9333	0,9333	0,8333	0,7667	0,9000	0,9000	0,9000
PKD	0,7273	0,5263	0,6250	0,8750	0,8824	0,7500	0,6667	0,8235	0,8333	0,8235
NKD	0,7714	0,7778	0,8000	0,9333	0,9655	0,9615	0,9200	0,9310	0,9643	0,9310
Dengeli Doğruluk	0,7000	0,6625	0,7125	0,9042	0,9354	0,8854	0,8208	0,8875	0,9188	0,8875
F1 Ölçüsü	0,5926	0,5714	0,6250	0,8750	0,9091	0,8333	0,7568	0,8485	0,8824	0,8485
EAA	0,7125	0,7083	0,7354	0,9667	0,9833	0,9417	0,9229	0,9750	0,9625	0,9670

Tablo 4.7’de RO ile eksik veri ataması ve SMOTE ile sınıf dengelemesi yapılan veri setindeki model performanslarına bakıldığında; tüm modellerin neredeyse hepsinde seçicilik metriği hariç tüm metriklerde artış gözlemlenmiştir. Buna rağmen yüksek doğruluk, hem elit hem de aday grubu belirlemede ve tahmin uyumunda iyi sonuçlar veren modellere ulaşılmıştır.

En iyi performans gösteren modeller;

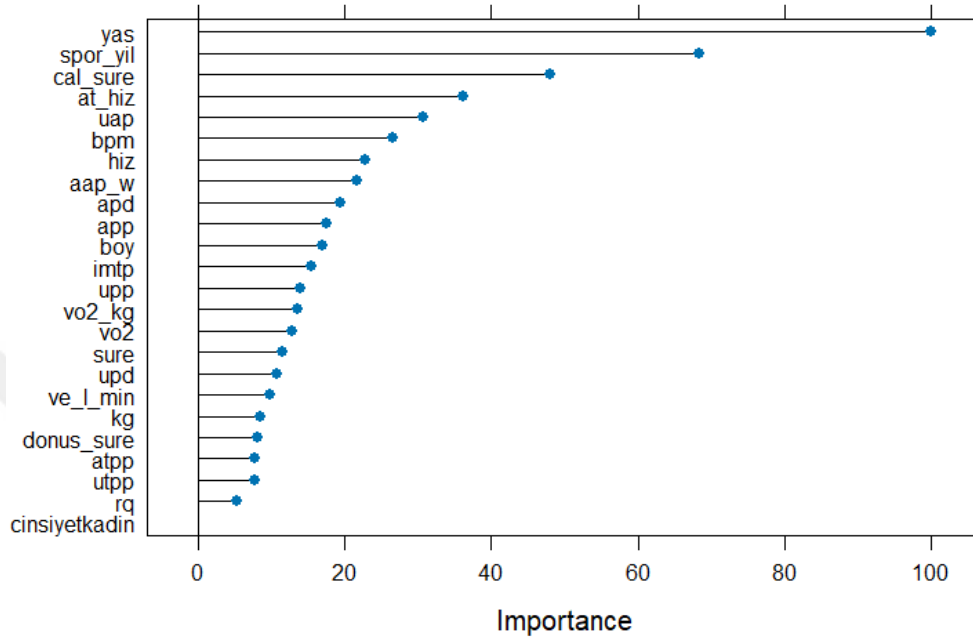
RO (doğruluk:0,93; Kappa: 0,85; F1 ölçüsü:0,90; EAA:0,98)

DVM (doğruluk:0,91; Kappa: 0,81; F1 ölçüsü:0,88; EAA:0,96)

CART Bagging (doğruluk:0,91; Kappa: 0,80; F1 ölçüsü:0,87; EAA:0,96)

şeklinde sıralanmaktadır.

RO modelinin performans testlerine katılan sporcuların Elit olma durumlarını tahmin etmek için oluşturduğu final sınıflandırma değişken önem dereceleri dikkate alınarak çıkarılan en önemli faktörler Şekil 4.3'te verilmiştir.

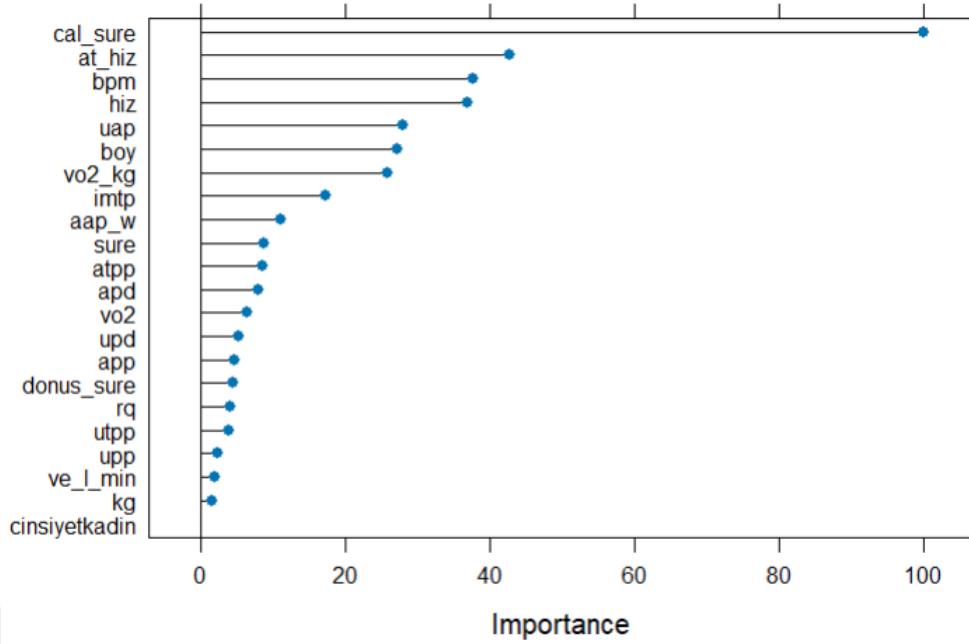


Şekil 4.3. RO modeline göre değişkenlerin önem dereceleri (RO+SMOTE uygulanan veri setinde).

Şekil 4.3'te görülen, performans testlerine katılan sporcuların kategorisinin tahminine etki eden faktörlerin önem dereceleri yorumlandığında; sporcunun yaşı, sporcunun ölçüm tarihi itibarıyla aktif olarak spor yaptığı yıl sayısı, reaktif çeviklik testini bitirme süresi, anaerobik eşikteki hızı ve üst ekstremitte anaerobik güç testinde ulaştığı ortalama gücü sırası ile elitlik kategorisi tahminine etki eden en önemli kriterlerdir.

Sporcunun yaşı ve aktif olarak spor yaptığı yıl sayısı her senaryo ve modelde en yüksek ayırt ediciliğe sahip değişkenler olduğu görülmektedir.

Karşılaştırma açısından aynı veri setinden (RO+SMOTE) yaş ve aktif spor yaptığı yıl sayısı çıkartılarak elde edilen faktörlerin önem dereceleri Şekil 4.4'te verilmiştir.



Şekil 4.4. RO modeline göre yaş ve aktif spor yaptığı yıl sayısı değişkenleri çıkartılarak elde edilen faktörlerin önem dereceleri (RO+SMOTE uygulanan veri setinde).

Şekil 4.4'te görülen, performans testlerine katılan sporcuların kategorisinin tahminine etki eden faktörlerin önem dereceleri yorumlandığında; sporcunun yaşı, sporcunun ölçüm tarihi itibariyle aktif olarak spor yaptığı yıl sayısı çıkarıldığında yaş ve spor yılından sonra gelen değişkenlerin önem derecelerinin aynı sırayla olmasa da ilk 5'te yer aldığı görülmektedir. Bu değişkenler, çok yönlü reaktif çeviklik testi bitirme süresi, izometrik orta uyluk çekiş testi ve alt ekstremitte wingate anaerobik güç testi yorgunluk yüzdesi, üst ekstremitte anaerobik güç testi ortalama güç, aerobik güç ve kapasite testinde anaerobik eşikteki hızı ve maksimum hız olarak belirlenmiştir.

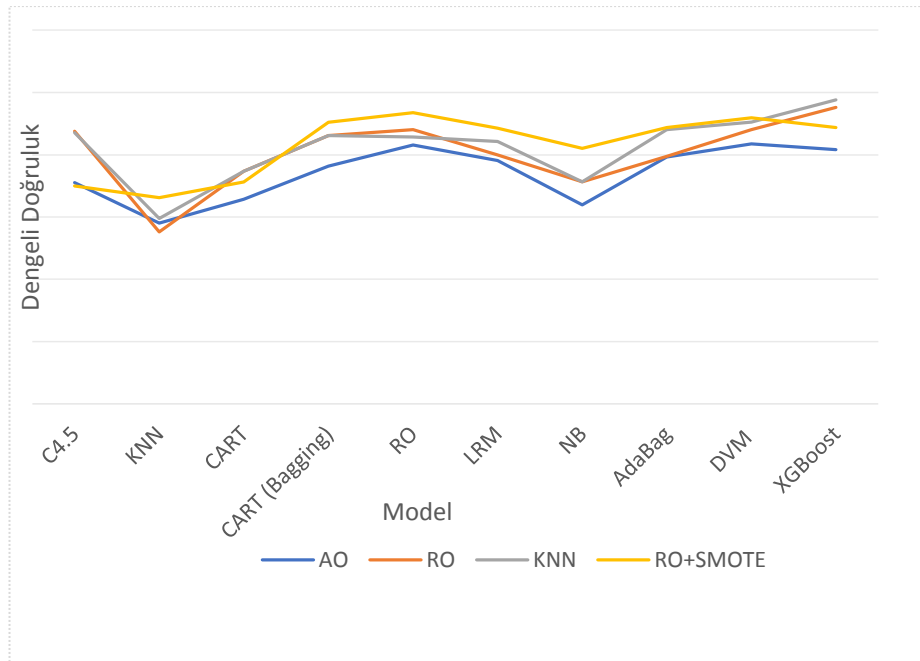
Tablo 4.8 ve Tablo 4.9'da sırasıyla modellerin genel performansa göre ve uygulanan senaryoya bağlı karşılaştırmaları verilmiştir.

Tablo 4.8. Modellerin genel performans karşılaştırması.

Model	Yorum
XGBoost/RO	Tüm senaryolarda en yüksek doğruluk ve tutarlılıkla lider görünmektedir.
AdaBag / CART(Bagging)	Ensemble yapılar olduğundan tahmin performansları yüksektir, özellikle RO eksik veri atamasından sonra iyi sonuçlar vermiştir.
Naive Bayes / KNN	Duyarlılık ve F1 skorları diğer modellere göre düşüktür; sınıf dengesizliği ve veri yapısına karşı duyarlıdır.
DVM	Tutarlı ve özellikle RO ataması sonrası yüksek performanslı olduğu gözlenmiştir.

Tablo 4.9. Modellerin uygulanan senaryoya bağlı performans karşılaştırması.

Senaryo	Lider modeller ve en güçlü metrikler
AO ile eksik gözlem atama	RO (doğruluk:0,89; ceçicilik:0,97; F1 ölçüsü:0,78; EAA: 0,92)
RO ile eksik gözlem atama	XGBoost (Doğruluk: 0,95; Kappa: 0,91; dengeli doğruluk ve F1 ölçüsü: 0,95; EAA:0,97)
KNN ile eksik gözlem atama	XGBoost (Doğruluk: 0,97; Kappa: 0,95; dengeli doğruluk:0,93 ve F1 ölçüsü: 0,90; EAA:0,98)
RO ile eksik gözlem atama+SMOTE (RO+SMOTE)	RO (Doğruluk: 0,93; Kappa: 0,85; dengeli doğruluk ve F1 ölçüsü: 0,95; EAA:0,98)

**Şekil 4.5.** Uygulanan senaryolarda dengeli doğruluk ölçütüne göre modellerin karşılaştırılması.

Tablo 4.10. Uygulanan senaryolarda dengeli doğruluk ölçütüne göre modellerin karşılaştırılması.

Senaryo/Model	C4.5	KNN	CART	CART (Bagging)	RO	LRM	NB	AdaBag	DVM	XGBoost
AO	0,71	0,58	0,66	0,76	0,83	0,78	0,64	0,79	0,83	0,82
RO	0,88	0,55	0,75	0,86	0,88	0,80	0,71	0,79	0,88	0,95
KNN	0,87	0,60	0,75	0,86	0,86	0,84	0,71	0,88	0,90	0,98
RO+SMOTE	0,70	0,66	0,71	0,90	0,94	0,89	0,82	0,89	0,92	0,89

Şekil 4.5 ve Tablo 4.10'da uygulanan senaryolarda model test performans ölçütlerinden dengeli doğruluk değerleri verilmiştir. Veri seti evrende bulunduğu haliyle dengesiz bir veri seti özelliği taşıdığından, genel doğruluk yerine dengeli doğruluk ve F1 ölçüleri model ve senaryo kıyaslamasında dikkate alınmıştır.

Aritmetik ortalama ile eksik gözlem atanan ilk senaryoda basit atama yöntemi istikrarlı modeller olan DVM ve RO modellerini öne çıkarmıştır. En zayıf performansı KNN modeli göstermiş ve her sınıf için doğru tahmin yeteneği XGBoost, LRM ve AdaBag modellerinin gerisinde kalmıştır.

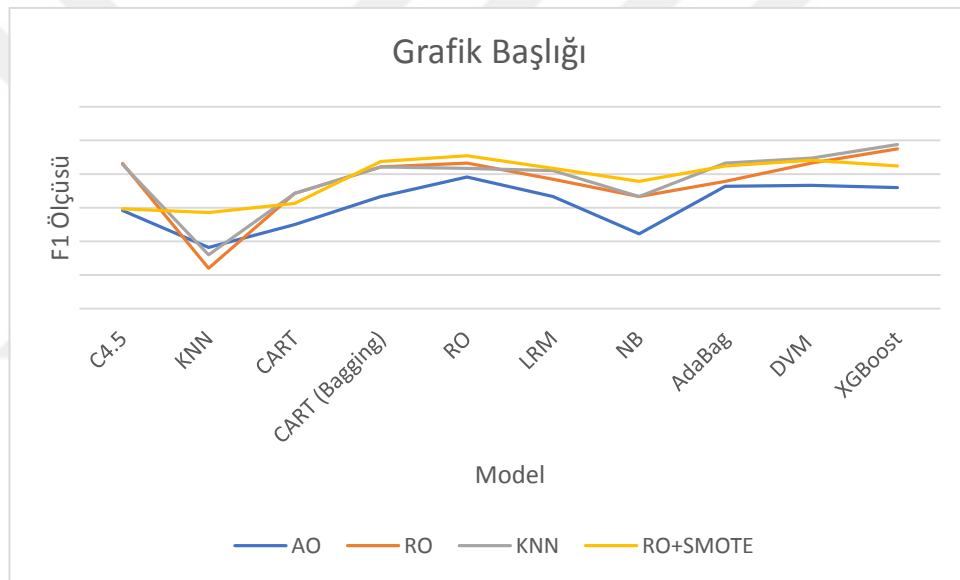
RO algoritması ile eksik gözlem atanan senaryoda en iyi performansı XGBoost modeli göstermiş ve ardından RO ve DVM modelleri performans olarak öne çıkmaktadır. RO atması XGBoost modeli ile mükemmel uyum sağlamış şekilde yorumlanabilir. Aynı şekilde RO atmasının RO modeli lehine bir durum yaratmış olması da söz konusudur. KNN modeli tüm senaryolar içinde en düşük dengeli doğruluk değerine bu senaryoda sahip olması RO ile oluşturulan veri yapısı, KNN 'in komşuluk mantığı ile uyum sağlamış şekilde yorumlanabilir. C4.5 modeli bir önceki senaryoya göre dengeli doğruluk performansını %23,5 oranında artırmıştır.

KNN algoritması ile eksik gözlem atanan senaryoda en iyi performansı yine XGBoost algoritması göstermektedir. RO modeli kendi atama yöntemine göre düşüş yaşamıştır.

RO algoritması ile eksik gözlem atanan veri setinde SMOTE veri dengeleme analizi yapılmış senaryoda RO modeli en yüksek dengeli doğruluk ölçütüne sahip

modeldir. Veri dengesizliğinin giderilmesi RO ve DVM modellerinin sınıf ayırım gücünü artırmıştır. NB modeli basit atama senaryosuna göre bu ölçütte %28,4 artış göstermiştir. XGBoost modelinin ise KNN senaryosundaki ayırt edicilik gücü düşüş yaşamıştır.

Özet olarak XGBoost modeli tüm senaryolarda en tutarlı model olarak karşımıza çıkmakta, DVM ise tüm senaryolarda en iyi üç model arasında yer almaktadır. RO özellikle SMOTE ile birlikte güçlü bir tahmin modeli olduğu yorumu yapılabilir. CART ve AdaBag modelleri yine öne çıkan modeller arasında yer almaktadır.



Şekil 4.6. Uygulanan senaryolarda F1 ölçütüne göre modellerin karşılaştırılması.

Tablo 4.11. Uygulanan senaryolara göre modellerin dengeli F1 ölçülerinin karşılaştırılması.

Senaryo/Model	C4.5	KNN	CART	CART (Bagging)	RO	LRM	NB	AdaBag	DVM	XGBoost
AO	0,58	0,36	0,50	0,67	0,78	0,67	0,44	0,73	0,73	0,72
RO	0,86	0,24	0,69	0,84	0,86	0,77	0,67	0,76	0,86	0,95
KNN	0,86	0,32	0,69	0,84	0,83	0,82	0,67	0,86	0,89	0,98
RO+SMOTE	0,59	0,57	0,63	0,88	0,91	0,83	0,76	0,85	0,88	0,85

Şekil 4.6 ve Tablo 4.11’de uygulanan senaryolarda model test performans ölçütlerinden F1 ölçüsü değerleri verilmiştir. Veri seti evrende bulunduğu haliyle dengesiz bir veri seti özelliği taşıdığından, genel doğruluk yerine dengeli doğruluk ve F1 ölçüleri model ve senaryo kıyaslamasında dikkate alınmıştır.

Aritmetik ortalama ile eksik gözlem atanan ilk senaryoda F1 ölçüsü değerlendirmesinde en iyi performans gösteren model RO modeli ve ardından DVM ve AdaBag modelleridir. Basit atama yöntemleri ensemble modellerin gücünü ortaya koymuştur. En zayıf performans KNN modeline aittir.

RO algoritması ile eksik gözlem atanan senaryoda F1 ölçüsü değerlendirildiğinde en yüksek performansa sahip olan modelin XGBoost olduğu görülmektedir. Ardından C4.5 ve DVM modelleri gelmektedir. RO algoritması ile veri ataması ağaç tabanlı modeller ile yüksek uyum göstermiş olabilir. KNN modeli bir önceki senaryoya göre performansında düşüş göstermiş, C4.5 ise %48 oranında performans artışı göstermiştir.

KNN algoritması ile eksik gözlem atanan senaryoda en iyi F1 ölçüsü XGBoost modeline aittir. RO modeli kendi atama senaryosuna göre düşüş göstermiştir.

RO algoritması ile eksik gözlem atanan veri setinde SMOTE veri dengeleme analizi yapılmış senaryoda en iyi F1 ölçüsü RO modeline aittir. CART Bagging yüksek F1 ölçüsü ile bu senaryoda ön plana çıkmaktadır. Veri dengesizliğinin giderilmesi bagging modellerin gücünü artırmıştır. NB modeli basit atama yöntemine göre F1 ölçüsünde %70 lik bir artış göstermiştir. XGBoost, KNN algoritması ile eksik gözlem atanan bir önceki senaryoya göre F1 ölçüsünde düşüş yaşamıştır.

Özet olarak; tüm senaryolar içerisinde tutarlı performansı gösteren model XGBoost, özellikle SMOTE ile yüksek tahmin performansı gösteren RO, tüm senaryolarda stabil olarak DVM modelleri iyi tahmin modelleri olarak belirlenmiştir.

5. TARTIŞMA

Bu tez çalışmasında performans testlerine katılan toplam 188 Güreş ve Judo sporcusuna ait anonimleştirilmiş gerçek veriler kullanılmıştır. Karar Ağaçları (CART, CART Bagging, C4.5), XGBoost, AdaBag, Rastgele Orman, KNN, Lojistik Regresyon Modeli, NaiveBayes ve Destek Vektör Makinaları olmak üzere 10 adet elit sporcu sınıflamasına yönelik model kurulmuştur. Elit sporcu olmaya etki eden faktörler belirlenmiştir. Tüm senaryolar içerisinde tutarlı performansı gösteren model XGBoost, özellikle SMOTE ile yüksek tahmin performansı gösteren RO, tüm senaryolarda stabil olarak DVM modelleri iyi tahmin modelleri olarak belirlenmiştir.

Analizlerde temel olarak *caret* paketi tercih edilmiş, sınıflama doğruluğunu arttırmak amacıyla on katlı 3 tekrarlı çapraz geçişleme yöntemi kullanılmıştır. Model seçimleri literatürde sınırlı sayıda bulunan sporcu seçimi ile alakalı çalışmalarda kullanılan modellerden ve bunu yanı sıra sınıflama problemlerinde sık kullanılan modeller arasından yapılmıştır.

Elit sporcu sınıflama model performansları; doğruluk, Kappa, duyarlılık, seçicilik, kesinlik, dengeli doğruluk, F1 ölçüsü ve EAA göstergeleri ile karşılaştırılmıştır.

RO modeli karar ağaçlarının birleşimi ile çalışan, yüksek doğruluk sağlayan ve aynı zamanda aşırı uyum riskini düşük tutan bir model olduğu için tercih edilmiş ve tüm sınıflama model performans ölçülerinde yüksek değerlere sahip tahmin modeli olmuştur.

XGBoost modeli, özellikle doğrusal olmayan ilişkiler içinde olan verilerin oluşturduğu veri setlerinde yüksek doğruluk sağladığından ve özellikle dengesiz sınıflar ile çalışırken yüksek performans gösterdiğinden tercih edilmiştir. Fakat fazla parametre ayarı gerektirdiğinden eğitimi süreci zordur ve uzun sürmektedir.

DVM modeli özellikle yüksek boyutlu veri setlerinde iyi performans gösterdiğinden ve doğrusal olmayan ilişkiler içeren veri setlerinde modelleme kapasitesine sahip olduğu için tercih edilmiştir.

KNN ve NB modelleri basit ve hızlı modeller olduğundan tercih edilmiştir. KNN ve NB modellerinin doğruluk, Kappa, duyarlılık ve F ölçülerine bakıldığında diğer modellere kıyasla düşük performans sergiledikleri görülmüştür. NB modeli, bağımsız özellikleri olan veri setlerinde iyi performans gösterir. Aksi halde model performansı düşebilir.

CART Bagging, AdaBag ve XGBoost gibi topluluk (ensemble) yöntemleri tekli modeller ile kıyaslandığında genel performansta belirgin bir artış sağladığı görülmüştür.

LRM modeli yüksek doğruluk, duyarlılık ve EAA değeri ile yüksek bir tahmin performansı göstermiştir. Yüksek bir negatif tahmin değeri vermesine rağmen pozitif tahmin değerinin diğer modellere göre düşük olması nedeniyle elit sınıfı kaçırma olasılığı yüksek olarak değerlendirilmiştir.

Yaklaşık tüm performans değerlendirme ölçütlerinde RO ve XGBoost modeli diğer modellere göre daha yüksek değerlere sahip olduğundan elit sınıflamasını en yüksek performans ile yapan modeller olmuştur. Ayrıca yüksek seçicilik oranları ile aday grubu da başarılı bir şekilde tahmin etmişlerdir. Sporcuların yetiştirilmesindeki maliyet ve müsabaka başarısı göz önüne alındığında elit sporcu seçim modellerinde aday grubun doğru tahmini de yüksek avantaj sağlayacaktır.

Tahmin modellerinde ortak olarak elit sporcu olma durumunu etkileyen faktörlerin başında; sporcunun yaşı, sporcunun ölçüm tarihi itibarıyla aktif olarak spor yaptığı yıl sayısı, çok yönlü reaktif çeviklik testi bitirme süresi, anaerobik eşikteki hızı ve aerobik güç testinde ulaştığı hızı, izometrik orta uyluk çekiş testi ve wingate anaerobik güç testi yorgunluk yüzdesi ve ortalama güç değerleri olduğu görülmüştür. Sporcunun yaşı ve aktif olarak spor yaptığı yıl sayısı tüm modellerde belirleyici faktörler olarak ön plana çıkmıştır. Bu durumun; sporcunun vücudunun branşına

adaptasyon yetisinin, dayanıklılığın, deneyimin, müsabaka tecrübelerinin, performans testlerine katılma oranının yaş ve antrenman yaşı ile doğru orantılı olmasıyla alakalı olduğu şeklinde yorumlanması mümkün olmakla beraber bu değişkenlerin elit grupta daha yüksek olması olasıdır. Bu değişkenlerin sınıflamada kullanılması diğer değişkenlerin önem derecelerini nasıl etkilediğinin incelenmiştir. Yaş ve spor yılı değişkenleri analiz dışında bırakılarak yapılan sınıflama analizlerinde faktörlerin önem dereceleri hesaplanmış ve ilk beş kriterin çok kapsamlı bir değişiklik göstermediği sonucuna varılmıştır. Böylelikle, sporcuların gerçek ölçüm verilerini kullanarak klasik sporcu seçimine alternatif olabilecek sınıflama modellerinin uygulanması, kıyaslanması ve eksik gözlem atama stratejileri değerlendirilmiş ve spor bilimlerinde uygulanabilirliği gösterilmiştir.

Peki ülkemizde spor branşlarına ait elit sporcu seçme kriterleri nelerdir? Her spor branşının gereklilikleri ve dinamikleri farklı olmasından kaynaklı federasyonlara göre elit sporcu seçme kriterleri değişiklik göstermektedir ve bu kriterler ilgili federasyonların yönetmeliklerinde açık bir şekilde belirlenmiştir. Bunlar genelde sporcu performansı, ulusal ve uluslararası müsabakalardaki başarılar, teknik yeterlilikler, disiplin ve etik değerlere uygunluk kriterlerini içeren çerçevede yapılandırılmıştır. Federasyonlar, sporcularını genellikle ulusal turnuvalardaki performanslarına göre değerlendirerek uluslararası müsabakalarda yer alacak sporcuları seçmek üzere sporcu seçme kampları düzenler. Kamplarda; fiziksel performans, teknik ve taktik değerlendirilmeler yapılır. Seçim kriterleri branşlara göre farklılık göstermektedir. Birçok federasyon, ulusal şampiyonalarda veya uluslararası turnuvalarda başarı elde eden sporculara öncelik verir. Gençlik kategorilerinde elit aday kadrolar genellikle sporcu potansiyeline göre büyükler kategorisinde ise performans ve başarı kriterlerine göre seçilmektedir. İlgili yönetmelikte temel kriter belirtilmekte ve uluslararası organizasyonlarda ülkeyi temsil eden sporcular elit sporcu olarak kabul edilmektedir. Özetle elit sporcu seçimi;

- Performans (ulusal ve uluslararası başarılar)
- Teknik yeterlilik
- Federasyon kararları ve yönetmelikte bulunan kriterler

- Disiplin, etik değerler ve yönetmelikte belirtilen engel haller (doping ihlali, şike, kanuna aykırı bahis oyunları oynatma vs.)

Seçim kriterleri tez kapsamında bulunan judo ve güreş branşlarına özelleştirildiğinde;

Güreş branşında elit sporcu olma kriterleri Türkiye Güreş Federasyonu (TGF) tarafından düzenlenir. Elit sporcu seçimlerinde performans ve müsabaka dereceleri olarak; ulusal (Türkiye Şampiyonası) ve uluslararası (Avrupa, Dünya şampiyonaları, Akdeniz Oyunları vb.) müsabakalarda derece yapan sporcular öncelikli olarak kadroya alınır. Türkiye Güreş Ligi'nde başarılı olan sporcular federasyon tarafından takip edilir. Teknik antrenörler ve federasyon yetkilileri tarafından sporcuların teknik becerileri, taktik anlayışları ve fiziksel dayanıklılıkları değerlendirilir. Aday sporcular, sporcu seçim kamplarında performans testlerine tabi tutulur. Her sporcu kendi yaş grubunda ve bu yaş grubunda belirlenen müsabakalarda değerlendirilir (U15, U17 Yıldızlar Şampiyonaları; U20 Avrupa ve Dünya Gençler Şampiyonaları; Büyükler ulusal ve uluslararası başarıları). Aday sporcular disiplin cezası almamış olmalı ve federasyon etik kurallarına uygun davranmalıdır.

Judo branşında elit sporcu olma kriterleri Türkiye Judo Federasyonu (TJF) tarafından düzenlenir. Elit sporcu seçimlerinde performans ve müsabaka dereceleri olarak; ulusal (Türkiye Şampiyonası) ve uluslararası (Avrupa Kupası, Avrupa ve Dünya Şampiyonaları) müsabakalarda derece yapan sporcular öncelikli olarak kadroya alınır. Judo branşında sporcuların uluslararası sıralaması (IJF dünya sıralaması) da değerlendirilir. Büyükler kategorisinde bu sıralama önem arz etmektedir. Sporcu seçme kamplarında sporcuların nage-waza (atış teknikleri), ne-waza (yer tek teknikleri) gibi becerileri değerlendirilir. Ayrıca kondisyon, dayanıklılık ve mental hazırlık durumları da sporcu seçiminde önemli kriterlerdir. Ümitler ve (U18) ve Gençler (U21) uluslararası turnuvalarda gösterdiği performansa göre Büyükler kategorisinde ise performans ve uluslararası dereceler ile birlikte IJF puan sıralaması önemli arz etmektedir. Judo branşı için takım seçmelerine katılım zorunludur. Seçmelere katılmayan sporcular aday kadroya dahil edilmez.

Her iki branşta da antrenör önerisi, federasyon yönetmeliği ve uluslararası temsil elit sporcu olma adımı büyük öneme sahiptir. Güreş branşında fiziksel güç, dayanıklılık ve müsabaka başarısı öne çıkarken, Judo branşında teknik yeterlilik, puanlama sistemi ve dünya sıralaması daha belirleyicidir. Sıralama oluşturulurken yarışma tarihinden önceki iki yıl dikkate alınır ve yarışma tarihinden önceki 13-24 ay ile yarışma tarihinden önceki 12 aylık süre olarak iki döneme ayrılır (112). Judo sporcuları uluslararası müsabakalarda derece alarak nihai hedef olan olimpiyat yeterliliğine sahip olmak için oyunlardan önce uluslararası tüm müsabakalara katılması önemli bir kriterdir (113).

Alanyazında klasik elit sporcu seçme yöntemlerine alternatif olarak veri madenciliği yöntemlerinin kullanılabilmesine dair az sayıda çalışma bulunmaktadır. Örneğin; Nagovitsyn ve ark. tarafından 2023 yılında yapılan bir çalışmada “Rusya Spor Ustası Adayı” veya “Rusya Spor Ustası” ünvanına sahip 19 sporcu, III-I spor kategorisine sahip 21 sporcu ve küçük spor kategorisine sahip yada kategorisi olmayan 32 sporcu olmak üzere toplam 72 güreş sporcusuna ait 36 özellik içeren bireysel veriler güreşçilerin arşivlenmiş verilerinden (1988-2008), sosyal ağlardan ve telefon anketlerinden toplanarak 12 özellik spor alanı, 13 özellik kalıtsal veriler ve 11 özellik bireysel başarılar üç temel kategoriye ayrılmıştır (114). Yapay sinir ağları, rastgele orman ve lojistik regresyon algoritmaları aracılığıyla en yüksek standartta performans gösteren ve bu standardı başaramayan sporcular şeklinde belirlenen iki tahmin kategorisine göre Orange programında sınıflama yapılmıştır. Model yardımıyla, kontrol grubunu oluşturan 18 güreşçiden belirli bir güreşçinin rekabetçi performansının yalnızca ikisinde güvenilir bir şekilde hata kaydedilmiş böylece güreşçilerin rekabetçi başarısına ilişkin tahminlerin uygulanmasında %11 oranında yanlış cevap verme olasılığını verdiği sonucuna ulaşmışlardır. Ayrıca oluşturulan model yüksek bir spor sonucu elde edemeyen 10 sporcudan oluşan bir kontrol grubu için test edildiğinde yalnızca tek bir hata ile tahmin performansı göstermiştir. Belirlenen özellikleri kategoriye göre sınıflandırması alanındaki işlevselliğine gelince, modelin %100 olasılıkla genç bir güreşçinin gelecekteki spor performansının sonuçlarını güvenilir şekilde etkileyen temel özellik kategorilerini belirlediği ve böylece veri madenciliği algoritmalarının kullanımının spor seçiminin kalitesini

arttırmakta olduğu ve bu da genç güreşçilerin eğitim sürecinin zamanında kişiselleştirilmesine ve iyileştirilmesine olanak sağlayacağı sonucuna ulaşmışlardır.

Değer ve Süel (113), tarafından 2022 yılında yapılan bir çalışmada, YSA ile judo müsabakalarına ait sonuçların kestirilmesi amaçlanmıştır. Sporculara ait veriler (sıklet, ülke, derece, müsabaka sayısı, kazanma oranı, paun ve ceza puanları vs.) açık erişimle ulaşılabilen Uluslararası Judo Federasyonu web sitesinden (<https://judobase.ijf.org/>) alınmıştır. Verilerin analizi için Python'un 3.8.2 versiyonu ile NumPy kütüphanesinin 1.20.3 versiyonundan yararlanılmıştır. Tahmin amacıyla YSA da katman, nöron sayıları ve optimizasyon yöntemleri (SGD, RMSprop, Adam, Adadelta, Adagrad, Adamax, Nadam) hiperparametreleri ile 21 farklı model oluşturmuşlardır. 2017-2021 tarihleri arasında 53775 büyükler judo müsabakalarında yarışan 7758 sporcuya ait her bir sporcu için 14 farklı öznelik hesaplanmıştır. Rakip 2 sporcuya ait 28 adet girdi verisi ve bir adet çıktı verisi belirlenmiştir. Model başarısı en yüksek olanın (%78,6 doğruluk) tek katmanda 64 nöron, optimizasyon yöntemi olarak RMSprop kullanılan model olduğu belirlenmiştir. Bu çalışma ile Judo müsabaka sonuçlarının veri madenciliği yöntemleri ile tahmin edilebileceğini ortaya koymuşlardır. Araştırmacıların yaptıkları literatür taramasında araştırmalarına örnek teşkil edecek bir çalışmaya rastlanmadığına ek olarak doğrusal model uygulanan çalışmalarda %26 doğrulukla tahmin yapıldığına değinmişlerdir (115).

Sporla başarıyı yakalamak için oyuncuların analizi ve potansiyellerinin keşfi gerekmektedir. Geleneksel yöntemlerle yapılan bu seçimler teknolojinin gelişmesiyle yerini veri odaklı seçime bırakmıştır. Bu da sporcu seçiminde daha nesnel kararların alınmasını mümkün kılmaktadır (116).

Ivankovic ve arkadaşlarının (2010) çalışması, Sırbistan Basketbol Liginde ait 2005-2010 sezonları arası verileri çok katmanlı algılayıcı (MLP) tipi yapay sinir ağı modeliyle, oyuncu seçim sürecine yönelik istatistiksel bir çözüm önermektedir. Modelde giriş değişkenleri olarak sayı, asist, ribaunt gibi temel oyuncu performans metrikleri kullanılmış, çıkış değişkeni olarak ise "potansiyel oyuncu değeri" öngörülmüştür. Modelin doğruluk oranı %85'in üzerinde raporlanırken, hata kareler ortalaması (MSE) ve korelasyon katsayısı (R) gibi istatistiksel performans göstergeleri

kullanılmıştır. Ancak çalışmanın sınırlı veri seti üzerinde gerçekleştirilmiş olması ve modelin çapraz doğrulama dışında başka bir genellenebilirlik testine tabi tutulmaması, sonuçların geniş ölçekli uygulamalara doğrudan aktarılabilirliğini sorgulatmaktadır. Sırbistan Basketbol Liginde ait 5 sezonluk (2005-2010) verileri sinir ağı algoritması kullanılarak analiz edilmiş ve bu veriler sporcu seçiminde de kullanılmıştır (117).

Benzer şekilde Barron ve arkadaşları (2018), profesyonel futbolcuların transfer süreçlerinde karar destek mekanizmalarında katkı sağlamak amacıyla yapay sinir ağları (NN) kullanmıştır. ProZone MatchViewer sistemi ve çeşitli veri tabanları aracılığıyla isabetli pas, top kaybı, top kazanma, şut ve koşu mesafesi gibi diğer birçok performans göstergesi model girdisi olarak kullanılmış ve oyuncular başarı potansiyeline göre sınıflandırılması ve hangi oyuncuların Premier Lig seviyesine yükselebileceğinin belirlenmesi hedeflenmiştir. Kullanılan 5 katlı çapraz doğrulama yöntemi ve ROC eğrisi (EAA = 0.91) gibi istatistiksel göstergeler, modelin yüksek doğrulukla çalıştığını göstermektedir. Ayrıca Cohen's Kappa katsayısı ile sınıflandırma tutarlılığı ölçülmüş ve anlamlı düzeyde uyum saptanmıştır. Ancak modelde kullanılan değişkenlerin seçimi ve normalizasyon süreci üzerine detay verilmemesi, model şeffaflığını ve tekrarlanabilirliğini sınırlayan bir unsurdur (6).

Pappalardo ve arkadaşları (2019) tarafından geliştirilen "*PlayeRank*" ise çok boyutlu ve pozisyon bazlı oyuncu değerlendirmesi yapabilmektedir. Bu model, oyuncuların saha içi performanslarını değerlendirerek, her oyuncuya sayısal bir derecelendirme vermektedir. 19 milyondan fazla aksiyon verisini işleyen bu sistemde destek vektör makineleri (DVM), regresyon analizleri ve karar ağaçları gibi farklı makine öğrenimi algoritmaları kıyaslanmıştır. Model, oyuncuların sahadaki farklı rollerine göre özelleştirilmiş performans skorları üretmektedir ve bu skorlar Spearman korelasyon katsayısı ile istatistiksel olarak anlamlı bulunmuştur. Ayrıca varyans analizi (ANOVA) ile pozisyonlar arası farklılıklar test edilmiş, modelin uzman değerlendirmeleri ile yüksek düzeyde örtüştüğü bildirilmiştir ($p < 0.01$). Bu çalışma, özellikle model girdilerinin çeşitliliği ve algoritma karşılaştırmaları açısından örnek teşkil edici niteliktedir (118).

Genel olarak deęerlendirildięinde, incelenen alıřmalarda yapay zeka tabanlı veri madencilięi yntemlerinin sporcu performans analizinde yksek doęruluk ve istatistiksel tutarlılık saęladıęı grlmektedir. Bununla birlikte, bazı alıřmalarda veri n iřleme srecinin eksik aktarılması, model girdilerinin seilme kriterlerinin net belirtilmemesi ve dıř geerlik testlerinin sınırlılıęı, akademik literatrde daha fazla metodolojik aıklıęa ihtiya duyulduęunu ortaya koymaktadır. zellikle veri madencilięi yntemlerinin spor bilimlerine entegrasyonunda, model performansları kadar bu modellerin řeffaflıęı, tekrarlanabilirlięi ve etik boyutları da gelecekte tartıřılması gereken bařlıca konular arasında yer almaktadır.

Bu tez alıřmasında veri madencilięi yntemleri ile elitlik tahmin modeli geliřtirmek amacıyla, elit sporcu sınıflandırmasında altın standart olarak alınan federasyonların klasik yntemler ile setikleri elit ve elit olmaya aday havuzdaki sporcuların, belirli periyotlar ile rutin olarak toplanan etiketli performans test verileri kullanılmıř ve yksek tahmin performansına sahip modellere ulařılmıřtır. Bu veriler sporcu performans lm yapaan tm merkezlerin veri tabanlarında mevcuttur. Bylelikle, sporcuların gerek lm verilerini kullanarak klasik sporcu seimine alternatif olabilecek bir sınıflama modelinin uygulanabilirlięi gsterilmiřtir. Kullanılan elitlik sınıflandırma modeli sporcu seiminde kullanılan msabaka dereceleri ve sporcuların performans lmlerinin gerekleřtirildięi sporcu performans lm merkezlerinin yapısına gre performansı etkileyen dięer kriterlerin (psikolojik, sosyolojik, beslenme vb.) lm ile beraber veri setinin daha fazla eřitlilik iermesini saęlayabilir. Bylece daha fazla gzlem ve seim kriterlerini oluřturan daha fazla zellik eklenerek modellerin genelleme yeteneęi arttırılabilir. alıřmamızdaki verileri sadece fizyolojik faktrler, motorsal ve fiziksel parametreleri ieren test verilerinde oluřması alıřmamızın sınırlılıkları arasındadır.

6. SONUÇ ve ÖNERİLER

Sporcu performansını ölçmeye yönelik saha ve laboratuvar ortamında sporcuların belirli periyotlarda tekrarlanan testlerden elde edilen verilerin veri madenciliği yöntemleri ile incelenip sporcuların sınıflandırılması amacıyla elde edilen modeller elit sporcu seçimi sürecinde ve bu seçimde etki eden performans değişkenlerinin belirlenmesinde antrenörler ve federasyon yetkililerinin karar almasında katkı sağlayacaktır.

Bu çalışmada elit ve elit olmaya aday sporcuların sınıflandırılmasında kullanılan veri madenciliği yöntemlerinin performanslarının incelenmiştir. RO ve XGBoost modelleri, elitlik sınıflandırmasında en iyi sonuç veren modeller olmuştur.

Bu çalışmada uygulanan elitlik sınıflandırma tahmin modelleri, sadece bir performans ölçme değerlendirme laboratuvarında mücadele sporları kategorisindeki judo ve güreş branşındaki sporcularından alınan fizyolojik faktörler, motorsal ve fiziksel değişkenleri içeren test verileri ile oluşturulmuştur. Elitlik sınıflandırma tahmin modellerinin çok sayıda branşa ait ve daha fazla sportif performans test verisi ile kurulması ve yine aynı gruptan sporcu verileri ile modelin test edilmesi ve güncellenmesi modellerin genellenebilirliğini arttıracaktır.

Çalışmamızda, sporcu sınıflandırmasında performans test sonuçları gibi sürekli sayısal değişken içeren veri setlerinde güçlü sonuçlar veren sınıflandırıcılar kullanılmıştır. Bu modellerin özellik seçimindeki başarısı ve genelleme yetenekleri sporcu seçimlerinde karar destek sistemlerine entegre edilerek daha isabetli seçimler yapılmasına yardımcı olabilir. Bu modeller, sporcuların performansını, fiziksel durumunu ve potansiyelini doğru şekilde analiz ederek, antrenörlere ve spor yöneticilerine, sporcu seçiminde yardımcı olarak nesnel karar alınabilmesinde fayda sağlayacaktır. Sporcu seçiminin sadece veriye dayalı karar vermeden ibaret olmadığı göz önüne alınarak, sporcu seçimlerinde takım içi iletişim, motivasyon, psikolojik destek, birebir ilgi ve durum değerlendirmeleri gibi antrenörlerin sporcu ile birebir iletişimde olduğu durumların önemini kavrayarak, karar verme süreçlerini ve stratejik planlamalarını destekleyen güçlü ve nesnel araçlar olarak veri madenciliği yöntemleri

ile spor bilimlerinde bir uygulama amaçlanmıştır. Bu analiz yöntemleri antrenörlerin yerini alacak kadar bağımsız çalışmamakla beraber veriye dayalı yaklaşımlar ancak sporcu performansı artırma ve sporcu seçimi karar süreçlerinde sporculara, antrenörlere, federasyonlara ve spor yöneticilerine yardımcı olacaktır.



7. KAYNAKLAR

1. Oliver JL, Ayala F, Croix MBDS, Lloyd RS, Myer GD, Read PJ. Using machine learning to improve our understanding of injury risk and prediction in elite male youth football players. *Journal of science and medicine in sport*. 2020;23(11):1044-1048
2. Bilek G, Ulas E. Predicting match outcome according to the quality of opponent in the English premier league using situational variables and team performance indicators. *International Journal of performance analysis in sport*. 2019;19(6):930-941
3. Campbell PG, Stewart IB, Sirotic AC, Drovandi C, Foy BH, Minett GM. Analysing the predictive capacity and dose-response of wellness in load monitoring. *Journal of sports sciences*. 2021;39(12):1339-1347
4. Min B, Kim J, Choe C, Eom H, McKay RB. A compound framework for sports results prediction: A football case study. *Knowledge-Based Systems*. 2008;21(7):551-562
5. Ivanković Z, Racković M, Markoski B, Radosav D, Ivković M, editors. Analysis of basketball games using neural networks. 2010 11th international symposium on computational intelligence and informatics (CINTI); 2010: IEEE;
6. Barron D, Ball G, Robins M, Sunderland C. Artificial neural networks and player recruitment in professional soccer. *PloS one*. 2018;13(10):e0205818
7. Sheng B, Li P, Zhang Y, Mao L, Chen CP. GreenSea: visual soccer analysis using broad learning system. *IEEE Transactions on Cybernetics*. 2020;51(3):1463-1477
8. Solieman OK. Data mining in sports: A research overview. *Dept of Management Information Systems*. 2006
9. Guth LM, Roth SM. Genetic influence on athletic performance. *Current opinion in pediatrics*. 2013;25(6):653-658
10. Iso-Ahola S. Intrapersonal and interpersonal factors in athletic performance. *Scandinavian journal of medicine & science in sports*. 1995;5(4):191-199
11. Feltz DL. Self-confidence and sports performance. *Exercise and sport sciences reviews*. 1988;16(1):423-458
12. Pedersen DM. Perceived relative importance of psychological and physical factors in successful athletic performance. *Perceptual and motor skills*. 2000;90(1):283-290
13. MacArthur DG, North KN. Genes and human elite athletic performance. *East African Running*. 2007:241-257

14. Hrysomallis C. Balance ability and athletic performance. *Sports medicine*. 2011;41:221-232
15. Ostrander EA, Huson HJ, Ostrander GK. Genetics of athletic performance. *Annual review of genomics and human genetics*. 2009;10(1):407-429
16. Heishman AD, Curtis MA, Saliba E, Hornett RJ, Malin SK, Weltman AL. Noninvasive assessment of internal and external player load: implications for optimizing athletic performance. *The Journal of Strength & Conditioning Research*. 2018;32(5):1280-1287
17. Scantlebury S, Till K, Beggs C, Dalton-Barron N, Weaving D, Sawczuk T, et al. Achieving a desired training intensity through the prescription of external training load variables in youth sport: More pieces to the puzzle required. *Journal of sports sciences*. 2020;38(10):1124-1131
18. Paulson TA, Mason B, Rhodes J, Goosey-Tolfrey VL. Individualized internal and external training load relationships in elite wheelchair rugby players. *Frontiers in physiology*. 2015;6:388
19. Guest NS, Horne J, Vanderhout SM, El-Sohemy A. Sport nutrigenomics: personalized nutrition for athletic performance. *Frontiers in nutrition*. 2019;6:8
20. Crowcroft S, Slattery K, McCleave E, Coutts AJ. Do athlete monitoring tools improve a coach's understanding of performance change? *International journal of sports physiology and performance*. 2020;15(6):847-852
21. Lindberg K, Solberg P, Bjørnsen T, Helland C, Rønnestad B, Frank MT, et al. Strength and power testing of athletes: Associations of common assessments over time. *International Journal of Sports Physiology and Performance*. 2022;17(8):1280-1288
22. McMaster DT, Gill N, Cronin J, McGuigan M. A brief review of strength and ballistic assessment methodologies in sport. *Sports Medicine*. 2014;44:603-623
23. Fleming P, Young C, Dixon S, Carré M. Athlete and coach perceptions of technology needs for evaluating running performance. *Sports Engineering*. 2010;13:1-18
24. Grehaigne J-F, Godbout P, Bouthier D. Performance assessment in team sports. *Journal of teaching in Physical Education*. 1997;16(4):500-516
25. Dimitrijevic M, Paunovic V, Zivkovic V, Bolevich S, Jakovljevic V. Body Fat Evaluation in Male Athletes from Combat Sports by Comparing Anthropometric, Bioimpedance, and Dual-Energy X-Ray Absorptiometry Measurements. *BioMed Research International*. 2022;2022(1):3456958
26. Mayhew J, Piper F, Ware J. Anthropometric correlates with strength performance among resistance trained athletes. *The Journal of sports medicine and physical fitness*. 1993;33(2):159-165

27. Brocherie F, Girard O, Forchino F, Al Haddad H, Dos Santos GA, Millet GP. Relationships between anthropometric measures and athletic performance, with special reference to repeated-sprint ability, in the Qatar national soccer team. *Journal of sports sciences*. 2014;32(13):1243-1254
28. Wixted AJ, Thiel DV, Hahn AG, Gore CJ, Pyne DB, James DA. Measurement of energy expenditure in elite athletes using MEMS-based triaxial accelerometers. *IEEE Sensors Journal*. 2007;7(4):481-488
29. Reilly T, Bangsbo J, Franks A. Anthropometric and physiological predispositions for elite soccer. *Journal of sports sciences*. 2000;18(9):669-683
30. Currell K, Jeukendrup AE. Validity, reliability and sensitivity of measures of sporting performance. *Sports medicine*. 2008;38:297-316
31. Tanner R, Gore C. *Physiological tests for elite athletes: Human kinetics*; 2012.
32. Jones MT, Matthews TD, Murray M, Van Raalte J, Jensen BE. Psychological correlates of performance in female athletes during a 12-week off-season strength and conditioning program. *The Journal of Strength & Conditioning Research*. 2010;24(3):619-628
33. Winter EM, Jones AM, Davison RR, Bromley PD, Mercer TH. *Sport and exercise physiology testing guidelines: volume I—Sport testing: the British association of sport and exercise sciences guide: Routledge*; 2006.
34. Cates W, Cavanaugh J. *Advances in rehabilitation and performance testing. Clinics in Sports Medicine*. 2009;28(1):63-76
35. Kazman JB, de la Motte S, Gribbin TC, Galecki J, Deuster PA, O'Connor F. Improving physical performance tests: time to include a psychologist. *BMJ Publishing Group Ltd and British Association of Sport and Exercise Medicine*; 2016. p. 1290-1291
36. Duda J. *Advances in sport and exercise psychology measurement: Fitness Information Technology*; 1998.
37. Comfort P, Dos' Santos T, Beckham GK, Stone MH, Guppy SN, Haff GG. Standardization and methodological considerations for the isometric midhigh pull. *Strength & Conditioning Journal*. 2019;41(2):57-79
38. Brady CJ, Harrison AJ, Comyns TM. A review of the reliability of biomechanical variables produced during the isometric mid-thigh pull and isometric squat and the reporting of normative data. *Sports biomechanics*. 2020
39. Townsend JR, Bender D, Vantrease WC, Hudy J, Huet K, Williamson C, et al. Isometric midhigh pull performance is associated with athletic performance and sprinting kinetics in division I men and women's basketball players. *The Journal of Strength & Conditioning Research*. 2019;33(10):2665-2673

40. Thomas C, Comfort P, Chiang C-Y, Jones PA. Relationship between isometric mid-thigh pull variables and sprint and change of direction performance in collegiate athletes. *Journal of trainology*. 2015;4(1):6-10
41. James LP, Roberts LA, Haff GG, Kelly VG, Beckman EM. Validity and reliability of a portable isometric mid-thigh clean pull. *The Journal of strength & conditioning research*. 2017;31(5):1378-1386
42. Zupan MF, Arata AW, Dawson LH, Wile AL, Payn TL, Hannon ME. Wingate anaerobic test peak power and anaerobic capacity classifications for men and women intercollegiate athletes. *The Journal of strength & conditioning research*. 2009;23(9):2598-2604
43. Zacharogiannis E, Paradisis G, Tziortzis S. An evaluation of tests of anaerobic power and capacity. *Medicine & Science in Sports & Exercise*. 2004;36(Supplement):S116
44. Coppin E, Heath EM, Bressel E, Wagner DR. Wingate anaerobic test reference values for male power athletes. *International journal of sports physiology and performance*. 2012;7(3):232-236
45. Bringhurst RF, Wagner DR, Schwartz S. Wingate anaerobic test reliability on the Velotron with ice hockey players. *The Journal of Strength & Conditioning Research*. 2020;34(6):1716-1722
46. Yousaf R, Marwat M. Effect of moderate level aerobic training on maximum oxygen consumption and cardiovascular fitness of non-athlete students. *Global Educational Studies Review*, VII. 2022
47. Bosquet L, Léger L, Legros P. Methods to determine aerobic endurance. *Sports medicine*. 2002;32:675-700
48. Edvardsen E, Hem E, Anderssen SA. End criteria for reaching maximal oxygen uptake must be strict and adjusted to sex and age: a cross-sectional study. *PloS one*. 2014;9(1):e85276
49. Ahmaidi S, Collomp K, Caillaud C, Prefaut C. Maximal and functional aerobic capacity as assessed by two graduated field methods in comparison to laboratory exercise testing in moderately trained subjects. *International journal of sports medicine*. 1992;13(03):243-248
50. Born D-P, Zinner C, Düking P, Sperlich B. Multi-directional sprint training improves change-of-direction speed and reactive agility in young highly trained soccer players. *Journal of Sports Science & Medicine*. 2016;15(2):314
51. Spasic M, Krolo A, Zenic N, Delextrat A, Sekulic D. Reactive agility performance in handball; development and evaluation of a sport-specific measurement protocol. *Journal of sports science & medicine*. 2015;14(3):501

52. Sheppard J, Young WB, Doyle T, Sheppard T, Newton RU. An evaluation of a new test of reactive agility and its relationship to sprint speed and change of direction speed. *Journal of science and medicine in sport*. 2006;9(4):342-349
53. Sekulic D, Krolo A, Spasic M, Uljevic O, Peric M. The development of a New Stop'n'go reactive-agility test. *The Journal of strength & conditioning research*. 2014;28(11):3306-3312
54. Reilly T, Williams AM, Richardson D. Identifying talented players. *Science and soccer*: Routledge; 2003. p. 315-334
55. Lorenz DS, Reiman MP, Lehecka B, Naylor A. What performance characteristics determine elite versus nonelite athletes in the same sport? *Sports health*. 2013;5(6):542-547
56. Aouadi R, Jlid M, Khalifa R, Hermassi S, Chelly MS, Van Den Tillaar R, et al. Association of anthropometric qualities with vertical jump performance in elite male volleyball players. *The Journal of sports medicine and physical fitness*, 2012;52(1):11-17
57. Armstrong N, McManus A. Exercise testing elite young athletes. *Med Sport Sci*. 2011;56:106-125
58. Abdelkrim NB, Chaouachi A, Chamari K, Chtara M, Castagna C. Positional role and competitive-level differences in elite-level men's basketball players. *The Journal of Strength & Conditioning Research*. 2010;24(5):1346-1355
59. Hrysomallis C, Buttifant D. Influence of training years on upper-body strength and power changes during the competitive season for professional Australian rules football players. *Journal of science and medicine in sport*. 2012;15(4):374-378
60. Ahmetov II, Rogozkin VA. Genes, athlete status and training—An overview. *Med Sport Sci*. 2009;54(4):43-71
61. Augustsson J, Thomee R. Ability of closed and open kinetic chain tests of muscular strength to assess functional performance. *Scandinavian journal of medicine & science in sports*. 2000;10(3):164-168
62. Broussal-Derval A. *Training and conditioning for judo*: Human Kinetics Publishers; 2021.
63. Franchini E, Del Vecchio FB, Matsushige KA, Artioli GG. Physiological profiles of elite judo athletes. *Sports medicine*. 2011;41:147-166
64. Mirzaei B, Curby DG, Rahmani-Nia F, Moghadasi M. Physiological profile of elite Iranian junior freestyle wrestlers. *The Journal of strength & conditioning research*. 2009;23(8):2339-2344

65. Latyshev M, Tropin Y, Pryimakov O, Curby D, Dokmanac M, Balic M, et al. Greco-Roman Wrestling on the World Stage: Performance Trends and Country Comparisons. *Ido Movement for Culture Journal of Martial Arts Anthropology*. 2024;24(4):33-39
66. Ersöz F, Çınar Y. Veri madenciliği ve makine öğrenimi yaklaşımlarının karşılaştırılması: Tekstil sektöründe bir uygulama. *Avrupa Bilim ve Teknoloji Dergisi*. 2021(29):397-414
67. Larose DT. *Data mining and predictive analytics*: John Wiley & Sons; 2015.
68. Han J, Kamber M, Pei J. *Data mining: Concepts and Techniques*, Waltham: Morgan Kaufmann Publishers. 2012
69. Fayyad UM, Piatetsky-Shapiro G, Smyth P, editors. *Knowledge Discovery and Data Mining: Towards a Unifying Framework*. KDD; 1996
70. Sullivan R. *Introduction to data mining for the life sciences*: Springer Science & Business Media; 2012.
71. Torgo L. *Data mining with R: learning with case studies*: Chapman and Hall/CRC; 2011.
72. Aaboub F, Chamlal H, Ouaderhman T. Statistical analysis of various splitting criteria for decision trees. *Journal of Algorithms & Computational Technology*. 2023;17:17483026231198181
73. Kabak M, ÇINAR Y. *Çok Kriterli Karar Vermenin Esasları*. 2020
74. Zhang H, Gu C-L, Gu L-w, Zhang Y. The evaluation of tourism destination competitiveness by TOPSIS & information entropy—A case in the Yangtze River Delta of China. *Tourism management*. 2011;32(2):443-451
75. Lee J-S. AUC4. 5: AUC-based C4. 5 decision tree algorithm for imbalanced data classification. *IEEE Access*. 2019;7:106034-106042
76. Sharma DN, Iqbal S. Applying Decision Tree Algorithm Classification and Regression Tree (CART) Algorithm to Gini Techniques Binary Splits. *Int J Eng Adv Technol*. 2023;12(5):77-81
77. Tibshirani RJ, Efron B. *An introduction to the bootstrap*. Monographs on statistics and applied probability. 1993;57(1):1-436
78. Breiman L. Bagging predictors. *Machine learning*. 1996;24:123-140
79. Freund Y. Boosting a weak learning algorithm by majority. *Information and computation*. 1995;121(2):256-285

80. Chen T, Guestrin C, editors. XGBoost: A scalable tree boosting system. Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining; 2016
81. Jiang H, He Z, Ye G, Zhang H. Network intrusion detection based on PSO-XGBoost model. IEEE Access. 2020;8:58392-58401
82. Alfaro E, Gamez M, Garcia N. adabag: An R package for classification with boosting and bagging. Journal of Statistical Software. 2013;54:1-35
83. Breiman L. Random forests. Machine learning. 2001;45:5-32
84. Liang J, Bi G, Zhan C. Multinomial and ordinal Logistic regression analyses with multi-categorical variables using R. Annals of translational medicine. 2020;8(16):982
85. Vapnik V. The nature of statistical learning theory: Springer science & business media; 2013.
86. Olson DL, Delen D. Advanced data mining techniques: Springer Science & Business Media; 2008.
87. Aydın Haklı D. Sınıf dengesizliği sorununu çözmek için kullanılan algoritmaların farklı sınıflandırma yöntemlerinde performanslarının karşılaştırılması. 2018
88. Han J, Pei J, Tong H. Data mining: concepts and techniques: Morgan kaufmann; 2022.
89. Wang J. Encyclopedia of data warehousing and mining: iGi Global; 2005.
90. Mitchell TM, Mitchell TM. Machine learning: McGraw-hill New York; 1997.
91. Jiarui H, Dinghan K, Yimeng M, editors. An Improved Pseudo Nearest Neighbor: Minkowski Distance-weighted Classification Algorithm. 2024 IEEE 4th International Conference on Electronic Technology, Communication and Information (ICETCI); 2024: IEEE;
92. Kalra V, Kashyap I, Kaur H, editors. Effect of distance measures on K-nearest neighbour classifier. 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA); 2022: IEEE;
93. Wiley J. Data mining for business intelligence. Concepts, Techniques, and Applications in Microsoft Office Excel with Xlminer. 2010
94. Taşcı E, Onan A. K-en yakın komşu algoritması parametrelerinin sınıflandırma performansı üzerine etkisinin incelenmesi. Akademik Bilişim. 2016;1(1):4-18

95. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*. 2002;16:321-357
96. He H, Garcia EA. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering*. 2009;21(9):1263-1284
97. Balaban M, Kartal E. Veri Madenciliği Süreci ve Futbol Maç Sonuçlarının Öngörülmesine İlişkin Bir Uygulama. R ile Veri Madenciliği Uygulamaları İstanbul: Çağlayan Kitapevi. 2016
98. KARAAĞAOĞLU A, KARAKAYA KARABULUT J, KILIÇKAP M. Tanı testlerinin değerlendirilmesinde istatistiksel yöntemler. 2016
99. KILIÇ S. ROC analysis in clinical decision making. *Psychiatry and Behavioral Sciences*. 2013;3(3):135
100. Tharwat A. Classification assessment methods. *Applied computing and informatics*. 2021;17(1):168-192
101. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *biometrics*. 1977:159-174
102. Lin TY, Yao YY, Zadeh LA. *Data mining, rough sets and granular computing: Physica*; 2013.
103. Araújo D, Couceiro M, Seifert L, Sarmiento H, Davids K. *Artificial intelligence in sport performance analysis: Routledge*; 2021.
104. Jayal A, McRobert A, Oatley G, O'Donoghue P. *Sports analytics: Analysis, visualisation and decision making in sports performance: Routledge*; 2018.
105. Li W. *Applied mathematics and nonlinear sciences. Appl Math Nonlinear Sci*. 2024;9(1):1-15
106. Gennadievich BA. Machine learning and data mining activity results when using projectiles in different sports. *International Journal*. 2020;9(3)
107. Dong X, Sun Y. Injury Risk Prediction and Prevention Algorithm for Athletes Based on Data Mining. *Journal of Electrical Systems*. 2024;20:1717-1728
108. Murathan T, Devocioğlu S. VERİ MADENCİLİĞİ VE SPOR ALANINDAKİ UYGULAMALARI. *Spor Bilimleri Dergisi*. 2018;29(3):147-156
109. DiCesare CA, Green B, Yuan W, Diekfuss JA, Barber Foss KD, Dudley J, et al. Machine learning classification of verified head impact exposure strengthens associations with brain changes. *Annals of biomedical engineering*. 2020;48:2772-2782

110. Ówiklinski B, Gielczyk A, Choraś M. Who will score? a machine learning approach to supporting football team building and transfers. *Entropy*. 2021;23(1):90
111. Miljković D, Gajić L, Kovačević A, Konjović Z, editors. The use of data mining for basketball matches outcomes prediction. *IEEE 8th international symposium on intelligent systems and informatics*; 2010: IEEE;
112. Biricik YS. Türkiye Olimpiyat Hazırlık Merkezi (TOHM) sporcularının tükenmişlik düzeyleri ile psikolojik iyi oluşları arasındaki ilişkide dayanıklılığın aracılık etkisi. *The Online Journal of Recreation and Sports*. 2023;12(2):70-84
113. Değer Ö, Süel E. Gerçek judo müsabaka sonuçlarının yapay sinir ağları yöntemi yolu ile karşılaştırılması. *Spor Eğitim Dergisi*. 2022;6(2):88-98
114. Nagovitsyn RS, Valeeva RA, Latypova LA. Artificial intelligence program for predicting wrestlers' sports performances. *Sports*. 2023;11(10):196
115. Franchini E, Julio UF. The judo world ranking list and the performances in the 2012 London Olympics. *Asian journal of sports medicine*. 2015;6(3):e24045
116. Dinca-Panaitescu T, Dinca-Panaitescu S. Artificial intelligence in the sports industry. *AI and Society: Chapman and Hall/CRC*; 2023. p. 113-125
117. Ivanković Z, Racković M, Markoski B, Radosav D, Ivković M. Appliance of neural networks in basketball scouting. *Acta Polytechnica Hungarica*. 2010;7(4):167-180
118. Pappalardo L, Cintia P, Ferragina P, Massucco E, Pedreschi D, Giannotti F. PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. *ACM Transactions on Intelligent Systems and Technology (TIST)*. 2019;10(5):1-27

8. EKLER

EK-1: Hacettepe Üniversitesi Etik Komisyon İzni

Tarih: 18/01/2024 14:13
Sayı: E-16969557-030.01.04-
00003302430



00003302430



HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ARAŞTIRMA ETİK KURULU

KURUL KARARI

<u>OTURUM TARİHİ</u>	<u>OTURUM SAYISI</u>	<u>KARAR SAYISI</u>
09.01.2024	2024/01	2024/01-11
Araştırma Numarası : SBA 24/017		Değerlendirme Tarihi : 09.01.2024

Üniversitemiz Tıp Fakültesi Biyoistatistik Anabilim Dalı öğretim üyelerinden Prof. Dr. Erdem KARABULUT'un sorumlu araştırmacı olduğu, Adeviye Gamze AKIN'ın yüksek lisans tezi olan, SBA 24/017 kayıt numaralı *"Elit ve Elit Olmaya Aday Sporcuların Performanslarını Etkileyen Değişkenlerin Veri Madenciliği Yöntemleri ile İncelenmesi ve Sporcuların Sınıflandırılması"* başlıklı araştırma önerisi gerekçe, amaç, yaklaşım ve yöntemleri dikkate alınarak incelenmiş olup, 01 Ocak 2018 – 31 Aralık 2023 tarihleri arasındaki arşiv kayıtlarının 10 Ocak 2024 – 10 Ocak 2025 tarihleri arasında geçerli olmak üzere incelenmesi etik açıdan **uygun bulunmuştur**.

Çalışma tamamlandığında sonuçlarını içeren bir rapor örneğinin Etik Kurulumuza gönderilmesi gerekmektedir.

Prof. Dr. Nüket
PAKSOY ERBAYDAR
Kurul Başkanı

Prof. Dr. Güzide Burça
AYDIN
Kurul Üyesi

Prof. Dr. Mehmet Özgür
UYANIK
Kurul Üyesi

Prof. Dr. Ayşe KİN
İŞLER
Kurul Üyesi

Prof. Dr. Burcu Balam
DOĞU
Kurul Üyesi

Prof. Dr. Tolga
YILDIRIM
Kurul Üyesi

Prof. Dr. İpek GÜRBÜZ
Kurul Üyesi

Prof. Dr. Betül ÇELEBİ
SALTIK
Kurul Üyesi

Doç. Dr. Merve BATUK
Kurul Üyesi

Doç. Dr. Gülten IŞIK
KOÇ
Kurul Üyesi

Dr. Öğr. Üyesi Melike
Hacer ÖZKAN
Kurul Üyesi

Dr. Öğr. Üyesi Müge
DEMİR
Kurul Üyesi

Dr. Öğr. Üyesi Burcu
Ersöz ALAN
Kurul Üyesi

EK-2: Turnitin Orijinallik Raporu Ekran Görüntüsü

Elit Ve Elit Olmaya Aday Sporcuların Sınıflandırılmasında Kullanılan Veri Madenciliği Yöntemlerinin Performanslarının İncelenmesi

ORJİNALLİK RAPORU

% 6	% 5	% 3	% 1
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	openaccess.hacettepe.edu.tr İnternet Kaynağı	% 2
2	acikbilim.yok.gov.tr İnternet Kaynağı	% 1
3	openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	<% 1
4	dergipark.org.tr İnternet Kaynağı	<% 1
5	www.openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	<% 1
6	Mersin, Hacer Tuba. "Yüzücülerde Miyofasyal Gevşeme Tekniği İle Uygulanan Farklı ısınma Egzersiz Modellerinin Ayak Bileği Esnekliğine ve Yüzme Performansına Olan Akut Etkilerinin Karşılaştırılması", Marmara Üniversitesi (Turkey), 2023 Yayın	<% 1
7	Ozlem Yalcin, Alpaslan Erman, Sedat Muratli, Melek Bor-Kucukatay, Oguz K. Baskurt. "Time course of hemorheological alterations after heavy anaerobic exercise in untrained human subjects", Journal of Applied Physiology, 2003 Yayın	<% 1

EK-3: Dijital Makbuz**Dijital Makbuz**

Bu makbuz ödevinizin Turnitin'e ulaştığını bildirmektedir. Gönderiminize dair bilgiler şöyledir:

Gönderinizin ilk sayfası aşağıda gönderilmektedir.

Gönderen: Adeviye Gamze AKIN
Ödev başlığı: Elit Ve Elit Olmaya Aday Sporcuların Sınıflandırılmasında Kulla...
Gönderi Başlığı: Elit Ve Elit Olmaya Aday Sporcuların Sınıflandırılmasında Kulla...
Dosya adı: Adeviye_Gamze_Ak_n_18.06.20205.pdf
Dosya boyutu: 2.79M
Sayfa sayısı: 102
Kelime sayısı: 20,990
Karakter sayısı: 130,814
Gönderim Tarihi: 18-Haz-2025 12:08ÖS (UTC+0300)
Gönderim Numarası: 2701631599

T.C.
HACETTEPE ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

ELİT VE ELİT OLMAYA ADAY SPORCULARIN
SINIFLANDIRILMASINDA KULLANILAN VERİ
MADENCİLİĞİ YÖNTEMLERİNİN PERFORMANSLARININ
İNCELENMESİ

Adeviye Gamze AKIN

Biyostatistik Programı
YÜKSEK LİSANS TEZİ

ANKARA
2025

9. ÖZGEÇMİŞ

Adı Soyadı: Adeviye Gamze AKIN	
Doğum Tarihi:	
E-Posta:	

EĞİTİM

Derece	Alan	Üniversite	Yıl
Lisans			
Yüksek Lisans			

MESLEKİ DENEYİM

Meslek	Kurum	Yıl

KONGRE, ÇALIŞTAY, EĞİTİM VE KONFERANSLAR

Tarih	Kongre, Çalıştay, Eğitim ve Konferanslar