

**T.C.**  
**BAHÇEŞEHİR ÜNİVERSİTESİ**  
**LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**  
**İŞLETME MÜHENDİSLİĞİ ANA BİLİM DALI**  
**İŞLETME MÜHENDİSLİĞİ DOKTORA PROGRAMI**

**RASTGELE ORMAN ALGORİTMASI TEMELLİ NESNEL BİR ÖLÇÜT**  
**AĞIRLIKLANDIRMA YÖNTEMİ ÖNERİSİ VE İNSAN KAYNAKLARI ANALİTİĞİ**  
**İÇİN DİNAMİK BİR UYGULAMA**

**DOKTORA TEZİ**  
**BATUHAN BİLENLER**

**İSTANBUL 2025**

T.C.  
BAHÇEŞEHİR ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ  
İŞLETME MÜHENDİSLİĞİ ANA BİLİM DALI  
İŞLETME MÜHENDİSLİĞİ DOKTORA PROGRAMI

RASTGELE ORMAN ALGORİTMASI TEMELLİ NESNEL BİR ÖLÇÜT  
AĞIRLIKLANDIRMA YÖNTEMİ ÖNERİSİ VE İNSAN KAYNAKLARI ANALİTİĞİ  
İÇİN DİNAMİK BİR UYGULAMA

DOKTORA TEZİ  
BATUHAN BİLENLER

TEZ DANIŞMANLARI  
DOÇ. DR. SAİT GÜL  
DR. ÖĞR. ÜYESİ TAMER UÇAR

İSTANBUL 2025

**DOKTORA TEZ SAVUNMA ONAY FORMU**

<b>Adı Soyadı</b>	Batuhan Bilenler
<b>Öğrenci No</b>	2100214
<b>Program Adı</b>	İşletme Mühendisliği
<b>Tezin Adı</b>	Rastgele Orman Algoritması Temelli Nesnel Bir Ölçüt Ağırlıklandırma Yöntemi Önerisi ve İnsan Kaynakları Analitiği İçin Dinamik Bir Uygulama
<b>Tez Savunma Tarihi</b>	23.5.2025

Bu tezin Doktora tezi olarak gerekli şartları yerine getirmiş olduğu Lisansüstü Eğitim Enstitüsü tarafından onaylanmıştır.

**Doç. Dr. Yücel Batu Salman**  
**Enstitü Müdürü**

Bu Tez tarafımızca okunmuş, nitelik ve içerik açısından bir Doktora tezi olarak yeterli görülmüş ve kabul edilmiştir.

<b>Doktora Tez Savunma Jürisi</b>			
<b>Tez Sınav Jürisi</b>	<b>Unvanı Adı Soyadı</b>	<b>Kurumu</b>	<b>İmza</b>
Tez Danışmanı	Doç. Dr. Sait Gül	Bahçeşehir Üniversitesi	
Tez İzleme Komitesi Üyesi	Dr. Öğr. Üyesi M. Aslı Aydın	Bahçeşehir Üniversitesi	
Tez İzleme Komitesi Üyesi (Kurum Dışı)	Dr. Öğr. Üyesi Adem Özyavaş	İstanbul Atlas Üniversitesi	
Üye	Doç. Dr. C. Okan Şakar	Bahçeşehir Üniversitesi	
Üye (Kurum Dışı)	Prof. Dr. Ferdi Sönmez	İstanbul Aydın Üniversitesi	

**Bu tezdeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.**

Ad, Soyad : Batuhan Bilenler

İmza :

## ÖZET

# RASTGELE ORMAN ALGORİTMASI TEMELLİ NESNEL BİR ÖLÇÜT AĞIRLIKLANDIRMA YÖNTEMİ ÖNERİSİ VE İNSAN KAYNAKLARI ANALİTİĞİ İÇİN DİNAMİK BİR UYGULAMA

BİLENLER, Batuhan

İşletme Mühendisliği Doktora Programı

Tez Danışmanları: Doç. Dr. Sait GÜL, Dr. Öğr.

Üyesi Tamer UÇAR

Mayıs 2025, 78 sayfa

Bu çalışmada, araştırma yapılması için paylaşılmış açık kaynak bir veri seti kullanılarak personelin işten ayrılmasının tahminlenebilmesi ve özniteliklerin bu modele hangi oranda katkı sunduğuna dair nesnel bir ağırlıklandırma yöntemi (RAFOW) önerilmektedir. Makine öğrenmesinin veriden anlam çıkararak, probleme özgü öznitelik ağırlıkları belirleyebilmesi literatüre sağladığı yeni bir katkıdır. Rastgele orman tekniği ile ağaçlar oluşturulurken düğümlere öznitelikleri rastgele olarak atama özelliğinden faydalanarak, ağaçların seviye 0, seviye 1 ve seviye 2 seviyelerindeki bölünme/aday oranları hesaplanmış ve o model üzerindeki ağırlıklar normalize edilerek gösterilmiştir. Hali hazırda Sklearn kütüphanesinin özellik önem hesaplama tekniği olan her ağaçtaki safsızlık azalması prensibi ile ağırlık çıkarılması yöntemiyle birlikte bu çalışmada ileri sürülen yöntemin karşılaştırılması yapılmıştır. RAFOW yönteminin, sklearn kütüphanesi kullanılarak elde edilen özellik önem dağılımına göre daha dengeli bir ağırlandırma yapabildiği, önemsiz gibi görünen özelliklerinin öneminin daha yüksek olabildiğini gösteren kazanımlar sağlamıştır. İnsan kaynakları analitiğinde, RAFOW ve ARAS yöntemlerinin birlikte kullanımında açıklanabilirlik, kolaylık, anlaşılabilirlik ve uygulanabilirlik gibi faydalar bu çalışmada ortaya çıkarılmıştır. Personel sayısı değiştikçe, oluşan yeni veri seti üzerinde oldukça kısa bir sürede yeni

öznitelik ağırlıkları dinamik şekilde oluşturulabileceđi bu alıřmada uygulamalı şekilde ortaya konulmuřtur.

**Anahtar Kelimeler:** Makine öğrenmesi, RAFOW, Karar destek sistemleri, İnsan kaynakları analitiđi



## ABSTRACT

# A RANDOM FOREST ALGORITHM-BASED OBJECTIVE ATTRIBUTE WEIGHTING METHOD PROPOSITION AND ITS DYNAMIC APPLICATION ON HUMAN RESOURCES ANALYTICS

BİLENLER, Batuhan

PhD Program in Management Engineering

Supervisors: Doç. Dr. Sait GÜL, Dr. Öğr. Üyesi

Tamer UÇAR

May 2025, 78 page

In this study, an objective weighting method (RAFOW) is proposed to predict employee turnover and determine the contribution levels of features to this model using a publicly available open-source dataset. The ability of machine learning to derive insights from data and assign problem-specific feature weights represents a novel contribution to the literature. By utilizing the feature assignment characteristic of the random forest technique during tree construction, split/candidate ratios at levels 0, 1, and 2 were calculated, and the weights on the model were normalized and presented. A comparison was made between the proposed method and the impurity reduction principle-based feature importance calculation technique in Scikit-learn's library. The study highlights the benefits of using RAFOW and the ARAS methods together in human resources analytics, emphasizing advantages such as explainability, simplicity, comprehensibility, and applicability. Furthermore, it is demonstrated through practical implementation that dynamic feature weights can be generated in a very short time for a new dataset as the number of employees changes.

**Keywords:** Machine learning, RAFOW, Decision support systems, Human resources analytics

Bu alıřmayı beni yetiřtirip byten annem Yeter BİLENLER ve rahmetli olan babam Demir BİLENLER'e ithaf ediyorum.

## TEŐEKKÜR

Bu tez alıőmasının planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteęini esirgemeyen, engin bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle alıőmamı bilimsel temeller ışığında őekillenmesine desteklerinden dolayı sayın hocalarım Do. Dr. Sait GÜL ve Dr. Öğr. Üyesi Tamer Uar'a sonsuz teőekkürlerimi sunarım.

Bu zorlu süreçte canım eőim ve kıymetli aileme gösterdikleri sabır, anlayış ve desteklerinden ötürü őükranlarımı sunarım.



## İÇİNDEKİLER

ETİK BEYAN.....	iii
ÖZET.....	iv
ABSTRACT.....	v
İTHAF.....	vi
TEŞEKKÜR.....	vii
İÇİNDEKİLER.....	viii
TABLolar LİSTESİ.....	x
ŞEKİL LİSTESİ.....	xi
KISALTMALAR LİSTESİ.....	xii
Bölüm 1: Giriş.....	1
1.1 Çalışmanın Kapsamı.....	2
1.2 Çalışmanın Amacı ve Önemi.....	3
Bölüm 2: İnsan Kaynakları Analitiği Yazını.....	8
Bölüm 3: Makine Öğrenmesi Teknikleri.....	13
3.1 Gözetimli Öğrenme Teknikleri.....	15
3.1.1 Regresyon Teknikleri.....	15
3.1.1.1 Doğrusal Regresyon.....	15
3.1.1.2 Lojistik Regresyon.....	18
3.1.1.3 Ridge ve Lasso Regresyonu.....	18
3.1.2 Destek Vektör Makineleri (SVM).....	19
3.1.3 Karar Ağaçları.....	20
3.1.3.1 Rastgele Orman (RO) Yöntemi.....	22
3.1.3.2 Gradyan Artırma Yöntemi (Gradient Boosting).....	25
3.1.3.3 Ekstrem Gradyan Artırma Yöntemi (XGBoosting).....	27
3.1.3.4 CHAID Yöntemi.....	30
3.1.3.5 CART Yöntemi.....	31
3.1.4 Naive Bayes.....	32
3.1.5 Bayes Ağları.....	33
3.2 Makine Öğrenmesinde Model Değerlendirme Ölçütleri.....	33
3.2.1 Regresyon Modellerinin Değerlendirme Ölçütleri.....	34
3.2.2 Sınıflandırma Modellerinin Değerlendirme Ölçütleri.....	35
Bölüm 4: Çok Kriterli Karar Verme.....	37

4.1 ÇKKV Yöntemlerinde Ağırlıklandırma Yöntemleri.....	40
4.1.1 Öznel (Subjektif) Ağırlıklandırma Yöntemleri.....	41
4.1.2 Nesnel (Objektif) Ağırlıklandırma Yöntemleri.....	42
4.2 ARAS Yöntemi.....	43
Bölüm 5: Yeni Bir Nesnel Ölçüt Ağırlıklandırma Yöntemi Önerisi: RAFOW.....	47
Bölüm 6: İnsan Kaynakları Analitiğinde Bir Uygulama.....	49
6.1 Veri Ön İşleme .....	51
6.2 Makine Öğrenmesi Yöntemleri ile Tahminleme Yapılması.....	51
Bölüm 7: RAFOW Destekli Dinamik ARAS.....	62
Bölüm 8: Ek Analizler.....	65
8.1 Sadece Ağırlığı En Yüksek Çıkan 4/5/6 Sayıda Öznitelik ile Modelin Çalıştırılması.....	65
8.2 Personel Sayısının Değişmesi Durumu .....	68
8.3 Yeni Bir Veri Setinde Yöntemin Uygulanması .....	69
8.4 SHAP Yöntemi ile Ağırlıklandırma Analizi.....	70
Bölüm 9: Tartışma ve Sonuçlar .....	74
KAYNAKÇA.....	79
EKLER	
A. Personelin İşten Ayrılmasını Etkileyen Kriterler ve Çalışmalar.....	93

## TABLULAR LİSTESİ

### TABLULAR

Tablo 1 İnsan Kaynakları Analitiği Yazın Özeti.....	9
Tablo 2 Makine Öğrenmesi ile İK Analitiğinde Yapılan Sınıflandırma ve Regresyon Çalışmaları.....	10
Tablo 3 AHP tercih ölçeği .....	39
Tablo 4 Veri setini oluşturan özniteliklerin yapısı .....	49
Tablo 5 Literatürde personelin işten ayrılmasını etkileyen parametreler.....	50
Tablo 6 Verilerin temel istatistikleri.....	51
Tablo 7 Yöntemlerin performans sonuçları.....	52
Tablo 8 Sınıflandırma metriklerinin gösterimi.....	53
Tablo 9 Önerilen RAFOW tekniğine göre özellik önemlerinin gösterilmesi.....	55
Tablo 10 RAFOW ve Sklearn kütüphanesine göre bulunan öznitelik ağırlıkları .....	55
Tablo 11 CRITIC, Entropi ve RAFOW yöntemleriyle kriter ağırlıklarının gösterimi.....	57
Tablo 12 Beş anahtar personele ilişkin normalize karar matrisi.....	59
Tablo 13 Personellerin ARAS skorlarının gösterimi.....	61
Tablo 14 Farklı parametre sayılarına göre elde edilen değerlendirme metrikleri .....	67
Tablo 15 Ek analiz için değerlendirme metrikleri.....	67
Tablo 16 Personel sayısının değişmesi durumunda ağırlıkların değişimi.....	68
Tablo 17 Yeni veri seti için bulunan ağırlıkların karşılaştırması.....	69
Tablo 18 SHAP yöntemiyle elde edilen ağırlıkların gösterimi.....	71

## ŞEKİLLER LİSTESİ

### ŞEKİLLER

Şekil 1 İki farklı kategoride olan verilerin düzleme yerleştirilmiş gösterimi.....	19
Şekil 2 İki farklı kategorideki verileri ayıran eğrinin gösterimi.....	19
Şekil 3 Dönüşüm sonrası verilerin birbirinden ayrılabilmesi.....	20
Şekil 4 Verilen örneğin AHP yöntemine göre mimari yapısının gösterimi.....	38
Şekil 5 Oluşturulan ağaç modellerinden birinci sıradaki ağacın gösterimi.....	53
Şekil 6 ARAS skorunu hesaplayabilmek için yapılan geliştirmeler.....	60
Şekil 7 Karar Destek Sisteminin Modellenmesi.....	64

## KISALTMALAR LİSTESİ

ARAS	Additive Ratio Assessment System
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
KV	Karar Verici
OOB	Özellik Önem Skoru
PCA	Principal Component Analysis
RAFOW	Random Forest Algorithm-based Objective Attribute
SVM	Support Vector Machine
XGBoosting	Extreme Gradient Boosting

## Bölüm 1

### Giriş

Bir şirketin hedeflerine ulaşması için insan kaynağı yönetimine önem vermesi gerekmektedir. Hızlı değişen piyasa koşullarına göre dinamik şekilde karar alabilmeli ve şirketi hedeflerine taşıyacak personelin mutluluğunu göz ardı etmemesi gereklidir. Özellikle anahtar rolde bulunan personeli şirkete tutundurma amacıyla şirketlerin üst yönetimleri çeşitli metotlar uygulayarak personellerin tatminliğini artırıcı hamleler yaptığı görülmektedir. Gün geçtikçe farklılaşan sosyal, ekonomik ve teknolojik koşullar, her alanda olduğu gibi insan kaynakları alanında da yeni ihtiyaçları doğurmakta ve bu ihtiyaçlar yeni kavramların, teorilerin ve uygulamaların ortaya çıkmasına neden olmaktadır (Reşitoğlu, 2011).

Kurum kültürünü taşıyan ve yansıtan personellerin geçmiş verileri kullanılarak, işten ayrılma kararının tahminlenebilmesi şirket yönetimi için oldukça önemli bir veridir. Personeller firmalar için aynı zamanda bir iç müşteri niteliği taşımaktadırlar. Müşterileri elinde tutma, sadece en iyi ve faydalı müşterileri rakiplerden korumak değil, aynı zamanda onlar arasındaki ilişki ve bağlantıları sürdürmek ve böylece hizmet kullanımını ve buna bağlı geliri korumaktır (Gülpınar, 2013). İşletmeler için önemli değerlerden birisi de kayıpların önceden doğru şekilde tahmin edilebilmesidir (Yalçın, 2019).

İşletmeler için insan kaynağı, sürdürülebilir rekabet avantajı sağlamada kritik bir faktör olarak öne çıkmaktadır. İnsan kaynakları, işletmenin stratejik hedeflerine ulaşmasını destekleyen en önemli unsur olarak kabul edilir (Noe, Hollenbeck, Gerhart ve Wright, 2015). Çalışanların bilgi, beceri ve yetkinlikleri, işletmenin pazarda konumlanmasını ve sürekli gelişim göstermesini mümkün kılar. Bu bağlamda, insan kaynağına yapılan yatırımlar, işletmelerin uzun vadeli başarısını ve karlılığını doğrudan etkiler (Boxall ve Purcell, 2016).

İnsan kaynaklarının yönetimi, işletmenin genel performansını artıran önemli bir süreçtir. Çalışan motivasyonu, performans değerlendirme sistemleri ve kariyer geliştirme fırsatları gibi unsurlar, işletmenin etkinliğini doğrudan etkilemektedir (Kramar, 2014). Ayrıca, insan kaynağının iyi yönetilmesi, çalışanların işletmeye olan bağlılığını artırarak yüksek çalışan

devir oranının önüne geçilmesini sağlar. Günümüzün dinamik iş dünyasında, yetenek yönetimi ve insan kaynaklarının stratejik rolü her zamankinden daha büyüktür (Sparrow, Brewster, Chung, 2016).

İnovasyon ve yaratıcılık, insan kaynağı ile doğrudan ilişkilidir. Özellikle bilgiye dayalı sektörlerde, çalışanların bilgi birikimi ve yenilikçi düşünce yapısı, işletmenin rekabet gücünü artırmaktadır (Stone ve Deadrick, 2015). İyi eğitilmiş ve motive çalışanlar, yeni ürün ve hizmetlerin geliştirilmesinde, işletmelerin değişen müşteri taleplerine yanıt verme kapasitesini artırmaktadır. Bu durum, insan kaynaklarının işletmeler için stratejik bir değer taşıdığını ortaya koymaktadır.

Teknolojik gelişmeler ve dijital dönüşüm, insan kaynaklarının yönetimini daha da karmaşık hale getirmiştir. Özellikle uzaktan çalışma, esnek çalışma modelleri ve dijital beceri gereksinimleri, işletmelerin insan kaynağı stratejilerini yeniden şekillendirmelerini zorunlu kılmaktadır (Cascio ve Montealegre, 2016). Dijitalleşen dünyada başarılı olabilmek için işletmelerin, çalışanlarının dijital becerilerini geliştirmesi ve onları sürekli eğitmesi gerekmektedir. Bu da insan kaynaklarına yönelik stratejik yatırımların önemini artırmaktadır.

İşletmelerde kapsayıcılık ve çeşitlilik yönetimi, insan kaynağının değerini daha da yükseltmiştir. Çeşitli yeteneklere sahip bir iş gücü, işletmenin küresel pazarlarda daha esnek ve yaratıcı olmasını sağlar (Shen, Tang ve D'Netto, 2010). Bu tür bir çeşitlilik, farklı bakış açılarını bir araya getirerek yenilikçi çözümler üretme potansiyelini artırır. Dolayısıyla, işletmelerin insan kaynaklarına yönelik stratejileri hem çalışan memnuniyetini hem de genel iş performansını yükselten bir unsurdur.

## **1.1 Çalışmanın Kapsamı**

Bu çalışmada, araştırma yapılabilmesi amacıyla açık kaynak olarak paylaşılmış bir insan kaynakları veri seti kullanılarak makine öğrenmesi yöntemlerinden olan Rastgele Orman (RO) yöntemi ile işten ayrılacak olan personelin tahminlemesi yapılmaktadır. RO yönteminin rastgele olarak düğümlere atama yapması prensibinden faydalanarak objektif ve yeni bir öznitelik ağırlık hesaplama tekniği önerilmektedir. Mevcutta sklearn kütüphanesinin öznitelik önemi hesaplamak için kullandığı saflılık

tabanlı gini indeksi yaklaşımı ile sonuçlar karşılaştırmalı olarak irdelenmektedir. Önerilen yeni öznitelik ağırlıklandırma tekniği ile elde edilen ağırlıklar, ARAS yöntemine girdi olarak verilmekte ve dinamik bir sistem tasarımı oluşturulmaya çalışılmaktadır. Makine öğrenmesi ve ARAS yöntemlerinin birlikte kullanılmasıyla birlikte en ufak bir veri değişikliğinde karar destek sisteminin güncel sonuç üretebilmesi yetkinliği dinamikliğin temel kaynağıdır.

## 1.2 Çalışmanın Amacı ve Önemi

İşletme yönetimi, mühendislik, finans, iş güvenliği ve işçi sağlığı, insan kaynakları, enerji gibi alanlardaki birçok karar verme sorunu, çok kriterli karar verme (ÇKKV) problemi olarak kategorize edilebilir. Küreselleşmenin etkisiyle bugün, işletmeler ve değişik yapıdaki organizasyonlar ile hükümetler karar verirken pek çok unsuru birlikte değerlendirmek zorundadır. Yönetimsel karar verme sürecinde üç temel bileşen vardır: uzmanlık ve görüşlerine başvuru alan karar vericiler (KV), çözüm olarak değerlendirilen alternatifler ve kararı etkileyen ölçütler. Bu süreçte, söz konusu ölçütlerin nasıl ağırlıklandırılacağı önemli bir husustur.

ÇKKV problemlerinde, alternatifler her ölçüt açısından puanlanır ve bu verileri işlemek için çeşitli ÇKKV yöntemleri kullanılarak bir karara varılır. Süreçte her ölçütün ne kadar önemli olduğu dikkate alınmalıdır, çünkü her ölçüt, problemin özelliklerine ve gereksinimlerine göre farklı bir önem düzeyine sahip olabilmektedir. Bu niteliklerin değerlendirilmesi ve önem atfedilmesi süreci, ölçüt ağırlıklandırma olarak bilinir.

Bu doktora çalışmasının özgün değerini 2 altbaşlık halinde özetlenebilir:

*1. Rastgele orman algoritması kullanarak yeni bir nesnel ağırlıklandırma tekniği önerilmektedir.*

İnsan Kaynakları Analitiği'nde (İKA), çalışanların kararlarını hangi niteliklerin yönlendirdiğini ve bunların ne kadar önemli olduğunu anlamak, çözülmesi gereken kritik bir aşamadır (Aydoğdu ve Gül, 2022). Nitelik ağırlıklarının belirlenmesi için iki ana yöntem vardır. Bu yöntemlerden biri özneye dayalı yaklaşım olan kişinin uzmanlığına ve/veya görüşlerine dayanan tekniklerdir. Bunlar ikili karşılaştırma, doğrudan atama veya Simos

prosedürü gibi yöntemlerdir. Buna karşılık, ikinci grubu oluşturan nesnel yöntemler kişisel tercihlere dayanmaz. Bunun yerine, doğrudan alternatiflerin performans puanlarına odaklanarak, önyargılı veya manipülatif karar vericilerden kaynaklanabilecek riskleri azaltmaya ve uzun veri toplama sürelerinden kaçınmaya çalışır. Entropi tabanlı ağırlıklandırma, CRITIC, LOPCOW en çok atıf alan nesnel nitelik ağırlıklandırma yöntemlerindedir.

Bu çalışmada, İKA alanında bir makine öğrenmesi algoritması uygulanarak, hangi çalışan(lar)ın işten ayrılma eğiliminde olduğu ve bu eğilimi en çok hangi niteliklerin yönlendirdiği anlaşılmasına çalışılmaktadır. Niteliklerin önceliklendirilmesinin ardından, gerekli önlemler alınarak şirketin kilit personelini kaybetme riski azaltılabilir veya ortadan kaldırılabilir.

RO algoritması, makine öğrenmesinde yaygın olarak kullanılan ve güçlü bir sınıflandırma ve regresyon yöntemidir. Bu teknik, bağımsız olarak eğitilen birden fazla karar ağacından oluşur. Rastgele orman modelinin nihai çıktısı, bu bireysel ağaçlar tarafından yapılan tahminlerin ortalaması alınarak veya çoğunluk oyu kullanılarak belirlenir. Bu yapı, aşırı öğrenme riskini azaltır ve modelin genel performansını artırır. Özellikle verilerin yüksek boyutlu ve karmaşık olduğu durumlarda, RO yöntemi yüksek doğruluk sunar ve değişkenler arasındaki etkileşimleri etkili bir şekilde yakalar.

İnsan kaynakları alanında RO yöntemi, veri odaklı karar verme süreçlerini çeşitli uygulamalar aracılığıyla kolaylaştırır. Örneğin, çalışan performans değerlendirmelerini, terfi olasılıklarını veya çalışan kaybı riskini tahmin etmek için kullanılabilir. Bu tahminler, insan kaynakları departmanlarının daha etkili stratejiler geliştirmesine yardımcı olur, bu da çalışan memnuniyetini ve verimliliğini artırır. Ayrıca, işe alım süreçlerinde RO yöntemi, adayları değerlendirerek ve geçmiş veri setlerine dayanarak hangi adayların başarılı olma olasılığının yüksek olduğunu tahmin etmek için kullanılabilir. Bu sayede, organizasyonlar daha bilinçli ve nitelikli işe alım kararları alabilirler.

RO'da ağırlık belirleme, değişken önem derecesi ile ilişkilidir. Gini indeksine dayalı olarak yapılan bölünmelerin sayısı ve katkısı, bir özelliğin önem derecesini hesaplamak için kullanılmaktadır. Örneğin, bir özellik

ağacın birçok düğümünde büyük bir saflılık artışı sağlıyorsa, bu özellik daha önemli kabul edilir. Gini indeksi, RO'da ölçütlerin önem derecelerini belirlemeye dolaylı olarak katkıda bulunur. Daha fazla ve daha etkili bölünmeyi sağlayan özellikler, model tarafından daha yüksek ağırlıklarla değerlendirilir. Bu sayede, modelin hangi özelliklere daha fazla güvendiğini ve tahminlerinde bu özelliklere öncelik verdiğini gösterir.

RO yönteminde, her karar ağacı bir düğümde bir ayırım yapabilmek için rastgele seçilmiş bir alt özellik kümesi kullanır. Bu süreç sırasında, bir özelliğin belirli bir düğümde ne sıklıkla seçildiği ve o düğümdeki ayırımın modelin doğruluğuna ne kadar katkıda bulunduğu izlenir. Bu çalışmada ileri sürülen Rastgele Orman Algoritmasına Dayalı Nesnel Ölçüt Ağırlıklandırma (Random Forest-based Objective Attribute Weighting – RAFOW) yöntemindeki ayırım/aday (split/candidate) oranı, dolaylı olarak nitelik ağırlıklarının (özellik öneminin) belirlenmesinde kullanılmaktadır. Sonuç olarak, RAFOW mantığında bir ölçütün birçok ağaçta etkili ve sık kullanılması durumunda daha önemli kabul edilebileceği varsayımını dayanılır. Bu değerlendirme, önem göstergesi olarak kabul edilir ve böylelikle saflılık bakış açısından farklı bir bakış açısı ile önem ağırlığı değerlendirilir.

RAFOW yöntemi, literatürde ilk kez sunulan benzersiz bir bakış açısı sunmaktadır. RO tekniğiyle ağaçlar oluşturulurken niteliklerin düğümlere rastgele atanma özelliğinden yararlanılarak ağaçların 0. seviye, 1. seviye ve 2. seviyelerindeki ayırım/aday oranları hesaplanmış ve modelin üzerindeki etkileri gösterilmiştir. Bu bakış açısı, bir niteliğin bir düğüm için kaç kez rastgele şekilde önerildiği ve kaç kez seçildiğini oranlayarak, saflık perspektifinden farklılık göstermektedir. Bu nedenle, ÇKKV literatüründe nesnel nitelik ağırlıklandırma alanında tamamen yeni bir yaklaşım olarak önerilmektedir çünkü nitelik ağırlıkları, herhangi bir başka ağırlıklandırma prosedürü dikkate alınmaksızın, doğrudan bir makine öğrenmesi algoritmasının çalışma mekanizmasından ortaya çıkarılmaktadır. Yöntemin uygulanabilirliği, açık kaynaklı bir veri seti üzerinde sunulmuştur.

## 2. *Dinamik Şekilde Ölçüt Ağırlıklarını Güncelleyen Makine Öğrenmesi Algoritması Tabanlı ARAS Karar Destek Sisteminin Oluşturması.*

ARAS yönteminde kullanılan ölçüt ağırlıkları makine öğrenmesi modellerinden elde edilebilir çünkü ARAS yaklaşımın kendisinde zaten özgün bir ağırlık belirleme yaklaşımı bulunmamaktadır. Örneğin, çalışan memnuniyetinin bir organizasyondaki önemi zamanla değişebilir, yeni personeller gelebilir veya mevcut personeller ayrılabilir. Makine öğrenmesi modeli güncellenerek özneliklerin ağırlıkları herhangi bir  $t$  anında değişebilir ve ARAS yönteminde kullanılacak ağırlıklar güncellenebilir. Bu sayede, hem veriden anlam çıkarılarak güncel ağırlıklar bulunabilir hem de anahtar personellerin herhangi bir  $t$  anındaki ayrılma durumları ARAS ile sıralabilir.

ARAS yöntemi ile makine öğrenmesi tekniklerinin birlikte kullanarak elde edebileceği kazanımları aşağıdaki şekilde sınıflandırabiliriz:

- Şeffaflık ve karmaşık veri analizi kabiliyeti.
- Dinamik kriter ağırlıkları elde etme.
- Hibrit karar destek sistemlerinden gelen doğrulama ve güvenilirlik.
- Esneklik ve Uyarlanabilirlik.
- Genel Performans Artışı.

(Chen ve Guestrin, 2016) yapmış oldukları çalışmada, ARAS ve makine öğrenmesi tekniklerinin birlikte kullanılmasıyla daha güçlü ve esnek karar destek sistemleri oluşturulabileceğini söylemiştir. Bu sistemler hem veri analizinde hem de karar verme süreçlerinde daha kapsamlı çözümler sunabilirler. ARAS yöntemi, hangi kriterlerin ne kadar önemli olduğunun bilgisini karar analistinden talep eder. Bu nedenle ARAS genel olarak bir alternatif sıralama tekniği olarak sınıflandırılabilir (Turskis ve Zavadskas, 2010). ARAS yöntemi, farklı türdeki kriterlerle çalışabilir ve makine öğrenmesi ile kullanıldığında, veri türlerine ve problemlere daha esnek ve uyarlanabilir çözümler sunabilir (Turskis, Zavadskas ve Peldschus, 2009).

Makine öğrenmesi, büyük veri setlerindeki karmaşık ilişkileri modelleyebilir ve tahminler yapabilir. Bu, daha karmaşık ve veri yoğun

problemlerde üstün performans sağlar (Murphy, 2012). (Shmueli, Bruce, Yahav, Patel ve Lichtendahl, 2017) yaptıkları çalışmada, bir organizasyonda çalışanların işten ayrılma olasılığını tahmin eden makine öğrenmesi modelini ARAS yöntemi ile kullanılarak hangi çalışanların daha kritik olduğunun belirlenmesinde ve bu bilgilere dayalı proaktif önlemler alınabilmesinde kullanmışlardır.



## Bölüm 2

### İnsan Kaynakları Analitiği Yazını

İşletmelerin sürdürülebilir olmasında, insan kaynağını doğru yönetebilmeleri oldukça önemlidir. Kurum kültürüne uyum sağlamış başarılı bir personelin, şirketten ayrılmayı düşünüp düşünmediğini bilmek insan kaynakları yönetiminin öncelikli görevlerindedir. Değişen kültürel, ekonomik, sosyolojik ve teknolojik şartlarla birlikte personellerin öncelikleri ve memnuniyeti değişebilmektedir. Günümüzde, bu memnuniyeti ölçümleyebilmek amacıyla işletmeler çeşitli ileri analitik süreçleri, kendi iç problemlerini çözmek amacıyla kullanabilirler. İKA, insan kaynakları süreçlerinin her biri hakkında daha iyi içgörü sağlamak, ilgili verileri toplamak ve daha sonra bu süreçlerin nasıl iyileştirileceğine dair objektif kararlar almak amacıyla kullanılan süreç, hesaplama veya modellerdir.

İnsan kaynakları analitiği alanında literatür taraması gerçekleştirildiğinde analitik çalışmaların, personel seçim problemi ve mevcut personelin ayrılma eğiliminin tahminlenmesi üzerine yoğunlaştığı görülmektedir. Personelin işten ayrılma (staff turnover) tahmini probleminin ele alınış yöntemi olarak farklı farklı yaklaşımlara başvurulduğu görülmektedir. Shankar, Rajanikanth, Sivaramaraju ve Murthy (2018) , veri madenciliği tekniklerini kullanarak personelin işten ayrılma problemini istatistiksel bakış açısı ile tahminlemeye çalışmıştır. Alao ve Adeyemo (2013), karar ağacı yöntemini kullanarak geçmiş verileri analiz ederek personelin ayrılma tahmininin yapılabileceğini göstermiştir. Abdulsalam, Ajao, Balogun ve Arowolo (2022) ise, RO ve evrişimli sinir ağları (convolutional neural network) algoritmaları kullanarak performans analizi gerçekleştirmiş ve trend olan iki yöntemi derinlemesine inceleyerek performans karşılaştırması yapmıştır. Gao, Wen ve Zhang (2019) ise, RO yönteminin diğer makine öğrenmesi yöntemlerinden daha başarılı sonuçlar elde ettiğini gösteren bir karşılaştırmalı analiz sunmuşlardır. İKA’da ele alınan temel problemler ve kullanılan yöntemler Tablo 1’de gösterilmektedir.

Tablo 1  
*İnsan Kaynakları Analitiği Yazını Özeti*

<b>Yazar(lar), Yıl</b>	<b>Ele Alınan Problem</b>
Burgess, Lane ve Stevens (2001)	İş gücü akışları ve işten ayrılmalar
Clark (2001)	İş tatmini ve işten ayrılma nedenleri
Dale-Olsen (2006)	Ücretler ve işten ayrılma ilişkisi
Delfgaauw (2007)	İş tatmininin iş arama üzerindeki etkisi
Lévy-Garboua, Montmarquette ve Simonnet (2007)	İş tatmini ve işten ayrılma ilişkisi
Böckerman ve Ilmakunnas (2009)	İş olumsuzlukları ve işten ayrılma niyetleri
Cottini, Kato ve Westergaard-Nielsen (2011)	Olumsuz iş koşulları ve işten ayrılma
Jantan, Hamdan ve Othman (2010)	Personel yetenek yönetimi
Saradhi ve Palshikar (2011)	Çalışan devri tahmini
Yang, Wan ve Fu (2012)	Uluslararası otellerde çalışan devri ve tutma stratejileri
Alao ve Adeyemo (2013)	IT firmalarında çalışan devrini etkileyen faktörler
Davis ve Haltiwanger (2014)	İşgücü piyasası akışkanlığı ve ekonomik performans
Katsikea, Theodosiou, Perdikis ve Kehagias (2011)	Organizasyonel yapı ve iş özelliklerinin iş tatmini üzerindeki etkileri
Punnoose ve Ajit (2016)	Kurumlarda çalışan devrini tahmin etme
Parry (2016)	Çalışma yerinde kuşak çeşitliliği
Alamsyah ve Salma (2018)	Çalışan devrini tahmin etme
Fletcher, Alfes ve Robinson (2016)	Eğitim olanakları ve iş tatmini ilişkisi
Stamolampros, Korfiatis, Chalvatzis ve Buhalis (2019)	İş tatmini ve çalışan devri belirleyicileri
Yuan (2021)	Çalışan devri tahmin etme
Li ve Fox (2023)	Çalışan devri tahmin etme
Kumar, Gaikwad, Ramya, Tiwari, Tiwari ve Kumar (2023)	Çalışan devri tahmin etme

Tablo 1’de insan kaynağı analitiği alanında çözülen problemler ve bu problemleri çözme yöntemleri sınıflandırılmıştır. Görüldüğü gibi yazında çoğunlukla, personelin işten ayrılması ve personelin işe alımı konularında yoğunlaşmış olup, yöntemsel açıdan özellikle son yıllarda makine öğrenmesi tekniklerinin tercih edildiği görülmektedir.

Personelin işten ayrılma tahminleme problemi özelinde inceleme yapıldığı takdirde probleme daha çok makine öğrenmesi teknikleriyle yaklaşıldığı görülmüştür. Tablo 2’de görüldüğü gibi personelin işten ayrılması tahminleme problemini ele alırken makine öğrenmesi teknikleri uygulanmakta ancak çalışmanın türü, değerlendirme metrikleri ve

çalışmanın hangi veri seti üzerinde yapıldığı bilgileri gösterilmektedir.

Tablo 2  
Makine Öğrenmesi ile İK Analitiğinde Yapılan Sınıflandırma ve Regresyon Çalışmaları

Yazar(lar), Yıl	Çalışmanın Türü	Değerlendirme Metrikleri	Çalışma Yapılan Veri Seti
Alao ve Adeyemo (2013)	Sınıflandırma	Doğruluk oranı, F1 değeri	Nijerya Yüksek Enstitüsünün 1978-2006 arasında çalışan personel listesi
Alam, Mohiuddin, Islam, Hassan, Hoque, Allayear (2018)	Regresyon	Kök Ortalama Kare Hatası, Ortalama Mutlak Hata	Çalışmada kullanılan veri seti Kaggle üzerinde mevcut olan IBM HR Analytics Employee Attrition and Performance'dır
Alamsyah ve Salma (2018)	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
Alduayj ve Rajpoot (2018)	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık, F1 değeri	Çalışmada kullanılan veri seti IBM HR Analytics Employee Attrition and Performance veri seti olarak belirtilmiştir.
Reddy, Rajanikanth, Sivaramaraju ve Murthy (2018)	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık, F1 değeri	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
Gao, Wen ve Zhang (2019)	Sınıflandırma	Doğruluk oranı, ROC Eğrisi altındaki alan, F1 değeri	Çin'de görev yapan bir Telekom şirketinin personel veri seti kullanılmıştır.
Dutta ve Bandyopadhyay (2020)	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık, F1 değeri	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
Subhashini ve Gopinath (2020)	Sınıflandırma	Doğruluk oranı, ROC Eğrisi altındaki alan	Kaggle platformundaki IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.

Tablo 2 (devam)

Joseph, Udupa, Jangale, Kotkar, Pawar (2021)	Sınıflandırma	Doğruluk oranı, ROC Eğrisi altındaki alan, F1 değeri	IBM HR Analytics Employee Attrition and Performance veri setini kullanmışlardır.
Qutub, Al-Mehmadi, Al-Hssan, Aljohani, Alghamdi (2021)	Sınıflandırma	Doğruluk oranı, F1 değeri, ROC Eğrisi altındaki alan	IBM HR Analytics Employee Attrition veriseti kullanılmıştır.
Abdulsalam, Ajao, Balogun, Arowolo (2022)	Sınıflandırma	Doğruluk oranı, Keskinlik, Duyarlılık, F1 değeri	Bir telekomünikasyon şirketine ait gerçek veriler kullanılmıştır, ancak şirket adı belirtilmemiştir.
Atef, Elzanfaly ve Ouf (2022)	Regresyon	Kök Ortalama Kare Hatası, Ortalama Mutlak Hata	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
Poornappriya ve Gopinath (2021)	Sınıflandırma	Doğruluk oranı, F1 değeri, ROC Eğrisi altındaki alan	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.
Sari ve Lhaksmana (2022)	Sınıflandırma	Doğruluk oranı, F1 değeri, ROC Eğrisi altındaki alan	Çalışmada kullanılan veri seti IBM HR Analytics Employee Attrition and Performance veri seti olarak belirtilmiştir.
Chen, Lin, ve Zhan (2023)	Sınıflandırma	Doğruluk oranı, F1 değeri, ROC Eğrisi altındaki alan	Çalışmada belirtilen veri seti Kaggle'daki IBM HR Analytics Employee Attrition and Performance veri setidir.
Agarwal, Bhardwaj, Gatkamani, Gururaj, Darapaneni, ve Paduri (2023)	Sınıflandırma	Doğruluk oranı, F1 değeri, Eğrinin altında kalan alan	IBM HR Analytics Employee Attrition and Performance veri seti kullanılmıştır.

Çoğu çalışma “IBM HR Analytics Employee Attrition and Performance” veri setini kullanmıştır. Bu veri seti, Kaggle'da halka açık olarak bulunan ve yaygın olarak kullanılan bir veri setidir. Özel veri kaynakları kullanılan çalışmalarda ise şirket ismi genellikle belirtilmemiştir.

Atef vd. (2022) tarafından regresyon tabanlı yaklaşım ile personelin işten ayrılma tahmini ile ilgili yapılan çalışmada, ortalama hata karelerinin kökü değerini, destek vektör regresyon yöntemi ile 0,671; karar ağacı regresyon yönteminde 0,624; RO yöntemiyle 0,614 bulmuştur. Bu üç teknik arasında en başarılı sonucu RO tekniği vermiştir. Chen vd. (2023) ise personelin işten ayrılma tahminlemesi problemi için birden fazla makine öğrenmesi yöntemi kullanmış ve performans metriklerini incelemiştir. Çalışmanın sonucunda RO tekniğinin doğruluk oranı 0,985; Adaboosting tekniğinkine ise 0,984 olarak bulunmuştur. F1 değerleri ise 0,982 olarak aynı bulunmuştur. Bu da sınıflandırma tabanlı işten ayrılma tahminini her iki yöntemin de başarılı şekilde yapabileceğini göstermiştir.

Alamsyah ve Salma (2018), personelin işten ayrılma tahmini üzerine yöntemleri karşılaştıran bir çalışmada, en iyi sınıflandırma modelinin %97,5 doğruluk oranı ile RO olduğunu göstermiştir. İkinci en iyi yöntem, %96,6 doğruluk oranı ile Naïve Bayes'tir ve en düşük doğruluk oranına sahip sınıflandırma modeli ise %88,7 ile karar ağacı olarak çıkmıştır. Çalışma bu makine öğrenmesi teknikleri içerisinde, çalışan devrini tahmin etmek için en güvenilir ve doğru sınıflandırma modelinin RO algoritması olduğunu tespit etmiştir.

Literatürde sınıflandırma problemlerinde genellikle doğruluk, F1 skoru ve duyarlılık gibi performans metriklerinin öne çıktığını görülmektedir. Regresyon problemlerinde ise, Ortalama Mutlak Hata (MAE), Hata Karelerinin Ortalamasının Karekökü (RMSE) ve gerektiğinde R-kare gibi değerlerin kullanıldığı görülmektedir.

Personel ayrılma tahmini üzerine yapılan ampirik çalışmalar, insan kaynakları yönetimi açısından kritik bir rol oynamaktadır. Bu çalışmalarda genellikle makine öğrenmesi algoritmaları kullanılarak, çalışanların işten ayrılma riskini tahmin etmeye yönelik modeller oluşturulmaktadır. Araştırmacılar, IBM HR Analytics gibi yaygın veri setleri üzerinde çalışarak çalışanların yaş, kıdem, maaş, departman, iş memnuniyeti ve eğitim gibi faktörleri analiz ederek ayrılma olasılıklarını öngörmeye çalışmışlardır.

Personelin işten ayrılmasını etkileyen parametreler açısından literatür araştırması yapılmış olup işletme bakış açısıyla işten ayrılma kararı ile ilişkisi incelenen öznitelikler **EKLER** başlığı altında gösterilmiştir.

### Bölüm 3

## Makine Öğrenmesi Teknikleri

Yapay zekâ alanının çatısı altında yer alan makine öğrenmesi kavramı, insan beyninin hesaplayamayacağı kadar karmaşık model ve süreçler oluşturarak analitik çıktılar üretmeyi sağlayan bir teknikler topluluğudur. Karar alma sürecinde geçmiş verileri kullanarak tahmin sonuçları üretilir.

Makine öğrenmesi, bilgisayarların ve algoritmaların, belirli verilerden bağımsız bir şekilde öğrenmesine ve çıkarımlar yapmasına olanak tanıyan bir yapay zekâ dalıdır. Bu alan, algoritmaların veri setlerinden öğrenerek tahmin yapabilmesini veya karar alabilmesini sağlamak amacıyla geliştirilmiştir (Alpaydin, 2020). Makine öğrenmesi, birçok uygulamada insan müdahalesini en aza indirir ve algoritmaların "deneyim" yoluyla öğrenmesine olanak tanır, bu da büyük veri işleme kapasitesini artırır (Russell ve Norvig, 2016).

Makine öğrenmesinin temel prensibi, geçmiş verilerden öğrenmek ve bu bilgiyi gelecekteki verilere uygulamaktır. Bu öğrenme süreci, veri setlerinin toplanması, veri temizleme, özellik mühendisliği, model seçimi ve modelin eğitilmesi gibi adımları içerir. Özellikle büyük veri setlerinde çalışırken, makine öğrenmesi, verileri analiz ederek ve çıkarımlar yaparak insanların manuel olarak elde edemeyeceği sonuçlara ulaşır.

Makine öğrenmesi, üç ana kategoriye ayrılır: gözetimli öğrenme, gözetimsiz öğrenme ve pekiştirmeli öğrenme (Bishop, 2006). Gözetimli öğrenme, algoritmanın bir hedef değişkenin tahmininin yapılması amacıyla etiketlenmiş veri setleri üzerinde çalışır. Örneğin, satış verilerini analiz ederek bir ürünün gelecekteki satışlarını tahmin etmek gibi. Gözetimsiz öğrenme ise etiketlenmemiş veriler üzerinde uygulanır ve genellikle veri kümelerindeki gizli kalıpları veya kümeleri bulmak için kullanılır.

Makine öğrenmesi, finans, sağlık, eğitim ve daha birçok sektörde geniş bir uygulama alanına sahiptir. Örneğin, sağlık sektöründe, algoritmalar hastalıkların erken teşhisinde kullanılmakta ve böylece hastaların tedavi süreçlerinde büyük faydalar sağlanmaktadır (Obermeyer ve Emanuel, 2016). Eğitimde, öğrencilerin öğrenme biçimlerine göre özelleştirilmiş içerik

sağlanması yoluyla, öğrenme sürecini daha etkili hale getirmekte kullanılmaktadır (Baker ve Inventado, 2014).

Makine öğrenmesi, geleneksel algoritmalara göre daha hızlı ve doğru tahminler sağlar. Bu doğruluk ve hız, özellikle büyük veri setleri üzerinde çalışıldığında belirgin hale gelir (Jordan ve Mitchell, 2015). Örneğin, finans sektöründe hisse senedi fiyatlarının tahmininde veya organize suistimal tespitinde etkili bir performans gösterir. Bu gibi alanlarda, algoritmalar büyük veri kümelerinde hızlı ve etkili analizler yaparak işletmelerin stratejik kararlarını iyileştirmesine yardımcı olur (Brynjolfsson ve McAfee, 2014).

Makine öğrenmesinin temel avantajlarından biri de sürekli gelişme yeteneğidir. Algoritmalar yeni verilerle beslendikçe daha hassas hale gelir ve tahminleri sürekli olarak iyileştirilir (Murphy, 2012). Bu özellik, makine öğrenmesini özellikle dinamik ve hızlı değişen ortamlarda güçlü bir araç haline getirir. Örneğin, e-ticaret siteleri, müşterilerin satın alma geçmişine dayanarak öneriler sunar; böylece müşteri deneyimini ve memnuniyetini artırır (Ricci, Rokach ve Shapira, 2015). Makine öğrenmesi, aynı zamanda tahmin doğruluğunun artırılması, maliyet tasarrufu sağlanması ve insan hatalarının azaltılması gibi faydalar da sunar (Domingos, 2012). Örneğin, otomotiv endüstrisinde kullanılan makine öğrenimi tabanlı algoritmalar, sürücülerin ve yolcuların güvenliğini sağlamak için çarpışma tahminlerinde kullanılmaktadır (Kumar vd., 2023). Ancak, makine öğrenmesi bazı zorluklarla karşı karşıyadır. Verilerin güvenilir ve eksiksiz olması, modelin doğruluğunu doğrudan etkiler. Yetersiz veri veya veri eksiklikleri durumunda modelin performansı düşebilir (Russell ve Norvig, 2016). Ek olarak, bazı durumlarda algoritmalar "aşırı uyum" sorunuyla karşı karşıya kalabilir; bu da algoritmanın eğitim verisine aşırı derecede uyum sağlaması anlamına gelir ve genelleme kapasitesini azaltır.

Sonuç olarak, makine öğrenmesi, veri analizi ve tahmin süreçlerini dönüştüren güçlü bir araçtır. Bu teknoloji, verimliliği artırmak, iş süreçlerini iyileştirmek ve yeni bilgi keşifleri yapmak için kullanılmaktadır. Gelişen teknolojiyle birlikte, makine öğrenmesinin uygulama alanlarının daha da genişlemesi ve yeni alanlara entegre edilmesi beklenmektedir (Alpaydin, 2020).

### 3.1 Gözetimli Öğrenme Teknikleri

Bu teknikler, modelin önceden etiketlenmiş verilerden öğrenmesini sağlar. Model, belirli giriş-çıkış ilişkilerini öğrenerek gelecekteki veriler için tahmin yapar.

**3.1.1 Regresyon teknikleri.** İki ya da daha çok nicel değişken arasındaki ilişkiyi ölçmek için kullanılan analiz metodudur. Eğer tek bir değişken kullanılarak analiz yapılıyorsa buna tek değişkenli regresyon, birden çok değişken kullanılıyorsa çok değişkenli regresyon analizi olarak adlandırılır.

**3.1.1.1 Doğrusal regresyon.** Doğrusal regresyon, istatistik ve makine öğrenmesinde yaygın olarak kullanılan bir modelleme tekniğidir. Temel olarak, iki veya daha fazla değişken arasındaki ilişkileri inceleyerek, bağımlı bir değişkenin (y) bağımsız değişkenlere (x) bağlı olarak nasıl değiştiğini anlamaya çalışır. Doğrusal regresyon, verileri en iyi şekilde temsil eden bir doğru oluşturmayı hedefler ve bu doğruya "regresyon doğrusu" denir (Montgomery, Peck, ve Vining, 2012). Basit doğrusal regresyon modeli Denklem 1'de gösterilmektedir:

$$\hat{y} = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

Basit doğrusal regresyon denklemi düz bir çizgi olarak grafiklendirilir, bu grafikte;

- $\beta_0$ , regresyon doğrusunun y-kesişimidir.
- $\beta_1$ , eğimdir.
- $\hat{y}$ , belirli bir  $x$  değeri için  $y$ 'nin ortalama veya beklenen değeridir.

Rastgele hata ( $\varepsilon$ ), belirli bir  $x$  değeri için  $y$ 'nin gözlenen değeri ile  $y$ 'nin ortalama değeri arasındaki farkı temsil eder. Bu terim, hata terimini temsil eder ve modelin doğruluğunu etkileyen faktörleri içerir (Kutner, Nachtsheim, Neter ve Li, 2005).

Regresyon katsayısı  $\beta_1$ , bağımsız değişkendeki bir birimlik değişimin bağımlı değişkende yaratacağı beklenen değişimi temsil eder. Örneğin,  $\beta_1$  değeri pozitif ise, bağımsız değişken arttıkça bağımlı değişkenin değerinde de artış olur; negatif ise, bağımsız değişken arttıkça bağımlı değişkenin değeri azalacaktır (Draper ve Smith, 1998). Bu ilişkilerin doğruluğunu

değerlendirmek için çeşitli istatistiksel testler ve ölçümler kullanılır.

Doğrusal regresyonun temel varsayımları arasında doğrusal ilişki, bağımsız değişkenlerin normal dağılıma uygunluğu, hata terimlerinin bağımsız ve aynı dağılıma sahip olması yer alır. Bu varsayımlar, modelin güvenilirliği için kritik öneme sahiptir. Varsayımların ihlal edilmesi, modelin tahmin gücünü olumsuz etkileyebilir (Field, 2013).

Doğrusal regresyon, birçok alanda uygulama bulmuştur; ekonomi, sağlık, mühendislik ve sosyal bilimler gibi alanlarda veri analizi ve tahminleme için sıklıkla tercih edilegelmiştir. Örneğin, bir ekonomik modelde gelir ve tüketim arasındaki ilişkiyi incelemek için sıklıkla kullanılmıştır (Maddala, 2001).

Sonuç olarak, doğrusal regresyon, iki veya daha fazla değişken arasındaki ilişkiyi anlamak için güçlü bir araçtır. Modelin doğru bir şekilde kurulması, varsayımlarının kontrol edilmesi ve sonuçların dikkatlice yorumlanması, etkili tahminler yapmak ve stratejik kararlar almak için önemlidir (Cohen, Cohen, West ve Aiken, 2003).

Doğrusal regresyon yönteminin avantajları aşağıda özetlenmektedir.

- *Kolay Anlaşılabilirlik:* Modelin çıktısının yorumlanması kolaydır. Regresyon katsayıları, değişkenler arasındaki ilişkilerin net bir şekilde anlaşılmasına yardımcı olur.
- *Basit Uygulama:* Doğrusal regresyon, kullanıcıların karmaşık matematiksel modelleme becerilerine ihtiyaç duymadan verileri analiz edebilmelerini sağlar (Montgomery vd., 2012). Kullanıcı dostu istatistik yazılımları ile kısa sürede sonuç elde edilebilir.
- *Düşük Hesaplama Maliyeti:* Doğrusal regresyon, diğer daha karmaşık makine öğrenmesi algoritmalarına göre daha düşük hesaplama maliyetine sahiptir. Bu, büyük veri setleri üzerinde hızlı bir şekilde analiz yapabilme yeteneği sağlar (Field, 2013). Özellikle çok büyük veri kümeleriyle çalışırken, bu avantaj önemli bir faktördür.
- *Doğrusal İlişkileri Modelleme:* Doğrusal regresyon, bağımlı ve bağımsız değişkenler arasındaki doğrusal ilişkileri etkili bir şekilde modelleme yeteneğine sahiptir. Bu, birçok gerçek dünyadaki

durumun doğasına uygun bir yaklaşım sunar (Draper ve Smith, 1998). Ekonomiden sağlık sektörüne kadar birçok alanda uygulanabilir.

- *Model Değerlendirme:* Doğrusal regresyon, modelin başarısını değerlendirmek için çeşitli istatistiksel testler ve kriterler sunar. R-kare, AIC ve BIC gibi ölçümler, modelin iyi uyumunu değerlendirerek daha iyi tahminler yapmak için kullanılır (Greene, 2012). Bu, araştırmacılara ve analistlere modelin performansını artırma fırsatı sunar.

Doğrusal regresyon yönteminin dezavantajlarını aşağıdaki şekilde özetlemek mümkündür.

- *Doğrusal Olmayan İlişkilerde Başarısızlık:* Veriler arasındaki ilişkilerin doğrusal olduğu varsayılır. Bu varsayım, bazı durumlarda geçerli olmayabilir ve sonuçlar yanıltıcı hale gelebilir. Gerçek dünyada, birçok ilişki karmaşık ve doğrusal olmayan bir yapı gösterebilir (Cohen vd., 2003).
- *Aşırı Uyum Riski:* Model, veriye aşırı uyum sağladığında, gelecekteki tahminlerde başarısız olabilir. Bu, modelin karmaşıklığını artırarak veri setindeki gürültüyü öğrenmesine neden olur. Aşırı uyum, özellikle sınırlı veri ile çalışırken dikkat edilmesi gereken bir durumdur.
- *Heteroscedastisite Sorunu:* Doğrusal regresyon, hata terimlerinin sabit bir varyansa sahip olduğunu varsayar. Ancak, gerçekte, hata terimlerinin varyansı değişkenlik gösterebilir; bu durum, modelin güvenilirliğini azaltır (White, 1980). Bu tür durumlarda, modelin tahmin gücü olumsuz etkilenir.
- *Çoklu Doğrusal Bağlantı Sorunu:* Bağımsız değişkenler arasında yüksek bir ilişki varsa, çoklu doğrusal bağlantı problemi ortaya çıkar. Bu durum, regresyon katsayılarının güvenilirliğini azaltır ve tahminlerin doğru olmasını engeller (Mason ve Lind, 1999). Modelin stabilitesi açısından bu sorun kritik bir öneme sahiptir.
- *Dışsallık ve Zayıf Bağlantılar:* Doğrusal regresyon, ilişkilerin dışsal faktörlere veya değişkenler arasındaki zayıf ilişkilere

dayandığı durumlarda etkisiz olabilir. Model, tüm olası değişkenleri dikkate almazsa, yanıltıcı sonuçlar verebilir (Wooldridge, 2010). Bu nedenle, doğru değişkenlerin seçimi, doğrusal regresyonun başarısı için hayati önem taşır.

Sonuç olarak, doğrusal regresyon, güçlü bir modelleme aracı olmasına rağmen, bazı sınırlamaları ve riskleri de beraberinde getirir. Uygulayıcıların, bu avantaj ve dezavantajları dikkatlice değerlendirerek doğru kararlar alması önemlidir. Doğrusal regresyon, verilerin doğasına ve modelin kurulumuna bağlı olarak doğru bir analiz yapma yeteneğine sahiptir.

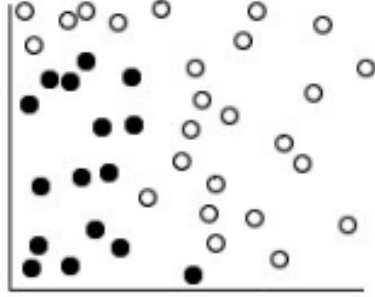
**3.1.1.2 Lojistik regresyon.** Lojistik regresyon, özellikle bağımlı değişkenin kategorik olduğu durumlarda yaygın olarak kullanılan bir istatistiksel modeldir. Bu model, bağımsız değişkenler ile bağımlı değişken arasındaki ilişkiyi tahmin etmek için kullanılır ve doğrusal regresyondan farklı olarak, bağımlı değişkenin kategorik sonuçlara sahip olmasını dikkate alır. Lojistik regresyon, olasılıkların logaritmik dönüşümünü kullanarak ikili sonuçları tahmin eder ve lojit fonksiyonuyla bağımlı değişkenin belirli bir kategoriye ait olma olasılığını hesaplar (Hosmer, Lemeshow ve Sturdivant, 2013). Bu sayede, hastalık tanısı, müşteri tercihleri gibi ikili sonuçlara sahip durumlarda yaygın olarak uygulanmaktadır.

Özellikle sosyal bilimler ve tıp alanlarında sıkça kullanılan lojistik regresyon, modelin esnekliği ve kategorik verilerle çalışabilme özelliği nedeniyle tercih edilmektedir. Bunun yanı sıra, bağımsız değişkenlerin lojistik regresyon üzerindeki etkisi, olasılık oranları kullanılarak yorumlanabilir. Odds oranları, bağımsız değişkenin bir birimlik artışının, bağımlı değişkenin belirli bir duruma ait olma olasılığını ne kadar etkilediğini gösterir (Menard, 2002). Böylelikle, lojistik regresyon modelleri karmaşık veri setlerinde bile anlamlı sonuçlar sağlayarak veri analizi süreçlerinde büyük bir katkı sunmaktadır.

**3.1.1.3 Ridge ve lasso regresyonları.** Ridge ve Lasso regresyonu, doğrusal regresyon modelinin genişletilmiş versiyonları olup, özellikle yüksek boyutlu verilerle çalışırken ve aşırı uyum (overfitting) riskini azaltmak için kullanılan düzenleme (regularization) yöntemleridir. Her iki yöntem de modelin karmaşıklığını sınırlandırarak daha güvenilir ve genellenebilir tahminler elde etmeyi amaçlar. Ancak, bunu yapma şekilleri

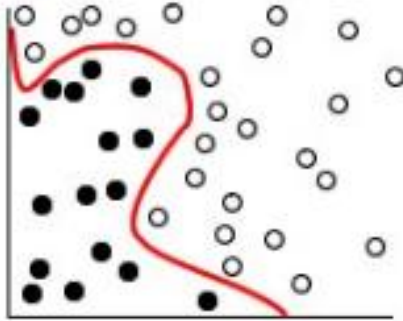
farklıdır.

**3.1.2 Destek vektör makineleri.** Destek Vektör Makineleri (Support Vector Machine – SVM), sınıflandırma ve regresyon analizi için kullanılan güçlü bir denetimli öğrenme yöntemidir. 1990’ların başında Vladimir Vapnik ve ekibi tarafından geliştirilen SVM, özellikle yüksek boyutlu verilerde etkili olmasıyla tanınmaktadır. Bu yöntem, verilerin doğru bir şekilde sınıflandırılmasını sağlamak için bir hiperdüzlem (hyperplane) kullanır. SVM'nin temel amacı, sınıflar arasındaki en büyük marjı maksimize ederek sınıflandırma hatasını minimize etmektir.



Şekil 1. İki farklı kategoride olan verilerin düzleme yerleştirilmiş gösterimi.

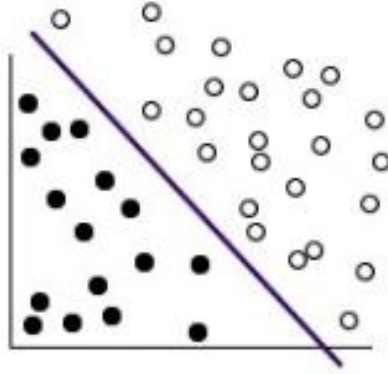
Şekil 1’de gösterildiği gibi, iki farklı sınıfı temsil eden verilerin iki boyutta yukarıdaki gibi gösterildiğini kabul edelim. Bu iki farklı kategorideki veriler Şekil 2’deki gibi eğri ile ayrılabilir.



Şekil 2. İki farklı kategorideki verileri ayıran eğrinin gösterimi.

Şekil 3’te bu ayırım net bir şekilde görülebilmektedir. En yaygın

kullanılan çekirdek fonksiyonları arasında lineer, polinom ve RBF (Radial Basis Function) çekirdekleri yer almaktadır.



Şekil 3. Dönüşüm sonrası verilerin birbirinden ayrılabilmesi.

**3.1.3 Karar ağaçları.** Karar ağaçları, veri madenciliği ve makine öğrenimi alanında yaygın olarak kullanılan sınıflandırma tekniklerindedir. Bu yöntem, verilerin özelliklerine dayanarak karar verme süreçlerini modellemek için hiyerarşik bir yapı kullanır. Karar ağaçları, veri analizi sırasında karmaşık ilişkileri basit ve görsel bir şekilde temsil etme yetenekleri ile dikkat çekmektedir.

Karar ağacının temel bileşenleri, düğümler, dallar ve yapraklardır. Düğümler, verilerin özelliklerini temsil ederken; dallar bu özelliklere göre verilen kararları gösterir. Yapraklar ise, sınıflandırma sonuçlarını veya regresyon tahminlerini temsil etmektedir. Bu yapı, karar alma süreçlerini açık bir şekilde görselleştirerek kullanıcıların modelin nasıl çalıştığını anlamalarına yardımcı olmaktadır.

Karar ağaçları, bölme (splitting) adı verilen bir süreçle çalışmaktadır. Veri seti, her bir düğümden en iyi özellik üzerinden iki veya daha fazla alt gruba ayrılmaktadır. Bu ayırım, genellikle bilgi kazancı (information gain), Gini indeksi veya varyans azaltma gibi kriterler kullanılarak yapılmaktadır. Bu yöntemler, hangi özelliğin veriyi daha iyi ayıracağını belirlemek için kullanılır. Her bir bölme, daha iyi bir tahmin veya sınıflandırma sağlamayı hedeflemektedir.

Ağaç algoritmalarında, önce kök düğüm, sonrasında ise ilk hangi nitelikten bölüneceği ve bölünme kurallarının neler olacağını tespit

edilmesini gerektirmektedir. En iyi ayırım noktalarının tespit edilmesi için ise gini indeks yöntemi kullanılabilir. Gini indeks değeri, bir sınıfın içindeki izafi sıklığı ifade eder. Her bir özellik için ayrı ayrı hesaplanan Gini indeks değerleri arasından en küçük olanı seçilir. Bölünme işlemi bu değere göre yapılmaktadır. Bu işlemler kalan veriler için de tekrar edilir ve diğer bölünmeler için hesaplamalar yapılır. Sonuç olarak, bir özneliğin gini indeks değeri ne kadar düşük ise önemi ve dolayısıyla ağırlığı o kadar yüksek olacaktır. Gini hesaplaması Denklem 2'deki şekilde formüle edilmiştir. Sklearn kütüphanesinin içerisinde yer alan Feature Importance fonksiyonu, özellik önemlerini model üzerinde otomatik olarak hesaplayabilmektedir.

$$\text{Gini} = 1 - \left( \sum_{j=0}^m p(j)^2 \right) \quad (2)$$

Burada  $p(j)$ , sınıf  $j$ 'nin gerçekleşme olasılığıdır ve toplam nüfus içerisindeki oranını temsil eder.  $m$  ise, toplam sınıf sayısıdır. Her sınıf için hesaplama yapılır ve çıkan sonuçların karelerinin toplamı birden çıkartılır. Gini değeri 0 ile 1 arasında bir sonuç alır ve sonuç 0'a ne kadar yakınsa o kadar iyi ayırım yapmış olur. Örneğin, sarı ve kırmızı toplam 4 bilyenin olduğunu kabul edelim.

Eğer 2 adet sarı ve 2 adet kırmızı bilye varsa:  $\text{Gini} = 1 - ((2/4)^2) + (2/4)^2 = 0.5$  olur.

Eğer 1 adet sarı ve 3 adet kırmızı bilye varsa:  $\text{Gini} = 1 - ((1/4)^2 + (3/4)^2) = 0.375$  olur.

Eğer 0 adet sarı, 4 adet kırmızı bilye varsa:  $\text{Gini} = 1 - ((0/4)^2 + (4/4)^2) = 0$  olur.

Görüleceği üzere, 4 bilyenin eğer 4'ü de kırmızı ise ağacın dalları çok net ayırım yapabilecektir. Bu da aynı zamanda saflılığın da bir göstergesidir.

Karar ağaçları hem kategorik hem de sürekli verilerle çalışabilme yeteneğine sahiptir. Kategorik veriler, belirli sınıflara ayrılan verileri temsil ederken, sürekli veriler bir dizi değer alabilen verilerdir. Bu esneklik, karar ağaçlarını farklı veri türleri üzerinde kullanılabilir hale getirir ve çok çeşitli uygulamalara olanak tanımaktadır.

Son yıllarda, karar ağaçları, topluluk (ensemble) yöntemleri ile bir

araya getirilerek daha güçlü modeller oluşturulmaktadır. Rastgele Orman ve Gradyan Artırma gibi yöntemler, birden fazla karar ağacını bir araya getirerek daha iyi tahminler sağlamaktadır. Bu teknikler, karar ağaçlarının her birinin avantajlarından faydalanırken, bireysel ağaçların zayıf yönlerini azaltmayı hedeflemektedir.

Chen ve Chen (2023), üniversitelerde finansal bilgi yönetimini iyileştirmek için karar ağacı algoritmalarını uygulamışlardır. Veri ön işleme, özellik seçimi ve model değerlendirmesi adımlarını içeren bu çalışma, karar ağaçlarının finansal yönetimde kullanımına bir örnek teşkil eder. Handayani ve ark. (2023), Jakarta Üniversitesi'nde öğrencilerin mezuniyet durumlarını sınıflandırmak için karar ağaçları ve topluluk yöntemlerini karşılaştırmışlardır. Topluluk yöntemleri, karar ağaçlarına kıyasla % 2-5 daha yüksek doğruluk sağlamıştır. Yapar ve ark. (2023), SEER veritabanından elde edilen 6234 hasta verisiyle primer malign kemik tümörlerinde sağkalım analizini gerçekleştirmişlerdir. Karar ağacı modeli, evre, yaş ve tümör derecesi gibi faktörleri belirleyerek bireysel hasta risk tahmininde etkili sonuçlar üretmiştir. Jaehn ve ark. (2023), yaşlı bireylerin günlük yaşam aktivitelerinde yardım ihtiyacını belirlemek için karar ağacı analizini kullanmışlardır. Bu yöntem ile, sosyal bakım hizmetlerinin daha etkili planlanmasına olumlu katkılar sağlanabileceği sonucuna ulaşmışlardır.

**3.1.3.1 Rastgele orman (RO) yöntemi.** RO yöntemi, karar ağaçlarına dayalı bir topluluk (ensemble) yöntemidir ve makine öğreniminde yaygın olarak kullanılmaktadır. Bu yöntem, birden fazla karar ağacından oluşan bir topluluk kurarak ve her ağacın tahmin sonuçlarını birleştirerek tahminlerde bulunmaktadır. Yani, hataları azaltmaya ve modelin genel doğruluğunu artırmaya odaklanır. Hem sınıflandırma hem de regresyon problemlerinde etkili sonuçlar veren rastgele orman algoritması, çeşitli veri tiplerinde başarıyla uygulanabilmektedir.

RO, çeşitli karar ağaçlarından oluşan bir yapıdadır. Bu yapı, torbalama (bagging) adı verilen bir teknikte birçok karar ağacının aynı veri seti üzerinden farklı alt kümelerle eğitilmesiyle elde edilmektedir. Rastgele orman modeli bir problemi çözmek için hem veri setinden hem de öznitelik setinden rastgele 10'larca ya da 100'lerce farklı alt-setler seçer ve bunları eğitir (Şimşek, 2022). Torbalama yöntemi, her ağacın farklı özellikler

üzerinde eğitilmesini sağlayarak modelin aşırı uyum gösterme ihtimalini azaltmaktadır. Ayrıca, bu yöntem veri setindeki gürültüyü dengelemek için oldukça etkilidir.

RO yönteminde kullanılan her bir karar ağacı, orijinal veri setinin farklı alt kümelerine göre eğitilmektedir. Bu alt kümeler, önyükleme örnekleme (bootstrap sampling) adı verilen bir teknikle rastgele seçilmektedir. Böylece her ağaç, veri setinin farklı bölümlerinden elde edilen örneklerle eğitilmiş olur. Karar ağacı algoritmaları hızlı sonuç vermelerine rağmen aşırı uyum problemi yaşayabilir; RO yöntemi ise bu durumu en aza indirmektedir.

Torbalama tekniği, veri setinin farklı alt kümelerini kullanarak modellerin eğitiminde çeşitlilik yaratmaktadır. RO'da, her bir karar ağacı birbirinden bağımsız olarak eğitildiği için ağaçlar arasındaki korelasyon azalmaktadır. Bu, modelin genel doğruluğunu artırarak verideki sapmayı ve varyansı dengelemede fayda sağlamaktadır. RO yöntemi, aşırı uyuma karşı oldukça dayanıklıdır. Torbalama ve özniteliklerin rastgele seçimi, modelin her ağaçta farklı özellikleri değerlendirmesini sağlayarak aşırı uyum riskini azaltmaktadır. Bu, özellikle yüksek boyutlu veri setlerinde sıkça karşılaşılan bir problemdir ve RO yöntemi sayesinde dengelenmektedir.

Her bir ağacın eğitimi sırasında kullanılan özellikler de rastgele seçilmektedir. Bu, ağacın her düğümünde belirli bir sayıdaki özelliğin içinden en iyi olanının seçilmesiyle sağlanmaktadır. Bu özellik rastgeleliği, modelin çeşitliliğini artırır ve aynı zamanda genelleme kabiliyetini yükseltmektedir.

RO, sınıflandırma ve regresyon problemlerinin her ikisini de çözebilmektedir. Sınıflandırma için, her bir ağacın sınıf tahmini alınır ve çoğunluk oyu prensibiyle nihai sınıf belirlenmektedir. Regresyon durumunda ise, her ağacın verdiği tahminlerin ortalaması alınarak son tahmin elde edilir. Her iki durumda da modelin doğruluğunun ve genelleme kabiliyetinin oldukça yüksek olduğu görülmektedir.

RO, her bir değişkenin model üzerindeki etkisini değerlendirme imkânı sağlamaktadır. Özellik önem skoru (OOB) adı verilen bu değerlendirme, modelde hangi özelliklerin daha fazla etkili olduğunu göstermektedir. Böylece, değişken seçimi yapılabilir ve model karmaşıklığı optimize edilebilir bir mahiyet kazanmaktadır.

RO yönteminde, her ağaç farklı bir veri alt kümesiyle eğitildiği için kalan veriler, model doğrulama amacıyla kullanılabilir. Bu, OOB hatası olarak bilinir ve modelin doğruluğunu hesaplamada oldukça yararlıdır. OOB hatası, genellikle doğrulama veri seti olmadan modelin doğruluğunu test etmede kullanılmaktadır.

RO'nun diğer avantajlarından biri de, yüksek doğruluğa sahip olmasıdır. Çeşitli veri kümeleri üzerinde yapılan farklı uygulamalarda RO yüksek performans göstermesiyle tanınmaktadır. Ayrıca, gürültülü veri setlerinde bile oldukça dayanıklıdır ve büyük veri setlerinde bile hızlıca uygulanabilmektedir. Bir başka avantajı ise, değişkenlerin önem sırasını verebilmesi ve veri analizine katkı sağlamasıdır.

Her ne kadar RO güçlü bir yöntem olsa da bazı dezavantajları da vardır. Özellikle çok sayıda karar ağacının oluşturulması, hafıza ve işlem gücü gerektirmektedir. Bu durum, büyük veri setleri üzerinde çalışıldığında daha fazla kaynak kullanımı anlamına gelmektedir. Ayrıca, modelin iç yapısı karmaşık olduğu için yorumlanabilirlik açısından zorluklar yaratabilmektedir.

RO algoritması, finans, biyoinformatik, sağlık ve pazarlama gibi birçok alanda kullanılmaktadır. Özellikle büyük veri setlerinde ve çoklu sınıflandırma gerektiren problemler için uygundur. Ayrıca, değişken önem sıralaması sunması, birçok sektörde veri analistlerine fayda sağlamaktadır. RO algoritmalarına alternatif olarak gradient boosting, XGBoost gibi diğer topluluk yöntemleri de kullanılabilir. Bu yöntemler de birden fazla modelin birleşiminden oluşur, ancak her bir modelin hata oranını minimize etmeye yönelik bir yaklaşımla çalışmaktadır. Bu nedenle, performans gereksinimleri ve veri yapısına göre yöntem seçimi yapmak önemlidir.

Martínez-Gramage ve ark. (2020), genç triatletlerde koşu yaralanmalarını azaltmak amacıyla rastgele orman tabanlı bir makine öğrenmesi çerçevesi geliştirmişlerdir. Bu model ile yaralanma riskini tahmin etmekte yüksek doğruluk sağlanmıştır. Rathore ve ark. (2021), RO algoritması ile kötü amaçlı Android uygulamalarını tespit etmişlerdir. Model, derin öğrenme ve diğer makine öğrenmesi algoritmalarına kıyasla daha yüksek AUC değerleri elde etmiştir. Başka bir çalışmada ise, Thomas ve Kaliraj (2024) RO algoritmasını kullanarak yazılım hatalarını tahmin

etmişlerdir. Model, NASA hata veriseti üzerinde uygulanmış ve yüksek doğruluk sağladığı görülmüştür. Ayrıca yazılım hatalarını tespit etmede bu yöntemin kullanılabileceği de gösterilmiştir. Nadkarni ve ark. (2023), RO ile hava folyo gürültüsünü tahmin etmişlerdir. Model, diğer makine öğrenmesi algoritmaları ile karşılaştırıldığında en yüksek determinasyon katsayısını vermiştir.

**3.1.3.2 Gradyan artırma yöntemi (Gradient Boosting).** Gradyan artırma yöntemi, makine öğrenimi alanında kullanılan güçlü ve popüler bir teknik olup, özellikle sınıflandırma ve regresyon problemlerinde başarılı sonuçlar elde etmek için kullanılmaktadır. Bu yöntemin temel amacı, bir dizi zayıf öğreniciyi art arda eğiterek bir güçlü öğrenici oluşturmaktır. Zayıf öğreniciler, genellikle hatalı tahminlerde bulunan basit modellerdir; ancak bu modeller art arda eğitilerek ve her bir modelin hatalarını gidermek için optimize edilerek güçlü ve başarılı bir tahmin modeli ortaya çıkarılmaktadır.

Gradyan artırma yöntemi, adım adım oluşturulan bir dizi modelle çalışmaktadır. İlk olarak, başlangıçta rastgele veya temel bir model oluşturulur ve veriler üzerinden tahmin yapılır. Bu ilk model, tahminlerde eksiklikler ve hatalar yapabilmektedir. Gradyan artırma yönteminin amacı, her bir adımda bu hataların üzerine odaklanarak ve yeni modeller ekleyerek hataları azaltmaktır. Bu, hatalar üzerinden öğrenme işlemi olarak adlandırılır ve her bir yeni model, bir önceki modelin hata yaptığı örneklerden daha iyi tahmin yapmayı öğrenebilmektedir.

Bu yöntem, aslında adından da anlaşılacağı üzere, gradyan azaltma algoritmasının bir varyasyonu olarak çalışır. Her bir adımda, modelin tahmin hatası üzerinden türev alınarak hangi yönde ilerlenmesi gerektiği belirlenmektedir. Bu türev bilgisi, modelin bir sonraki adıma hangi doğrultuda gitmesi gerektiğini gösterir ve böylece model adım adım hataları azaltacak şekilde güncellenmektedir. Gradyan artırma yönteminde kullanılan bu strateji, modelin hızlıca en iyi çözümü bulmasını sağlamaktadır.

Gradyan artırma algoritması genellikle ağaç tabanlı modellerle, özellikle karar ağaçları ile kullanılmaktadır. Bu durum, modelin doğrusal olmayan ilişkileri öğrenmesini sağlar ve veri setindeki karmaşık yapıları daha iyi anlamasına olanak tanımaktadır.

Karar ağaçları, verileri düğümlere ayırarak çalışmakta ve her bir

düğümde belirli bir sınırlama kriteri kullanılmaktadır. Gradyan artırmada, her yeni ağaç önceki ağaçların yaptığı hataları minimize etmeye odaklanarak eğitilir ve bu ağaçlar toplandığında güçlü bir model elde edilmektedir. Gradyan artırma yöntemi, birçok farklı veri türüyle uyumlu olarak çalışabilme avantajına sahiptir. Bunun en büyük nedeni, ağaç tabanlı yapısının kategori, sürekli ve karmaşık veri tiplerini işleyebilmesidir. Aynı zamanda, veri setindeki eksik verilerle de başa çıkabilir ve verilerin daha anlamlı hale getirilmesine katkı sağlamaktadır. Dolayısıyla, veri setlerinde eksik verilerin olduğu durumlarda veya karmaşık veri yapılarında gradyan artırma etkili sonuçlar sunmaktadır.

Bu yöntemin bir diğer önemli avantajı, aşırı öğrenme problemini azaltmasıdır. Gradyan artırma, adım adım hataları iyileştirerek çalıştığından, aşırı öğrenmeye yatkın bir model yapısından kaçınmaktadır. Özellikle hiper parametrelerin doğru ayarlanması durumunda, model veriyi daha iyi öğrenirken, aynı zamanda genelleme gücünü de artırmaktadır. Örneğin, öğrenme oranı adı verilen bir parametre, her adımda yapılan iyileştirmenin ne kadar büyük olacağını belirler ve bu sayede modelin dengeli bir şekilde öğrenmesi sağlanmaktadır. Bir diğer avantajı ise, büyük veri setlerinde yüksek doğruluğa ulaşabilmesidir. Büyük veri setlerinde modelin yüksek doğruluğa sahip olması kritik öneme sahiptir. Gradyan artırma yöntemi, bir dizi modeli optimize ederek eğittiği için veri setinin tamamını daha iyi öğrenir ve daha yüksek doğruluk elde eder. Bu nedenle, büyük veri setleri üzerinde sınıflandırma ve regresyon gibi görevlerde gradyan artırma sıklıkla tercih edilir.

Gradyan artırma algoritmasında dikkat edilmesi gereken en önemli noktalardan biri hiper parametre ayarlamasıdır. Öğrenme oranı, ağaç derinliği ve ağaç sayısı gibi parametreler, modelin başarısını büyük ölçüde etkiler. Bu parametrelerin yanlış seçilmesi durumunda, model ya aşırı öğrenme yapabilir ya da düşük performans gösterebilir. Dolayısıyla, gradyan artırma yönteminde hiper parametre optimizasyonu kritik bir adımdır ve model performansını maksimize etmek için dikkatle yapılmalıdır.

Gradyan artırma yöntemi, birden fazla alanda başarıyla uygulanabilmektedir. Örneğin, finansal tahminlerde, kredi risk analizi gibi durumlarda sıklıkla tercih edilmektedir. Bunun yanında, pazarlama

analitiğinde müşteri segmentasyonu, tıbbi teşhislerde hastalık sınıflandırma ve doğal dil işleme gibi farklı alanlarda da etkili sonuçlar sağlamaktadır. Bu alanlarda veri setlerindeki karmaşık ilişkileri öğrenebildiğinden, gradyan artırma yöntemi son yıllarda oldukça popüler hale gelmiştir.

XGBoost, LightGBM ve CatBoost gibi popüler gradyan artırma yöntemleri bulunmaktadır. XGBoost, en bilinen gradyan artırma algoritmalarından biri olup, hızlı çalışması ve yüksek performans göstermesi ile bilinmektedir. LightGBM, daha büyük veri setlerinde yüksek hızda çalışmasıyla dikkat çekerken, CatBoost ise özellikle kategorik verilerin fazla olduğu veri setlerinde etkili sonuçlar sunmaktadır. Bu algoritmalar, belirli durumlarda gradyan artırma yöntemini daha verimli hale getirmektedir.

Ying ve ark. (2022), çoklu görev öğrenimi için MT-GBM adlı bir gradyan artırma tabanlı yeni bir yöntem önermiştir. Bu yöntem, paylaşılan ağaç yapıları aracılığıyla görevler arası bilgi paylaşımını sağlamaktadır. Ayrıca gradyan artırma yönteminden daha yenilikçi bir yaklaşım ile öğrenme sağladığı gösterilmiştir. Başka bir çalışmada Jun (2021), Seul metropol alanında kentsel arazi kullanımı değişikliklerini modellemek için gradyan artırma, RO ve yapay sinir ağlarını karşılaştırmıştır. Gradyan artırma ve RO, yapay sinir ağlarına kıyasla daha yüksek tahmin gücü göstermiştir. Rizkallah (2025), regresyon problemlerinde gradyan artırma performansını artırmak için K-ortalamlar ve Bisecting K-ortalamlar kümeleme yöntemlerini entegre etmiştir. Bu hibrit yaklaşım, 40 veri seti üzerinde yapılan deneylerde tek başına gradyan artırma modeline kıyasla daha iyi sonuçlar üretmiştir. Yıldız ve Kalaycı (2025), tıbbi teşhislerde gradyan artırma algoritmalarının üç versiyonunun (XGBoost, CatBoost, LightGBM) performansını incelemiştir. Bu yöntemlerin, geleneksel makine öğrenimi ve derin öğrenme modellerine kıyasla daha yüksek doğruluk ve daha düşük hesaplama maliyeti ortaya çıkardığı deneyler sonucunda tespit edilmiştir.

**3.1.3.3. Ekstrem gradyan artırma yöntemi (XGBoosting).** Ekstrem Gradyan Artırma Yöntemi veya yaygın adıyla XGBoost (Extreme Gradient Boosting), makine öğrenimi alanında oldukça popüler olan, performansı yüksek bir gradyan artırma algoritmasıdır. XGBoost, veri biliminde sıklıkla kullanılan bir araç haline gelmiş olup, özellikle Kaggle gibi araştırma platformlarında oldukça çok tercih edilmektedir. Temel olarak, gradyan

artırma yönteminin geliştirilmiş bir versiyonudur ve yüksek doğruluk sağlamanın yanında hızlı çalışması ile öne çıkmaktadır. XGBoost, adından da anlaşılacağı gibi ekstrem seviyede performans sunmayı hedefleyen bir gradyan artırma yöntemidir.

XGBoost'un temel prensibi, klasik gradyan artırma yönteminin üzerine çeşitli optimizasyonlar ekleyerek daha verimli bir tahmin modeli oluşturmaktır. Bu optimizasyonlar arasında paralel işleme, bellek verimliliği, düzenleme teknikleri ve ağaç yapısının özelleştirilmesi gibi iyileştirmeler bulunmaktadır. Bu özellikler sayesinde XGBoost, diğer gradyan artırma algoritmalarından daha hızlı çalışır ve daha yüksek doğruluk oranları sunmaktadır. Dolayısıyla, büyük veri setlerinde ve karmaşık yapıları problemler için XGBoost oldukça uygundur.

XGBoost, tıpkı diğer gradyan artırma yöntemlerinde olduğu gibi, bir dizi zayıf öğreniciyi art arda ekleyerek güçlü bir model oluşturmaktadır. Bu zayıf öğreniciler genellikle karar ağaçlarıdır. Ancak XGBoost, her bir yeni ağaç eklendiğinde önceki ağaçların yaptığı hataları daha iyi bir şekilde optimize etmek için farklı stratejiler kullanmaktadır. Örneğin, her bir ağaç eğitim aşamasında mevcut modelin hata terimi üzerinde çalışmakta ve bu hataları minimize edecek yeni bir model eklemektedir. Bu süreç, modelin genel doğruluğunu artırır ve daha iyi tahminlerde bulunmasını sağlamaktadır.

XGBoost'un en önemli özelliklerinden biri, paralel işleme yeteneğidir. Paralel işleme sayesinde algoritma, karar ağaçlarını aynı anda birden fazla işlemci çekirdeğinde oluşturabilmektedir. Bu durum, özellikle büyük veri setleri üzerinde çalışırken hesaplama hızını ciddi şekilde artırmaktadır. Standart gradyan artırma yöntemleri, genellikle seri olarak çalışmakta olup bu nedenle XGBoost kadar hızlı değildirler. XGBoost'un paralel işleme kapasitesi, onu yüksek hızlı analiz gerektiren uygulamalarda öne çıkarmaktadır.

Bellek verimliliği, XGBoost'un öne çıkan bir diğer avantajlarından biridir. XGBoost, veri belleğini verimli bir şekilde kullanmak için optimizasyon teknikleri sunmaktadır. Bu, büyük veri setleri ile çalışırken algoritmanın hızlı çalışmasını sağlamaktadır. XGBoost, veri setlerini "in-place" işlemeye yönelik olarak tasarlanmış olup, bellekte gereksiz veri kopyalamalarını

azaltılmaktadır. Bu özellik, veri bilimcilerin düşük bellek alanında bile yüksek doğrulukla çalışmasına olanak tanımaktadır.

XGBoost'un sahip olduğu bir diğer gelişmiş özellik ise düzenleme yeteneğidir. Diğer gradyan artırma yöntemlerinde düzenleme sınırlı bir şekilde uygulanırken, XGBoost bu konuda daha esnekler. L1 ve L2 düzenleme terimleri, XGBoost'ta yerleşik olarak bulunur ve aşırı öğrenmeyi önlemek için kullanılır. Bu düzenleme teknikleri, modelin karmaşık veri yapılarını öğrenirken genelleme kabiliyetini artırmakta ve aşırı öğrenmeyi engellemektedir. Bu da XGBoost'u daha sağlam bir model haline getirmektedir.

XGBoost, eğitim sürecinde 'ağaç yapısı araması' ile de diğer gradyan artırma yöntemlerinden ayrılmaktadır. Standart gradyan artırma yöntemlerinde, yeni ağaçlar belirli bir yapı ile oluşturulmaktadır. Ancak XGBoost, farklı ağaç yapılarını dener ve en uygun yapıyı seçer. Bu, algoritmanın her bir adımda daha doğru tahminler yapmasını sağlamaktadır. Ağaç yapısı araması, modelin doğru karar ağaçlarını daha verimli bir şekilde oluşturmasına yardımcı olur ve tahmin performansını artırmaktadır.

Bir diğer özelliği ise ağırlıklı örnekleme (weighted quantile sketch) olarak bilinen tekniktir. XGBoost, veri setindeki ağırlıklı değerleri daha iyi işleyerek, modelin daha hassas tahminler yapmasını sağlamaktadır. Bu özellik, dengesiz veri setlerinde modelin doğru bir şekilde öğrenmesini sağlar. Örneğin, sınıflar arasında dengesizlik olan bir veri setinde XGBoost, dengesizliği gidermek için ağırlıklı örnekleme tekniğini kullanarak daha doğru sınıflandırmalar yapabilmektedir.

XGBoost'un önemli parametrelerinden biri olan öğrenme oranı, modelin her bir adımda yaptığı değişikliğin büyüklüğünü kontrol etmektedir. Öğrenme oranı genellikle küçük bir değerde ayarlanarak modelin yavaş ama doğru bir şekilde öğrenmesi sağlanmaktadır. Yüksek bir öğrenme oranı modelin hızlı öğrenmesine yol açabilir, ancak bu durumda aşırı öğrenme riski artmaktadır. Bu yüzden, XGBoost'ta öğrenme oranının dikkatlice ayarlanması gerekmektedir.

XGBoost, büyük veri setlerinde yüksek verimlilikle çalışabilmesi nedeniyle çok sayıda endüstriyel ve akademik uygulamada kullanılmaktadır. Finans sektöründe kredi riski analizi, pazarlama alanında müşteri

davranışlarını tahmin etme, tıbbi arařtırmalarda hastalık sınıflandırması gibi alanlarda etkili bir řekilde kullanılabilir. XGBoost, büyük veri setlerinde hızlı çalışması ve yüksek doğruluęu sayesinde bu alanlarda oldukça popülerdir. Dang ve ark. (2022), güneş lekesi sayısını tahmin etmek için XGBoost tabanlı bir derin öğrenme modeli (XGBoost-DL) geliřtirmiřtir. Model, dięer geleneksel ve derin öğrenme modellerine kıyasla daha düşük RMSE ve MAE deęerleri elde etmiřtir. Ma (2022), COVID-19 döneminde medikal sektör hisselerinin fiyatlarını tahmin etmek için XGBoost algoritmasını kullanmıřtır. Model, ARIMA ve LSTM modellerine kıyasla daha iyi genel performans sergilemiřtir. Engin ve Durer (2023) Borsa İstanbul BIST100 endeksindeki řirketlerin finansal sıkıntı yařayıp yařamayacaęını tahmin etmek için XGBoost algoritmasını kullanmıřtır. Inoue ve ark. (2020), servikal omurilik yaralanması olan hastalarda nörolojik iyileřmeyi tahmin etmek için XGBoost algoritmasını kullanmıřtır. Model, dięer geleneksel yöntemlere kıyasla daha yüksek doğruluk ve AUC deęerleri elde etmiřtir.

Hiper parametre optimizasyonu, XGBoost'ta başarı için kritik öneme sahiptir. XGBoost'ta ağaç sayısı, maksimum derinlik, öğrenme oranı, düzenleme parametreleri gibi birçok hiper parametre bulunur. Bu parametrelerin doğru ayarlanması modelin başarısını büyük ölçüde etkilemektedir. XGBoost, model performansını maksimize etmek için bu parametrelerin optimize edilmesi gereken esnek bir yapı sunmaktadır.

**3.1.3.4. CHAID yöntemi.** CHAID (Chi-squared Automatic Interaction Detector), verileri kategorilere ayırarak karar ağaçları oluřturan ve istatistiksel olarak anlamlı farklılıklar saęlayan bir yöntemdir. Özellikle sosyoloji, psikoloji, pazarlama gibi sosyal bilimlerde kullanılmaktadır. CHAID yöntemi, deęişkenler arasındaki iliřkileri inceleyerek veriyi daha anlamlı kategorilere böler ve bu sayede sınıflandırma ve tahmin amaçlı analizler yapabilmektedir. CHAID, veri setinde baęımlı deęişken ile baęımsız deęişkenler arasındaki iliřkiyi arařtırır ve veriyi belirli kurallara göre ayırarak bir karar ağacı oluřturur. Bu ayırım, baęımsız deęişkenlerin baęımlı deęişkeni en iyi řekilde açıklayabileceęi gruplar elde etmek amacıyla yapılmaktadır.

CHAID'in en büyük avantajı, baęımlı deęişkenin farklı baęımsız

değişken kategorileri üzerindeki etkilerini istatistiksel olarak anlamlı bir şekilde ortaya koymasındır. Ayrıca, büyük veri setlerinde oldukça hızlı sonuç vermektedir. CHAID yöntemi, yalnızca kategorik verilerle çalışabildiğinden, sürekli verileri analiz etmek için uygun değildir. Bu nedenle, sürekli verileri analiz etmek isteyen kullanıcılar önce bu verileri kategorik hale getirmek zorundadır.

Kuyumcu ve ark. (2024), Türkiye'deki trafik kazalarında sürücülerin yaralanma durumlarını analiz etmek için CHAID karar ağacı modelini kullanmıştır. Çalışma, sürücülerin yaşları, cinsiyetleri ve araç türleri gibi faktörlerin yaralanma durumları üzerindeki etkilerini belirlemiştir. Karakaya ve ark. (2018), lise öğrencilerinin sosyal medyaya yönelik tutumlarını etkileyen faktörleri belirlemek için CHAID analizini kullanmıştır. Çalışma, öğrencilerin hayal ettikleri mesleklerin sosyal medya tutumları üzerinde etkili değişken olduğunu ortaya koymuştur. Díaz-Pérez ve Bethencourt-Cejas (2017), İspanya'nın Tenerife adasındaki Teide Ulusal Parkı'na gelen ziyaretçilerin çevresel etkilerini incelemek için CHAID algoritmasını uygulamıştır. Çalışma, ziyaretçilerin ulaşım türleri ve parkta geçirdikleri sürelerin çevresel etkiyle ilişkili olduğunu ortaya koymuştur.

**3.1.3.5 CART yöntemi.** CART, veri setindeki bağımlı değişkenle ilişkili bağımsız değişkenleri belirler ve veriyi bu değişkenlere göre alt gruplara ayırarak bir karar ağacı oluşturmaktadır. Sınıflandırma için kullanıldığında veri kategorilere ayrılırken, regresyon amacıyla kullanıldığında sürekli değerler tahmin edilmeye çalışılmaktadır.

CART, veriyi en iyi şekilde ayıracak düğümü belirlemek için farklı bölme kriterleri kullanılmaktadır. Sınıflandırma problemlerinde genellikle Gini kirliliği veya entropi, regresyon problemlerinde ise en küçük kareler yöntemi kullanılır. Gini kirliliği, sınıflandırma problemlerinde kullanılan bir bölme ölçütüdür. Veriyi iki gruba ayırırken, her bir grubun homojenliğini artırmayı hedefler. Gini değeri düşükse, gruptaki veriler benzer özelliklere sahiptir. Hesaplama formülü Denklem 2'te verilmişti.

Entropi, bir grubun bilgi miktarını ölçen bir diğer sınıflandırma kriteridir. Daha homojen gruplar daha düşük entropiye sahiptir. CART, bazen entropi ölçütü kullanarak grupları ayırır, böylece bilgi kazancını en yüksek seviyede tutar. Entropi hesaplama formülü Denklem 3'te

gösterilmektedir.

$$H = -\sum p(x) \ln p(x) \quad (3)$$

Burada,  $p(x)$  belirli bir sınıfa ait grubun yüzdesini ve  $H$  ise entropiyi belirtmektedir. Karar ağacında entropi değerini en aza indirgeyen bölünmeler yapılması hedeflenmektedir. En iyi bölünmeyi belirlemek için de bilgi kazancı kullanılmaktadır. Bilgi kazancı en yüksek olan özellik en iyi olarak seçilmektedir. Hesaplama formülü, Denklem 4'te gösterilmektedir.  $p(i)$ ,  $i$  durumunun gerçekleşme olasılığıdır.  $n$  ise, durum sayısıdır.

$$G(s) = Entropi(s) - \sum_{i=1}^n (p(i) * Entropi(i)) \quad (4)$$

Klusowski (2020), CART algoritmasının istatistiksel özelliklerini inceleyerek, eğitim hatasının Pearson korelasyonu ile ilişkisini analiz etmiştir. Çalışma, model karmaşıklığı ve uyum arasındaki dengeyi ve ilişkiyi incelemiştir. Belli ve Vantini (2020), fonksiyonel veri analizinde kullanılmak üzere yeni bir CART tabanlı algoritma önermiştir. Bu yöntem, çoklu fonksiyonel girişleri işleyerek sınıflandırma ve regresyon görevlerini gerçekleştirebileceğini göstermiştir.

**3.1.4 Naive bayes.** Naive Bayes (NB), olasılık temelli bir sınıflandırma algoritmasıdır ve özellikle metin sınıflandırması gibi yüksek boyutlu verilerle çalışmada yaygın olarak kullanılmaktadır. Temeli Bayes Teoremi olup, sınıflar arasındaki bağımsızlık varsayımıyla çalışmaktadır. Naive Bayes, genellikle basit ve hızlı çalışır, aynı zamanda büyük veri kümelerinde yüksek performans sunmaktadır. Bu algoritmanın adı, özelliklerin bağımsız olduğu varsayımı nedeniyle "Naive" yani "Saf" ve Bayes Teoremi'ne dayandığı için "Bayes" olarak adlandırılmıştır.

Bayes Teoremi, bir olayın olasılığının öncül bilgilere dayanarak nasıl değiştirilebileceğini ifade etmektedir. Bayes Teoremi'nin matematiksel ifadesi Denklem 5'te gösterilmiştir.

$$P(A | B) = (P(B | A) \times P(A)) / P(B) \quad (5)$$

$P(A|B)$ , B olayı gerçekleştiğinde A olayının olasılığıdır. Naive Bayes yönteminde ise özellikler bağımsız olarak varsayılır ve bu nedenle olasılık hesabı basitleşmektedir.

Sedghpour ve Sedghpour (2020), web belgelerinin sınıflandırılmasında Naive Bayes algoritmasını Gizli Anlamsal Analiz (LSA) ile birleştirerek, sınıflandırma doğruluğunu artırmıştır. Bu yaklaşım, özellikle yüksek boyutlu ve seyrek veri setlerinde etkili olduğunu ortaya çıkarmıştır. Başka bir çalışmada ise, Bozuyula (2021) sosyal medyada sahte haberleri tespit etmek için Naive Bayes algoritmasını AdaBoost topluluk öğrenme yöntemiyle birleştirerek, doğruluk oranını %96,74'ten %98,2'ye yükseltmiştir. Bu yaklaşım, sahte haber tespiti için etkili bir yöntem sunmuştur.

**3.1.5 Bayes ağları.** Bayes Ağları olasılık temelli bir grafiksel modeldir ve değişkenler arasındaki olasılıksal ilişkileri modellemek için kullanılmaktadır. Yapay zekâ, makine öğrenmesi, veri madenciliği ve istatistik gibi birçok alanda tercih edilen Bayes ağları, olayların neden-sonuç ilişkilerini modelleyerek belirsizlikler altında karar verme süreçlerini desteklemektedir. Bayes ağları, düğümler ve yönlendirilmiş kenarlardan oluşan bir grafiktir ve düğümler rastgele değişkenleri temsil ederken, kenarlar ise bu değişkenler arasındaki nedensel bağımlılıkları ifade etmektedir.

Bayes ağlarının temel çalışma prensibi, bir önceki yöntemde de olduğu gibi Bayes Teoremi'ne dayanmaktadır. Bayes ağlarında her bir düğüm, bir olayın veya durumun gerçekleşme olasılığını içerir ve bu düğümlerin bağımlılık yapısı, sistemin genel olasılıksal modelini oluşturmaktadır. Her bir düğümün olasılığı, kendisine bağlı olan ebeveyn düğümlerin değerlerine göre belirlenmektedir. Böylece, bir olayın gerçekleşme olasılığı hesaplanırken, diğer olayların etkisi göz önünde bulundurulabilmektedir. Ordovas ve ark. (2021), İspanya'da toplanan büyük bir veri setini kullanarak kardiyovasküler risk tahmini için Bayes Ağı modelinin kullanılabileceğini uygulamalı olarak önermiştir.

### 3.2 Makine Öğrenmesinde Model Değerlendirme Ölçütleri

Makine öğrenmesi modeli oluşturulduktan sonra yapılan

tahminlemelerin doğru olup olmadığı, problemin sınıflandırma ya da regresyon problemi olmasına bağlı olarak bazı metrikler üzerinden değerlendirilmektedir. Bu metrikler modelin kararlı, doğru ve güvenilir olmasını test etmektedir.

**3.2.1 Regresyon modellerinin değerlendirme ölçütleri.** Regresyon tahmini yapılmak istenilen makine öğrenmesi modellerinde  $R^2$  skoru, ortalama mutlak hata, ortalama kare hatası ve düzeltilmiş R-Kare metriklerini kontrol etmek gerekmektedir.

$R^2$  Skoru, verilerin yerleştirilmiş regresyon hattına olan mesafesinin istatistiksel ölçüsüdür. Regresyon noktasına isabet edemeyen her noktaya artık adı verilmektedir. Bu artıkların her birinin karelerinin toplamına Artıkların Karelerinin Toplamı (AKT) ve her bir noktanın ortalamaya uzaklığının karesine de Ortalamaya Uzaklığın Kareler Toplamı (OUKT) adı verilir.  $R^2$  değeri, bu iki istatistikten hareketle Denklem 6'daki gibi hesaplanır.

$$R^2 = 1 - (AKT/OUKT) \quad (6)$$

$R^2$  değeri 1'e ne kadar yakın ise regresyon modelinin o kadar doğru olduğu anlaşılır. Bu metrikte dikkat edilmesi gereken husus  $R^2$  değeri eğer 1 çıkarsa modelin ezberlemiş olabileceğidir. 1 değerinden uzaklaşmasıyla modelin tahminleme gücünde azalma olduğu anlaşılmaktadır.

Düzeltilmiş R-Kare (Adjusted  $R^2$ ), modeldeki bağımsız değişken sayısına göre ayarlanmış bir  $R^2$  versiyonudur. Düzeltilmiş  $R^2$ , modele yeni bir değişken eklendiğinde artış gösterebilir. Bağımsız değişkenlerin sayısı arttıkça model daha karmaşık hale gelir, ancak bu da aşırı öğrenme (overfitting) gibi sorunlara yol açabilmektedir.

Ortalama Mutlak Hata (Mean Absolute Error), tahmin edilen değerler ile gerçek değerler arasındaki farktır. Veri setindeki tüm örnekler için, tahmin edilen değer ile gerçek değer arasındaki farkların mutlak değerlerinin ortalamasına bu isim verilir. Bu değerinin düşük olması, modelin daha az hata yaptığını ve dolayısıyla iyi tahminlerde bulunduğunu göstermektedir.

Ortalama Kare Hatası (Mean Squared Error), regresyon modellerinde en yaygın kullanılan hata ölçütlerinden biridir. Bu terim, tüm veri setindeki

tahmin hatalarının karelerinin ortalamasıdır. Bu değerin sıfıra yakın olması, modelin daha doğru tahminler yaptığını ifade etmektedir.

### 3.2.2 Sınıflandırma modellerinin değerlendirme ölçütleri.

Sınıflandırma tahmininde modelin doğruluğunu ve performansını ölçmek için birçok değerlendirme metriği kullanılmaktadır. Bu metrikler, modelin sınıfları doğru bir şekilde ayırt etme kabiliyetini değerlendirir ve hangi alanlarda iyileştirmeye ihtiyaç duyduğunu belirlemede önemli rol oynamaktadır.

Doğruluk (Accuracy), tüm tahminlerin içindeki doğru tahminlerin oranıdır. Doğru sınıflandırılmış örneklerin tüm örneklere oranı olarak hesaplanır. Ancak, sınıflar dengesizse yanıltıcı sonuçlar verebilir çünkü yüksek frekanslı sınıflara doğru tahmin oranı fazla olurken nadir sınıflar göz ardı edilebilmektedir. Matematiksel ifadesi Denklem 7’de gösterilmiştir.

$$\text{Doğruluk} = \text{Doğru tahmin sayısı} / \text{Toplam tahmin sayısı} \quad (7)$$

Kesinlik (Precision), modelin pozitif olarak tahmin ettiği örneklerin ne kadarının gerçekten pozitif olduğunu göstermektedir. Özellikle yanlış pozitif tahminlerin maliyetli olduğu durumlarda önemlidir (Denklem 8).

$$\text{Kesinlik} = (\text{Doğru Pozitif (TP)} + \text{Yanlış Pozitif (FP)}) / \text{Doğru Pozitif (TP)} \quad (8)$$

Duyarlılık (Recall): Modelin gerçekten pozitif olan örnekleri doğru şekilde pozitif olarak tahmin etme oranını gösterir. Yanlış negatif tahminlerin olumsuz etkilerinin fazla olduğu durumlarda (örneğin, hastalık tespiti) bu metrik önem taşımaktadır (Denklem 9).

$$\text{Duyarlılık} = (\text{Doğru Pozitif (TP)} + \text{Yanlış Pozitif (FP)}) / \text{Yanlış Negatif(TP)} \quad (9)$$

F1-skoru, duyarlılık ve kesinliğin harmonik ortalamasıdır ve dengesiz sınıflar olduğunda kullanılır. Hem yanlış pozitif hem de yanlış negatiflerin olumsuz etkilerini dengeleyerek modelin performansını değerlendirir. 1’e yakın bir F1-skoru hem kesinlik hem de duyarlılık açısından güçlü bir modele işaret etmektedir (Denklem 10).

$$F1\text{-skor} = (2 * (\text{Duyarlılık} * \text{Kesinlik})) / (\text{Duyarlılık} + \text{Kesinlik}) \quad (10)$$

ROC Eğrisi (Receiver Operating Characteristic Curve), modelin çeşitli eşik değerlerinde yanlış pozitif oranı (False Positive Rate) ile duyarlılık arasındaki ilişkiyi gösterir. AUC (Area Under Curve) ise eğri altındaki alan demek olup ROC eğrisinin altındaki toplam alanı ölçer ve 0 ile 1 arasında değer alır. 1'e yakın bir AUC, modelin pozitif sınıfı doğru bir şekilde ayırt etme gücünün yüksek olduğunu gösterir.

Özgüllük (Specificity), modelin gerçekten negatif olanları doğru bir şekilde negatif olarak tahmin etme oranıdır. Yanlış pozitifleri azaltmak için kullanılmaktadır (Denklem 11).

$$\text{Özgüllük} = \text{Doğru Negatif (TN)} / (\text{Yanlış Pozitif (FP)} + \text{Doğru Negatif (TN)}) \quad (11)$$

Bu metrikler, modelin genel performansını analiz etmede farklı bakış açıları sunarak, güçlü ve zayıf yönlerin belirlenmesine olanak tanır. Hangi metriklerin kullanılacağı, sınıflandırma probleminin doğasına ve yanlış sınıflandırmanın etkilerine göre belirlenmesi gerekmektedir.

## **Bölüm 4**

### **Çok Kriterli Karar Verme**

Yöneticiler, rutin olmayan kararları alırken kalitatif ve kantitatif tüm verileri değerlendirmekte, istatistiksel ve matematiksel teknikler ile ekonometrik modellerden yararlanmaktadır (Atağan, 2013). Çok Kriterli Karar Verme (ÇKKV) yöntemleri, birden fazla kriterin göz önüne alınması gereken karmaşık karar problemlerini çözmek için kullanılan sistematik yöntemlerdir. ÇKKV, özellikle işletme yönetimi, mühendislik, ekonomi ve kamu yönetimi gibi alanlarda yaygın olarak kullanılmaktadır. Bu yöntemler, bir karar vericinin farklı alternatifleri bir arada değerlendirerek en uygun çözümü seçmesine yardımcı olmaktadır. ÇKKV yöntemleri, analitik ve matematiksel tekniklerin uygulanması sayesinde, karar süreçlerini daha objektif ve sistematik hale getirmektedir.

ÇKKV, birden fazla kriterin dikkate alındığı karar verme problemlerinde, alternatiflerin kıyaslanarak en uygun seçeneğin belirlenmesini sağlamaktadır. Bu yöntemler, basit seçimlerden karmaşık stratejik kararlara kadar geniş bir yelpazede uygulanabilmektedir. Özellikle karar sürecinde hem nicel hem de nitel verilerin dikkate alınmasını sağladığı için birçok sektörde etkin bir çözüm olarak görülmektedir. ÇKKV yöntemlerine ilişkin yapılan derlemeler, bu yöntemlerin farklı alanlardaki farklı kullanımlarını raporlamışlardır (Zavadskas, Turskis ve Kildienė, 2014). Örneğin, dijital tedarik zinciri yönetiminde çok kriterli karar alma yöntemleri, mevcut literatürde geniş bir uygulama alanı bulmaktadır (Büyüközkan ve Göçer, 2018).

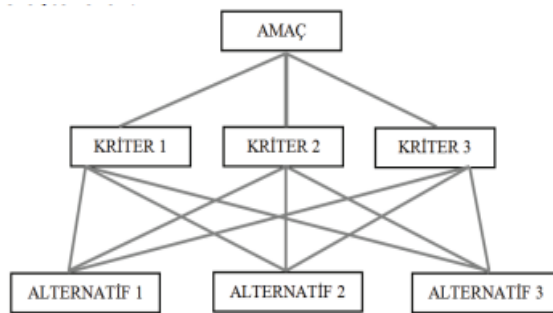
ÇKKV yöntemleri, 20. yüzyılın ortalarından itibaren matematik, istatistik ve karar teorisinin gelişmesiyle yaygınlaşmaya başlamıştır. Bu yöntemler, karar teorisi ve matematiksel modelleme tekniklerinin gelişimiyle daha da çeşitlenmiştir. Günümüzde, veri yoğunluklu karar problemlerinin çözümünde etkin bir şekilde kullanılmaktadır. ÇKKV yöntemleri, işletmelerde yatırım kararları, ürün geliştirme, tedarik zinciri yönetimi, stratejik planlama gibi birçok alanda tercih edilmektedir. Aynı zamanda

kamu sektöründe, politika geliştirme, altyapı projeleri, sağlık hizmetleri gibi alanlarda da kullanılmaktadır. Bu yöntemlerin esnek yapısı, çeşitli alanlarda uygulanabilirlik sağlamaktadır.

ÇKKV'nin temel unsurları, karar verici, alternatifler, kriterler ve kriter ağırlıklarıdır. Karar verici, bir problemi çözmek amacıyla alternatifleri ve kriterleri belirlemektedir. Alternatifler, değerlendirilecek seçenekleri ifade ederken; kriterler, bu alternatiflerin hangi özelliklere göre değerlendirileceğini belirlemektedir. Kriter ağırlıkları ise her bir kriterin önem derecesini göstermektedir.

Yazında pek çok ÇKKV tekniği bulunmaktadır. Bunların en bilinenlerinden birkaç örnek aşağıda tanıtılmaktadır.

Analitik Hiyerarşi Süreci (Analytical Hierarchy Process – AHP), en bilinen ÇKKV yöntemlerinden biridir. AHP, karmaşık karar problemlerini daha sade bir yapıya indirgeyerek kriterlerin ve alternatiflerin hiyerarşik bir yapıda karşılaştırılmasıyla karar önerilerinin elde edilmesini sağlar. AHP yönteminde öncelikle amaç belirlenmekte ve bu amaç doğrultusunda amacı etkileyen faktörler saptanmaya çalışılmaktadır (Dağdeviren, Akay ve Kurt, 2004). Yöntem, karar kriterlerinin ağırlıklandırılmasını ve alternatiflerin sıralanmasını olanaklı kılar. Bu süreç, karar vericilerin sezgilerini matematiksel bir modele entegre ederek güvenilir sonuçlar üretir. Örneğin, üç kriterle değerlendirilen yapıldığı ve üç firmanın değerlendirildiği bir karar sorununun hiyerarşik tasarımı Şekil 4'te verilmektedir.



Şekil 4. Hiyerarşi örneği (Öneren, Arar ve Çelebioğlu, 2013).

Hiyerarşik yapı kurulduktan sonra, hiyerarşideki her elemanın, üst seviyedeki kriter veya unsurlara göre önem derecesini belirlenir. Bu aşamada, her eleman ikili karşılaştırmalar yoluyla değerlendirilir; yani iki

kriter veya alternatif birbiriyle kıyaslanarak görelî önemleri elde edilir. Bu süreç, karar vericilerin her iki unsur arasındaki öncelik veya tercih seviyesini netleştirerek, genel bir ağırlıklandırma oluşturmasına olanak tanımaktadır. İkili karşılaştırmalı yargıların oluşturulmasında, bir başka ifade ile A kriterinin B kriterine göre ne kadar önemli olduğunun karar vericiye sorulmasında karar verici 1-9 aralığındaki tercih ölçeğinden faydalanır (Palaz ve Kovancı, 2008). Bu ölçek Tablo 3'te gösterilmektedir.

Tablo 3

*AHP Tercih Ölçeği (Saaty, 2004)*

<b>Önem Yoğunluğu</b>	<b>Tanım</b>
1	Eşit önem durumu
2	Zayıf veya biraz daha az önemli
3	Orta derecede daha önemli
4	Orta dereceden fazla önemli
5	Güçlü derecede önemli
6	Güçlü dereceden daha fazla önemli
7	Çok kuvvetli ya da kanıtlanmış derecede önemli
8	Çok çok güçlü derecede önemli
9	Aşırı önem durumu

İkili karşılaştırmalardan elde edilen değerlerle bir karşılaştırma matrisi oluşturulur. Bu matris, kriterler arasındaki göreceli önem derecelerini içerir. Daha sonra, bu matris kullanılarak her bir kriterin ağırlığı hesaplanır. Bu işlem, karar vericinin öznelliğini dikkate alarak kriterlerin görelî önemini belirler. Matrisin normalize edilmesi ve tutarlılığının kontrol edilmesi ile ağırlık değerleri daha güvenilir hale getirilir. Her bir alternatif, belirlenen kriterlere göre değerlendirilir. Bu değerlendirme sonucunda, her bir alternatifin toplam puanı elde edilir. En yüksek toplam puana sahip alternatif, en iyi seçenek olarak belirlenir.

TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) yöntemi, Hwang ve Yoon tarafından 1980 yılında geliştirilen bir ÇKKV yaklaşımıdır. Bu yöntem, karar verme süreçlerinde alternatiflerin değerlendirilmesi için pozitif ve negatif ideal çözümleri temel alır. TOPSIS yöntemi, karar seçeneklerini değerlendirirken, pozitif ideal çözüm ile olan mesafesi en kısa ve negatif ideal çözüm ile olan mesafesi en yüksek olan

seçeneđi bulmaya çalışır. Böylece, ideal ve ideal olmayan çözümler arasında bir karşılaştırma yapılmış olur.

Ada ve Çakır (2022) personel seçimi üzerine yapmış oldukları çalışmada, yedi kriterden oluşan değerlendirme ölçeđi çerçevesinde 10 adaydan oluşan örneklem setini AHP ve TOPSIS ile değerlendirmişlerdir. Tüm süreç sonunda en yüksek puanı alan aday, karar vericiye önerilmiştir. Soba ve Eren (2011) ise, özel bir otobüs firmasının finansal ve finansal olmayan verilerini göz önünde bulundurarak TOPSIS yöntemi ile şirketin yıllara göre başarı sıralamasını yapmışlardır. Özgül (2006), özel bir işletme için ERP yazılımı seçim sürecini TOPSIS ve AHP yöntemlerini kullanarak yapmıştır.

PROMETHEE (Preference Ranking Organization METHod for Enriched Evaluation), üstünlük temelli ÇKKV tekniklerinden biri olup alternatifleri belirli kriterlere göre sıralayarak en iyi çözümlü bulmayı amaçlamaktadır. PROMETHEE, alternatiflerin birbirine göre olan üstünlük durumlarını değerlendirip karşılaştırarak, en iyi alternatifin belirlenmesine yardımcı olmaktadır. PROMETHEE yöntemi, karar vericinin kriterlere verdiği öneme göre alternatifler arasında bir tercih sıralaması yapmaktadır. Bu yöntem, özellikle çok sayıda alternatifin ve kriterin olduğu durumlarda daha basit ve hızlı bir analiz sağlamaktadır. Alternatifler, pozitif ve negatif akımlarına veya net üstünlük akımına göre sıralanmaktadır. En yüksek net akıma sahip olan alternatif, en iyi seçenek olarak belirlenmektedir. PROMETHEE I ile kısmi öncelikler belirlenebilir. Alternatiflere ait pozitif ve negatif akım değerlerinin ikili olarak mukayese edilmesi sonucunda üç farklı durum ortaya çıkabilir: bir alternatifin diđer bir alternatife üstün olması, alternatiflerin birbirlerinden farksız olması ve alternatiflerin birbirleriyle karşılaştırılmaz olmasıdır (Atıcı ve Ulucan, 2009). PROMETHEE I bu üç öncelik durumunu da görselleştirerek kısmi bir sıralama imkânı sunar. PROMETHEE II ise pozitif ve negatif akım değerlerinin farkını alarak net akım hesaplar ve alternatifleri net akım değerlerine göre tam sıralamaya tabii tutar.

#### **4.1 ÇKKV Yöntemlerinde Ađırlıklandırma Yöntemleri**

Çok kriterli karar verme (ÇKKV) yöntemlerinde, kriterlerin önem

derecesini belirlemek için kullanılan ağırlıklandırma yöntemleri, karar sürecinin doğruluğunu ve güvenilirliğini artırmak açısından büyük önem taşır. Ağırlıklandırma, kriterlerin etkisini belirlemeyi sağlayan bir aşamadır.

**4.1.1 Öznel (Subjektif) ağırlıklandırma yöntemleri.** Öznel yöntemlerde, ağırlıklar karar vericilerin bilgi ve deneyimlerine dayanarak belirlenir. Kriterlerin ağırlıkları, karar vericilerin tercihlerini yansıtır. Karar vericilerin kriterlerin önem derecelerini sıralaması ve aralarındaki farkları belirlemesi tamamen kişisel yargılara dayanır. Farklı kişiler aynı problem üzerinde farklı sıralamalar ve aralıklar tanımlayabilir. Bu öznellik çözümün genelleştirilebilirliği açısından sorun teşkil edebilmektedir. Uzman kişiler değişirse, kriterlerin ağırlık setlerinde değişim olması oldukça olasıdır. Bazı öznel ağırlıklandırma yaklaşımları aşağıda özetlenmektedir.

- Doğrudan Ağırlıklandırma: Karar verici, her bir kritere doğrudan bir ağırlık atar. Bu yöntem basit ve hızlı olması nedeniyle tercih edilebilmektedir.
- Oran Yöntemi: Karar verici, kriterlerin birbirine göre göreceli önemlerini belirlemek için oranlar kullanır. Örneğin, bir kriterin diğerine göre iki kat daha önemli olduğu belirtilir.
- Simos Yöntemi: 1989 yılında Jean SIMOS tarafından geliştirilen bu yöntem, karar vericilerin kriterlerin önem derecelerini sezgisel ve görsel bir yaklaşımla ifade etmelerini sağlar. Yöntem, özellikle karmaşık ve sayısal olmayan kriterlerin yer aldığı durumlarda kullanılmaktadır.
- AHP (Analitik Hiyerarşi Süreci): AHP, kriterler arasındaki ikili karşılaştırmalarla kriterlerin göreceli önemlerini belirler. Bu yöntemde her kriter diğer kriterlerle karşılaştırılarak, bir matris oluşturulur ve matristen ağırlıklar türetilir. Fountzoula ve Aravossis (2021), AHP'nin kamu sektörü projelerinin değerlendirilmesinde nasıl uygulandığını ele almıştır. Bu çalışma, 2010-2020 yılları arasında yayınlanmış AHP uygulamalarını kapsamakta olup, özellikle şehir planlaması, güvenlik ve kamu hizmetleri gibi alanlarda kullanımlarını özetlemektedir.
- SWING Yöntemi: Karar vericiler, kriterleri en az önemli olandan en önemliye sıralar ve ardından en önemli kritere bir referans ağırlık atanır. Diğer kriterlerin ağırlıkları bu referansa göre belirlenir. Stojčić vd. (2019) yapmış oldukları çalışmada, Swing yönteminin, enerji ve altyapı projelerinde çevresel, sosyal ve ekonomik kriterleri ağırlıklandırmak için

uygun bir araç olduğunu vurgulamışlardır. Bu inceleme, sürdürülebilir projelerde ÇKKV yöntemlerinin rolünü değerlendirmektedir.

- SWARA Yöntemi: Karar vericiler her bir kriteri bir önceki kritere göre önem derecesine göre sıralar ve göreceli önem katsayılarını belirler. Bu nedenle SWARA yöntemi de kişisel yargılara ve uzman görüşlerine dayalı olarak ağırlıklandırma yaptığı için öznel bir yaklaşım olarak sınıflandırılmaktadır.

**4.1.2 Nesnel (Objektif) ağırlıklandırma yöntemleri.** Nesnel yöntemlerde ağırlıklar, veri setinin doğrudan analizine dayalı olarak belirlenir. Bu yöntemlerde karar verici müdahalesi minimum düzeydedir, bu nedenle daha tarafsız bir ağırlıklandırma olanağı tanır.

- Entropi Yöntemi: Veri setinin taşıdığı bilgi miktarı dikkate alınarak kriterlerin ağırlıkları belirlenir. Belirsizliği yüksek yani taşınan bilginin daha geniş aralıkta salındığı kriterler daha yüksek ağırlık sahibi olurlar. Zhang (2023), entropi temelli yaklaşımları kullanarak su izleme ağlarının nasıl daha etkin tasarlanabileceğini ele almıştır. Çalışmada özellikle yağış, akış ve su kalitesi ağları gibi farklı izleme türleri değerlendirilmiş ve entropi yönteminin ağ tasarımında veri doğruluğunu ve kapsamını artırmaya yönelik faydaları incelenmiştir. Entropi yöntemi, karar verici etkisinden bağımsızdır; yalnızca verinin dağılımını ya da bir diğer deyişle taşıdığı farklı bilgi hacmini dikkate almaktadır. Verideki çeşitliliği yakalamada oldukça etkilidir. Bununla birlikte bir takım zayıflıkları da söz konusudur, örneğin kriterler arasındaki ilişkiyi dikkate almaz. Girdi verilerindeki normalizasyon veya hata, ağırlıkları ciddi şekilde etkileyebilmektedir.
- CRITIC (Criteria Importance Through Intercriteria Correlation): Kriterler arasındaki korelasyonu ve bir kriterdeki verilerin standart sapmasını aynı anda dikkate alarak, daha objektif ağırlık değerleri sağlayan bir karar verme metodudur (Krishnan, Kasim, Hamid & Ghazali, 2021). Kriterlerin değişkenliği ve birbirleriyle olan ilişkileri dikkate alınarak ağırlıklar belirlenir. Kriterler arasındaki korelasyonlar hesaba katılır; yüksek korelasyon, bir kriterin diğerine göre önemini azaltır. CRITIC yöntemi, kriterler arasındaki bağımlılığı hesaba katar. Aynı zamanda kriterlerin ayırım gücünü de değerlendirir. Karar verici etkisi olmadan tamamen veriye dayalı ağırlıklandırma yapar. Entropiye kıyasla hesaplama süreci daha karmaşıktır.

Homojen verilerde CRITIC'in ayırım gücü azalabilir.

- Standart Sapma Yöntemi: Olasılık ve istatistik bilimlerinde bir olasılık dağılımının standart sapması, değerlerin yayılım ölçümü olarak değerlendirilebilir (Altıntaş, 2023). Kriterlerin veri içindeki standart sapması hesaplanarak ağırlıklandırma yapılabilir. Standart sapması yüksek olan kriterler daha fazla bilgi içerdiğinden, daha yüksek ağırlık alır. Standart Sapma Yöntemi, matematiksel olarak kolay ve hızlı yöntemlerden biridir. Karar verici etkisinden bağımsızdır ve tamamen verilere dayanır. Standart Sapma Yöntemi de kriterler arasındaki ilişkiyi dikkate almaz.

#### 4.2 ARAS Yöntemi

ARAS (Additive Ratio Assessment) yöntemi, ÇKKV süreçlerinde kullanılan bir yöntemdir. Bu yöntem, alternatiflerin çeşitli kriterler açısından değerlendirilmesi ve bu değerlendirmeler sonucunda en iyi alternatifin belirlenmesi için geliştirilmiştir. ARAS, karar vericilerin farklı kriterler arasında bir denge kurarak daha iyi bir karar almalarına yardımcı olmaktadır. Genellikle mühendislik, üretim, finans ve çevresel yönetim gibi çeşitli alanlarda uygulanmaktadır.

ARAS yöntemi, alternatiflerin ideal çözüme olan bağıl uygunluğunu esas alarak karar vermeyi sağlayan katkı oranlı bir değerlendirme yaklaşımı olarak tanımlanır (Liu ve Xu, 2021). ARAS yöntemi, karar vericilere alternatiflerin karşılaştırılabilir bir şekilde değerlendirilmesi için basit ve etkili bir araç sunmaktadır. Yöntem, alternatiflerin her birinin performansını belirlemek için kriterler arasında bir ilişki kurmaktadır. Bu süreçte, her kriterin önemi dikkate alınarak alternatiflerin toplam puanları hesaplanmaktadır. Yöntem, genellikle karmaşık ve belirsiz durumlarda bile karar verme sürecini kolaylaştırmaktadır.

ARAS yönteminin ilk adımı, karar verme sürecinde dikkate alınacak kriterlerin belirlenmesidir. Bu kriterler, karar vericilerin hedeflerine ulaşmalarına yardımcı olacak şekilde seçilmelidir. Kriterler belirlendikten sonra, değerlendirilecek alternatiflerin tanımlanması gerekmektedir. Bu alternatifler, karar verme sürecinin bağlamına göre değişiklik gösterebilir. Örneğin, bir proje için farklı tedarikçiler, malzemeler veya stratejiler alternatif olarak değerlendirilmektedir. Alternatifler ve kriterler

belirlendikten sonra, her alternatifin her bir kriterdeki performansını gösteren bir matris oluşturulmalıdır. Bu matris, alternatiflerin belirlenen kriterler açısından nasıl bir performans sergilediğini göstermektedir. Bu aşamada verilerin doğru ve güvenilir olması önemlidir.

Performans matrisinin oluşturulmasının ardından, her bir kriterin normalize edilmesi gerekmektedir. Fayda ve maliyet tipinde olması durumuna göre normalizasyon formülleri Denklem 12 ve Denklem 13'de gösterilmiştir. Normalizasyon, farklı ölçeklerdeki kriterlerin karşılaştırılabilir hale gelmesini sağlamaktadır. Genellikle, her kriter için en yüksek performansın 1, diğer alternatiflerin ise bu en yüksek performansa göre oranlarının alındığı bir yöntem kullanılmaktadır.

$$(x_{ij})' = \frac{x_{ij}}{\sum_{i=1}^m x_{ij}} \quad (\text{Fayda tipinde normalizasyon}) \quad (12)$$

$$(x_{ij})' = \frac{\sum_{i=1}^m x_{ij}}{x_{ij}} \quad (\text{Maliyet tipinde normalizasyon}) \quad (13)$$

Normalizasyonun ardından, her bir kriter için bir ağırlık belirlenmelidir. Kriter ağırlığı, ilgili kriterin önem derecesini yansıtmalıdır. ARAS'ın kendine has bir ağırlıklandırma yaklaşımı mevcut değildir. Bu nedenle uygun bir nesnel ya da öznel ağırlıklandırma yaklaşımı kullanılabilir. Daha sonra, her bir alternatifin bütünleşik performans puanı hesaplanır (Denklem 14). Bu hesaplamada alternatiflerin her bir kriterdeki performans değerleri o kriterin ağırlığı ile çarpılarak toplanır.

$$v_{ij} = (x_{ij})' * w_j \quad (14)$$

Bütünleşik performans puanlarının hesaplanmasının ardından, alternatifler toplam puanlarına göre sıralanmalıdır. Bu sıralama, en yüksek puana sahip alternatifin en iyi seçenek olduğunu göstermektedir. Böylece, karar vericiler hangi alternatifin tercih edilmesi gerektiğine dair bir karar alabilmektedirler. Her bir alternatif için toplam puan olan  $S_i$  hesaplanır (Denklem 15). Alternatifler,  $S_i$  değerlerine göre büyükten küçüğe doğru

sıralanırlar. En yüksek  $S_i$  değerine sahip alternatif en iyi alternatif olarak seçilir. ARAS, kriterlerin normalize edilmesini takiben, ağırlıklı toplam skorlar üzerinden her bir alternatifin ideal çözüme olan uzaklığını belirler (Liu ve Xu, 2021).

$$S_i = \sum_{j=1}^n v_{ij} \quad (15)$$

Alternatiflerin ideal çözüme göre görelî etkinliklerini göstermek amacıyla bağıl optimalite katsayısı hesaplanır.  $S_i$  değeri,  $S_0$  değerine bölünür ve buna bağıl optimalite katsayısı ( $K_i$ ) adı verilir.  $K_i$  değeri ideal çözüme olan yakınlık derecesini ifade eder;  $K_i$ 'nin 1'e yaklaşması, alternatifin optimal çözüme daha yakın olduğunu göstermektedir.

Son olarak, tüm alternatifler  $K_i$  değerlerine göre büyükten küçüğe sıralanır. En yüksek  $K_i$  değerine sahip olan alternatif, karar verici açısından en uygun seçenektir.

Hassasiyet analizi yapıldıktan sonra, karar verme süreci tamamlanmış olmaktadır. Bu aşamada, karar verici, en iyi alternatifin hangisi olduğuna dair son bir değerlendirme yapabilmektedir. Ayrıca, alınan kararın nedenleri ve sonuçları ile ilgili bir rapor hazırlanabilmektedir.

ARAS yöntemi, çeşitli uygulama alanlarına sahiptir. Örneğin, tedarikçi seçiminde, proje değerlendirmesinde, çevresel etki analizlerinde ve finansal yatırım kararlarında kullanılabilir. Bu yöntem, alternatiflerin çeşitli kriterler açısından sistematik bir şekilde değerlendirilmesini sağladığı için geniş bir yelpazede uygulanmaktadır.

Idaman, Amrullah ve Rolanda (2024), üretim şefinin seçimi için ARAS yöntemini uygulayarak karar destek sistemlerinin etkinliğini arttırmıştır. Reza ve Majid (2013), AHP ve ARAS'ı kullanarak bankaları internet bankacılığına yönelik güven algısı açısından değerlendirmişlerdir. Bu incelemenin sonuçları, algılanan gizli riskin en önemli kriter olduğunu göstermiştir. Aras ve Yıldırım (2020) ise, sosyo-ekonomik refah düzeylerini incelemek amacıyla G-20 ülkelerini ARAS yöntemi ile değerlendirmişlerdir. Çalışmada, 20 farklı gösterge dikkate alınmış ve 2018 verileri kullanılarak oluşturulan bir endeks üzerinde analiz yapılmıştır. Bu analiz sonucunda,

Avusturya, Kanada ve Amerika'nın sosyo-ekonomik refah açısından en yüksek seviyelerde olduğu bulunmuştur. Ayrıca, Akçakaya ve Akçakaya (2019) büyükşehirlerin çevresel performanslarını belirlemek için Entropi temelli ARAS ve COPRAS yaklaşımlarını kullanmışlardır. Bu çalışmalar, şehirlerin çevresel sürdürülebilirlik düzeylerini değerlendirmek için önemli bir temel sağlamaktadır. Benzer bir başka çalışma, Kenger ve Organ (2017) tarafından gerçekleştirilmiştir. Hatay ilinde bir banka için personel seçim sürecinde Entropi ve ARAS yöntemleri kullanılarak değerlendirme yapılmıştır. Kurşunmaden (2024) Borsa İstanbul'da işlem gören şirketlerin 2023 yılı kurumsal yönetim olgunluk endeksi verilerini kullanarak, CRITIC yöntemiyle kriter ağırlıklarını belirlemiş ve ARAS yöntemiyle sektörel performansları sıralamıştır. Sonuçlar, Merkezi Kayıt Kuruluşu tarafından hesaplanan genel puanlarla farklılık gösterdiği görülmüştür. Sivam ve Rajendran (2022), derin çekme kaplarının üretiminde yapay sinir ağı parametrelerinin optimizasyonu için Taguchi ve ARAS yöntemlerini birleştirmiştir. Elde ettikleri bulgularda, bu yöntemin üretim verimliliğini artırdığı sonucuna ulaşmıştır.

## Bölüm 5

### Yeni Bir Nesnel Ölçüt Ağırlıklandırma Yöntemi Önerisi: RAFOW

RO tekniği ile bir ağaç modeli oluşturulduğunda, öznitelikler düğümlere rastgele atanır. Bu atama işlemi  $n$  sayıdaki ağaç için ayrı ayrı uygulanmaktadır. Bu çalışmada, metrik değerler üzerinden başarılı sonuçlar elde edilmiş olan bir RO modeline farklı bir yaklaşım uygulanarak yeni bir nesnel ağırlık çıkarımı yaklaşımı önerilmektedir. RO yöntemi ile oluşmuş  $n$  adet ağacın her biri üzerinde seviye 0, seviye 1 ve seviye 2 seviyelerindeki düğümlere rastsal atanan özniteliklerin bölünme/aday oranlarının, o model üzerindeki önemine işaret edeceği fikrinden yola çıkılmıştır. Bu bakış açısı, saflılık bakış açısından farklı olarak bir kriterin bir düğüm noktasına kaç kere önerildiği ve bunların kaçında denenen düğümün yerini kazandığı dikkate alınmıştır. bölünme/aday oranları o özneliğin model üzerindeki etkisini gösteren önem değeri olarak dikkate alınmaktadır. Ağacın 3 seviyesi için bu oran, Knime Analytic Tool üzerinde formülize edilmiş olup Denklem 16'da gösterilmektedir.

$$\begin{aligned} \text{Özellik Önem Skoru} &= \left( \# \frac{\text{Bölünme}(\text{level } 0)}{\# \text{Aday}(\text{level } 0)} \right) + \\ &0,50 \left( \# \frac{\text{Bölünme}(\text{level } 1)}{\# \text{Aday}(\text{level } 1)} \right) + 0,25 \left( \# \frac{\text{Bölünme}(\text{level } 2)}{\# \text{Aday}(\text{level } 2)} \right) \quad (16) \end{aligned}$$

Nesnel ağırlık çıkarımını sağlayan önerilen RAFOW (RANdom Forest-based Objective Attribute Weighting) algoritmasının adımları şu şekildedir:

- Adım 1: Kullanılacak verilerin veri analizi ve ön işleme süreçlerini gerçekleştir.
- Adım 2: Veri ön işleme adımları gerçekleştirilmiş veri setini kullanarak test verisini ayır.
- Adım 3: Eğitim verisi ile RO modelini oluştur.
- Adım 4: Test verisi ile modelin değerlendirme metriklerini oluştur.
- Adım 5: Değerlendirme metrikleri tatmin edici derecede başarılı ise, Adım 6'ya geç. Metrikler yeterli çıktıyı

sağlamadıysa Adım 1'e geri dön.

- Adım 6: Başarılı model üzerinde Denklem 16'daki Özellik Önem Skoru değerlerini hesapla.
- Adım 7: Özellik Önem Skorlarını normalize et ve 0 ile 1 arasında özniteliklerin ağırlıklarını hesapla.

Karar ağaçları ikili (binary) yapıda olduğu için Seviye 0'da: 1 düğüm vardır; Seviye 1'de 2 düğüm; Seviye 2'de 4 düğüm bulunur. Yani her seviye, bir öncekinin 2 katı kadar düğüm içerir ve toplam düğüm sayısı geometrik olarak artar. Her seviye daha az veriyi temsil eder ve örnek sayısı azalmış olur. Her seviye aşağı indikçe, veri kümesinin yalnızca küçük bir kısmına karar verilir. Bu nedenle, alt seviyelerdeki bölünmelerin katkısı etki bakımından daha düşüktür.

Formüldeki her seviyenin ağırlığını hesaplarırken sondaki  $2^{-\text{seviye sayısı}}$  katsayısı ile çarpım işlemi, hem düğüm sayısındaki artışı hem de her düğümün temsil ettiği örnek oranındaki azalmayı eş zamanlı telafi eder. Aşağı seviyelere inildikçe ağırlığın azalması ilkesi gözetilmiş olur.

Önerilen yeni nesnel ağırlıklandırma tekniğinin süreç adımları elideki veri setine uygulanmış olup, personelin işten ayrılma kararını etkileyen özniteliklerin ağırlıkları hesaplanmıştır.

Bu çalışmanın en önemli katkısı, RAFOW olarak adlandırılan yenilikçi bir öznitelik ağırlıklandırma prosedürünün öneriliyor olmasıdır. Çoğu ÇKKV yöntemi özel bir öznitelik ağırlıklandırma adımı içermez ve nesnel ve/veya öznel ağırlıklandırma yöntemi ile entegre edilir. Özellikle büyük veri setleri için karar vericinin ekstra bir çaba göstermesine gerek kalmaksızın bu algoritma sayesinde nesnel ağırlıkların belirlenmesi mümkün olabilecektir. Karar verici, ek bir ağırlıklandırma prosedürü uygulama zorunluluğundan kurtulacak ve büyük veri setleri ile çalışan veri bilimcinin ÇKKV konusunda yeterli bilgi veya deneyime sahip olmaması durumunda bile bu analitik süreç uygulama kolaylığı ile karar vericiye bir karar desteği sunacaktır. Bunun için de Knime, SPSS gibi analitik araçlarda gömülü şekilde oluşturulacak bir düğüm sayesinde RAFOW yaklaşımı kolayca uygulanabilecektir.

## Bölüm 6

### İnsan Kaynakları Analitiğinde Bir Uygulama

Bu çalışmada, araştırma yapmak amacıyla açık kaynak olarak Kaggle platformunda bulunan HR\_comma\_sep.csv isimli veri seti kullanılmıştır. Çalışan personelin işten ayrılma tahmini yapabilmek, tahmin modeli üzerinden yeni bir ağırlıklandırma yöntemi kullanarak öznitelikleri ağırlıklandırmak ve yeni ağırlıklandırma yöntemi ile elde edilmiş ağırlıkları ARAS yöntemi ile birlikte kullanarak yeni bir karar destek sistemi oluşturmak bu çalışmanın amaçlarıdır. Bu karar destek sistemine yeni personeller eklenip veya çıkarıldığı takdirde istenilen herhangi bir  $t$  anında model yeniden çalıştırılıp, güncellenen dinamik ağırlıklara göre işten ayrılma kararları ARAS yöntemiyle kolaylıkla sıralanabilecektir.

Bu çalışmanın yapıldığı veri setindeki özniteliklerin yapısı Tablo 4’te verilmektedir. Veri seti 14.999 personel verisinden oluşmaktadır. Kullanılan veri setindeki öznitelikler literatürde yapılan çalışmalarda özniteliklerle büyük ölçüde eşleşmektedir. Literatürde işten ayrılma probleminde etkisi olan öznitelikler incelenmiş ve çalışmanın sonundaki **Ekler** başlığında detaylı olarak gösterilmiştir. Çalışmada kullanılan özniteliklerin, literatürde yapılmış çalışmalarda özniteliklerle büyük ölçüde örtüştüğü ortaya konuşmuştur. Literatürde işten ayrılmayı etkileyen kriterler 8 ana grupta toplanmış ve aşağıdaki Tablo 5’te gösterilmiştir.

Tablo 4  
*Veri Setini Oluşturan Özniteliklerin Yapısı*

Sütun Adı	Veri Tipi	Alt Sınır	Üst Sınır
Emp Id	Metin	-	-
Satisfaction_level	Ondalıklı sayı	0,09	1
Last_evaluation	Ondalıklı sayı	0,36	1
Number_project	Tamsayı	2	7
average montly hours	Tamsayı	96	310
time spend company	Tamsayı	2	10
Work_accident	Tamsayı	0	1
left	Tamsayı	0	1
promotion last 5years	Tamsayı	0	1
Department	Metin	-	-
salary	Metin	-	-

Tablo 5

*Literatürde Personelin İşten Ayrılmasını Etkileyen Parametreler*

Literatürde İşten Ayrılmayı Etkileyen Etkiler ve Ana Kriterler	Kapsamlar ve Açıklamalar
İş tatmini	Personelin yaptığı işten haz alması, mutluluğu.
Örgütsel Bağlılık	Kendini bulunduğu departmanın veya şirkete üye olarak hissetmesi.
İş Stresi, Çalışma Süresi ve Tükenmişlik	Baskı, huzursuz çalışma ortamı, kendine vakit ayıramama, dinlenememe.
Yönetim ve Liderlik	Adil bir yönetici, yapılan çalışmaların görülmesi ve onanması, yol gösterici etki.
Kariyer Gelişimi, Terfi ve Eğitim	Kişisel gelişme ve ilerlenebileceğine dair iç görüşü, yeni bilgiler edinebilme yol haritası.
Ücret, Performans Değerlendirmesi ve Ödüllendirme	Tüm maddi kazanç unsurları ve yönetici tarafından performans durumu.
Çalışma Koşulları ve Esneklik	Uzaktan/Hibrit çalışma, sağlığı tehdit edebilecek iş yeri fiziki koşulları.
Aile/İş Dengesi	Tatmin edici sürelerde kendine ve aileye vakit ayırabilme, yenilenebilme ihtiyacı.

Kullanılan veri setinde bu verilerin nasıl elde edildiği veya hesaplama metodu hakkında bilgi mevcut değildir. Veri setinin datacard bölümünde bir not veya bilgi göze çarpmamaktadır. Çalışma, verili halleri kullanılarak gerçekleştirilmiştir. Sütun isimleri ve veri yapısı içerilen veri hakkında bilgi sağlamaktadır. Verilerin temel istatistik bilgileri Tablo 6'da gösterilmiştir. Kriter tanımlamaları ise şu şekildedir:

- Emp\_Id: Personelin numarası.
- Satisfaction\_level: Personelin tatmin seviyesi.
- Last\_evaluation: Son değerlendirme puanı.
- Number\_project: Yer aldığı toplam proje sayısı.
- average\_monthly\_hours: Ortalama aylık saat cinsinden çalışma süresi.
- time\_spend\_company: Şirkette toplam yıl cinsinden çalışma süresi.

- Work\_accident: İş kazası geçirip geçirmediği.
- promotion\_last\_5years: Şirkette son 5 yılda terfi alma durumu.
- Department: Çalıştığı bölüm.
- Salary: Maaş seviyesi (Düşük/Orta/Yüksek)
- Left: İşten ayrılma bilgisi (Bağımlı değişken).

Tablo 6  
Verilerin Temel İstatistikleri

Öznitelikler	En küçük	En büyük	Ortalama	Standard Sapma	Varyans	Kayıp Veri
satisfaction level	0,09	1	0,612	0,248	0,061	0
last evaluation	0,36	1	0,716	0,171	0,029	0
number_project	2	7	3,803	1,232	1,5192	0
average montly hours	96	310	201,05	49,943	2494,313	0
time spend company	2	10	3,498	1,46	2,131	0
Work_accident	0	1	0,144	0,351	0,123	0
left	0	1	0,238	0,425	0,181	0
promotion last 5years	0	1	0,021	0,144	0,02	0

### 6.1 Veri Ön İşleme

Veri setinde personelin departman ve maaş bilgileri kategorik değişken şeklindedir. Maaş bilgisi düşük, orta ve yüksek olarak firma tarafından etiketli olarak paylaşılmıştır. Bu bilgileri kullanabilmek için “One-to-Many” düğümü yardımıyla veri manipülasyona (0-1) uğratılmıştır. Dönüşüm gerçekleştikten sonra modele eski halinin girdi olarak verilmemesi için veri setinden kaldırılmıştır. Satırlar üzerinde yapılan kontrolle boş verinin bulunmadığı görülmüştür. Herhangi bir anomali değerinin olmadığı, hayatın olağan akışına ters gelecek bir bilginin bulunmadığı ve modellemeye uygun hale bir veri setinin elde bulunduğu anlaşılmıştır.

### 6.2 Makine Öğrenmesi Yöntemleri ile Tahminleme Yapılması

Veri manipülasyonu sonrası verisetindeki Satisfaction\_level, Last\_evaluation, Number\_project, average\_montly\_hours, time\_spend\_company, Work\_accident, promotion\_last\_5years, Department, salary öznitelikleri modellere girdi olarak verilmiştir.

Yapay zekâ alanının çatısı altında yer alan makine öğrenmesi kavramı, insan beyninin hesaplayamayacağı kadar karmaşık model ve süreçler

oluşturarak analitik çıktılar üretmeyi sağlayan bir teknikler topluluğudur. Karar alma sürecinde geçmiş verileri kullanarak tahmin sonuçları üretilir. Bu çalışma kapsamında regresyon algoritmalarından olan Gradyan Artırma, Rastgele Orman, Ekstrem Gradyan Artırma regresyon teknikleri kullanılmış olup ağaç teknikleri üzerinde performans karşılaştırması da gerçekleştirilmiştir.

3 farklı makine öğrenmesi algoritması eldeki veri setine uygulanmış ve sonuçları karşılaştırılmıştır. İnsan kaynakları analitiği problemindeki özneliklerin önem seviyelerinin belirlenmesinden önce, modellerin performans göstergeleri Tablo 7’de paylaşılmaktadır.

Tablo 7

*Yöntemlerin Performans Sonuçları*

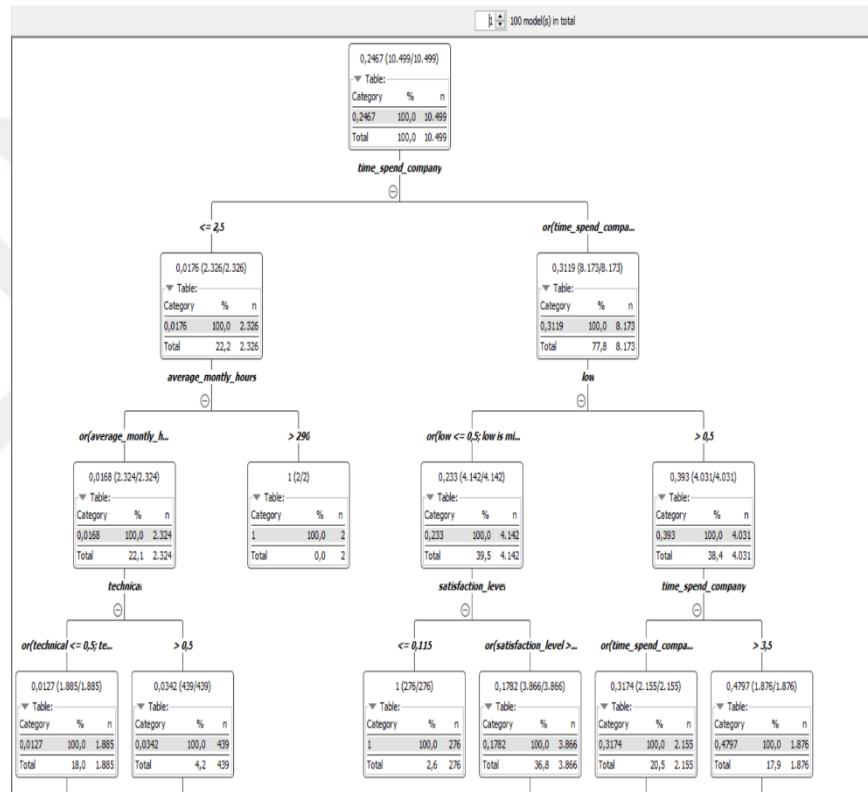
	Rastgele Orman Yöntemi (Ağaç sayısı=100, tree deep=10)	Gradyan Artırma Yöntemi (Ağaç sayısı=10, tree deep=10)	Ekstrem Gradyan Artırma Yöntemi (tree deep=10)
R <sup>2</sup> Skoru	0,921	0,854	0,870
Düzeltilmiş R <sup>2</sup>	0,906	0,853	0,869
Ortalama Mutlak Hata	0,014	0,026	0,022
Hata Kareleri Ortalamasının Karekökü	0,044	0,031	0,023

Her 3 modelin de oldukça başarılı tahminleme yapabildiği görülmektedir. Ortalama hata skorları birbirine oldukça yakın çıktığı için yöntem seçiminde R<sup>2</sup> değeri 1’e en yakın çıkan RO modeli üzerinden öznelik önemi belirlenmiştir.

R<sup>2</sup> Skoru, bir modelin açıklayıcı gücünü gösterir. 1’e ne kadar yakınsa, modelin hedef değişkenin varyansını o kadar iyi açıkladığı anlaşılır. RO 0,921 ile bu ölçütte en yüksek değeri sağlamaktadır. Bu değer, modelin hedef değişkeninin %92,1’ini açıkladığı anlamına gelmektedir. Ortalama mutlak hata, modelin tahmin ettiği değerler ile gerçek değerler arasındaki ortalama farkı ölçer. Daha düşük bir ortalama mutlak hata daha başarılıdır. RO, 0,014 ile en düşük oranı sağlamaktadır. Düzeltilmiş R<sup>2</sup> değerlerine göre de karşılaştırma sonucu benzerdir. RO 0,906 gibi yüksek bir değer ile verileri oldukça iyi açıklamaktadır. Hata kareleri ortalamasının karekökü, hataların

karekökünü alarak ortalamayı hesaplar ve büyük hatalara daha fazla ağırlık verir. RO 0,044 değeri ile en yüksek değere sahip yöntem olup büyük hatalar konusunda diğer iki modele göre bu performans ölçütünde daha zayıf kalmıştır. Bu ölçüt açısından en iyi performansı 0,023 ile Ekstrem Gradyan Artırma sağlamaktadır.

Bu performans analizi sonucunda genel olarak, RO algoritmasının veri setinin yapısına en uygun model belirlenmiş olmaktadır. Oluşturulan 100 adet RO ağaçlarından biri örnek olarak Şekil 5’te gösterilmektedir.



Şekil 5. Oluşturulan ağaç modellerinden birinci sıradaki ağacın gösterimi.

Eğer problem sınıflandırma bakış açısıyla ele alınırsa elde edilen RO modelinin değerlendirme metrikleri Tablo 8’deki gibi oluşmaktadır.

Tablo 8

Sınıflandırma Metriklerinin Gösterimi

	Doğru Pozitif	Yanlış Pozitif	Doğru Negatif	Yanlış Negatif	Duyarlılık	Kesinlik	Hassaslık	Özgüllük
1	1011	9	3419	61	0,943	0,991	0,943	0,997
0	3419	61	1011	9	0,997	0,982	0,997	0,943

Bu metriklerle ilişkin yorumlar ařađıda zetlenmektedir.

- Model iki sınıfı da başarılı bir şekilde ayırt edebilmektedir.
- Pozitif sınıf için kesinlik ok yüksek bulunmuřtur. Bu da modelin "pozitif" dediđi rneklerde yanılma payının az olduđu anlamına gelmektedir.
- Negatif sınıf için duyarlılık ve hassaslık ok yüksek bulunmuřtur. Buna gre, model "negatif" olanları neredeyse tamamen yakalayabilmektedir.
- Model, hem kesinlik hem duyarlılık aısından oldukça gl bir performans gstermiřtir.
- Dengesizlik veya ařırı đrenme belirtisi olarak deđerlendirilebilecek herhangi bir bulgu sz konusu deđildir.

### 6.3 RAFOW ile Nesnel Ađırlıklandırma

Ađa algoritmalarında nce kk dđmn belirlenmesi, sonrasında ise ilk hangi nitelikten blneceđi ve buna ilişkin kuralların ne olduđunun tespit edilmesi gerekmektedir. En iyi ayırım noktalarının tespit edilmesi için ise gini indeks yntemi kullanılabilir. Gini indeks deđeri, bir sınıfın iindeki izafi sıklıđı ifade eder. Her bir zellik iin ayrı ayrı hesaplanan Gini indeks deđerleri arasından en kk olanı seilir. Blnme iřlemi bu deđere gre yapılmaktadır. Bu iřlemler kalan veriler iin de tekrar edilir ve diđer blnmeler iin hesaplamalar yapılır. Sonu olarak, bir zniteliđin gini indeks deđeri ne kadar dřk ise nemi ve dolayısıyla ađırlıđı o kadar yüksek olacaktır. Sklearn ktphanesinin ierisinde yer alan Feature Importance fonksiyonu, zellik nemlerini model zerinde otomatik olarak hesaplayabilmektedir. Bu ađırlıklar Tablo 10'da RAFOW ađırlıkları ile karřılařtırmalı olarak verilmektedir.

"cret" ve "Departman" gibi kategorik deđiřkenler, ikili (0-1) deđerlere dnřtrlerek maniple edilmiřtir. Sonu olarak, stn sayısı artmıřtır. Kategorik deđiřkenlerin nihai stn biiminde oluřturulmuř versiyonuna gre karřılařtırmalar yapılmıřtır. Tablo 9'da her znitelik iin nem skoru normalleřtirilmemiř formda hesaplanmıřtır. zellikler arasında, alıřan memnuniyeti en nemli znitelik olarak bulunmuřtur. te yandan,

muhasebe departmanında olmak, çalışanın şirketten ayrılma kararında en az etkiye sahip olarak bulunmuştur. Bu sonuçlar, diğer herhangi bir organizasyonun özel koşullarına bağlı olarak değişiklik gösterebilir.

Tablo 9

*Önerilen RAFOW Tekniğine Göre Özellik Önemlerinin Gösterilmesi*

Öznitelikler	# bölünme (seviye 0)	# bölünme (seviye 1)	# bölünme (seviye 2)	#aday (seviye 0)	#aday (seviye 1)	#aday (seviye 2)	Özellik Önem Skoru
satisfaction_level	16	34	73	16	37	94	1,653
last_evaluation	11	20	39	17	32	73	1,093
number_project	16	33	41	17	45	76	1,442
average_monthly_hours	13	27	67	24	42	87	1,055
time_spent_company	14	18	42	19	29	76	1,185
Work_accident	8	16	13	19	36	87	0,680
promotion_last_5years	3	3	7	28	38	83	0,167
sales	0	1	8	17	42	85	0,035
accounting	0	0	5	21	40	72	0,017
hr	1	1	2	19	38	74	0,072
technical	0	1	8	20	37	66	0,043
support	0	0	4	25	51	76	0,013
management	0	1	4	19	39	82	0,025
IT	1	4	2	16	47	79	0,111
product_mng	0	1	8	17	50	71	0,038
marketing	0	2	5	16	33	83	0,045
RandD	3	2	2	24	38	60	0,159
low	6	12	15	22	39	81	0,472
medium	3	10	9	26	35	85	0,284
high	5	10	8	18	52	78	0,399

Ücret ve departman bilgilerini gösteren sütunlar toplandığında, kategorik değişkenin toplam ağırlığı elde edilebilir. Bu çalışmada, alt kırılım düzeyinde bir karşılaştırma yapılmıştır. Kategorik öznitelikler birleştirildiğinde, nihai ağırlık sonuçları Tablo 10'da verildiği şekilde belirlenmiş olmaktadır.

Tablo 10

*RAFOW ve Sklearn Kütüphanesine Göre Bulunan Öznitelik Ağırlıkları*

Öznitelikler	RAFOW ağırlıkları	RAFOW sıra	Saflılık tabanlı Sklearn ağırlıklar	Sklearn sıra
satisfaction_level	0,183	1	0,500	1
last_evaluation	0,121	5	0,124	2
number_project	0,160	2	0,105	4

Tablo 10 (devam)

average_monthly_hours	0,117	6	0,102	5
time_spend_company	0,131	3	0,144	3
work_accident	0,075	7	0,002	8
promotion_last_5years	0,018	9	0,0001	9
salary	0,127	4	0,007	7
departments	0,057	8	0,016	6

RAFOW ve Sklearn algoritmasının ortaya çıkardığı ağırlık sonuçlarına göre,

- satisfaction\_level ölçütü her iki yöntem tarafından da en önemli ölçüt olarak belirlenmiştir. Her iki yöntemde de 1. Sırada çıkmıştır. Bu öznelik, çalışanların memnuniyet düzeyinin çalışan bağlılığı veya işten ayrılma gibi hedef değişkenle doğrudan ve güçlü şekilde ilişkili olduğunu gösteriyor. Tutarlılık açısından çok güvenilir bir belirleyici olduğu görülmektedir.
- last\_evaluation, average\_monthly\_hours, time\_spend\_company öznelikleri iki yöntemde de orta önemde ve tutarlı ağırlıklara sahip olduğu görülmektedir. last\_evaluation, Sklearn'de 2. sırada, RAFOW'da 5. Sırada çıkmıştır. Bu fark, Sklearn modelinin bu değişkenin etkisini daha net ve ayrıştırıcı bulduğunu; RAFOW'un ise daha dengeli dağılım verdiğini gösterir. Genel olarak çalışan performans değerlendirmesinin davranışsal öngöründe güçlü bir rolü olduğunu göstermektedir.
- time\_spend\_company (şirkette kalma süresi) Her iki yöntemde 3. sırada çıkmıştır. Bu tutarlılık, kıdemin hedef değişken üzerindeki etkisinin anlamlı olduğunu göstermektedir.
- number\_project, RAFOW'a göre 2. sırada, Sklearn'de 4. Sırada çıkmıştır. Yüksek sayıda projeye meşgul olmak ya da fazla iş yükü, çalışan performansı veya tükenmişlik riski açısından belirleyici olabilir. Her iki yöntemde de ilk 5'te olması, önemli bir iş gücü metriği olduğunu göstermektedir.
- promotion\_last\_5years ve departments her iki yöntemde de

modele katkısının önemli olmadığı olarak sınıflandırılmıştır.  
Ölçüt bazlı bazı değerlendirmelerin yapılması da olasıdır.

satisfaction\_level:

- Saflık temelli yaklaşımda bu değişkene olağanüstü yüksek bir önem verilmiştir (%50).
- RAFOW bu açıdan daha dengeli bir yaklaşım ortaya koymakta ve herhangi bir kriterin kararı bu derece yüksek bir önemle etkilemesine izin vermemektedir.
- Bu farkın nedeninin, Gini'nin karar ağaçlarının veriyi bölme sırasına bağlı olarak "dominant" olan tek değişkene aşırı ağırlık vermesi olduğu değerlendirilmektedir.

salary:

- RAFOW görel olarak %12,7 oranında orta seviyeli bir önem belirtmiştir.
- Ancak, saflılık temelli yaklaşım bu özneliği neredeyse yok saymış (%0,07).
- Bu durumun, salary değişkeninin veri setindeki dağılımına ve kategorik bir değişken olup dönüşüm şekline bağlı olabileceği değerlendirilmektedir.

Eğer modelin doğruluğu kadar açıklanabilirliği de önemliyse, RAFOW yöntemi daha anlamlı bir yaklaşım sunmaktadır çünkü kriterler arasında önemi de ifade edecek şekilde daha dengeli bir ağırlık dağılımı sunmaktadır. Gini yöntemi, hızlı ve doğrudan ayırım gücüne göre sıralama yaparken bazı değişkenlerin önemini oldukça düşük gösterebilmektedir.

RAFOW yöntemi ile iki farklı nesnel kriter ağırlıklandırma tekniği olan CRITIC ve Entropi metodları kullanılarak bulunan kriter ağırlıkları da karşılaştırmalı olarak Tablo 11'de sunulmaktadır.

Tablo 11

*CRITIC, Entropi ve RAFOW Yöntemleriyle Kriter Ağırlıklandırma*

Columns	CRITIC	Entropi	RAFOW
satisfaction_level	0,046	0,002	0,183
last_evaluation	0,042	0,001	0,121
number_project	0,039	0,001	0,160

Tablo 11 (devam)

average_monthly_hours	0,037	0,001	0,117
time_spend_company	0,029	0,002	0,131
Work_accident	0,058	0,054	0,075
promotion_last_5years	0,024	0,109	0,018
sales(department)	0,081	0,036	0,003
accounting(department)	0,038	0,084	0,001
hr(department)	0,037	0,085	0,008
technical(department)	0,069	0,048	0,004
support(department)	0,063	0,053	0,001
management(department)	0,034	0,089	0,002
IT(department)	0,048	0,070	0,012
product_mng(department)	0,041	0,079	0,004
marketing(department)	0,040	0,081	0,005
RandD(department)	0,038	0,083	0,017
low(salary)	0,090	0,020	0,052
medium(salary)	0,087	0,023	0,031
high(salary)	0,047	0,070	0,044

RAFOW yöntemi, özellikle çalışanların davranışsal ve performans göstergelerine dayalı sayısal özniteliklere yüksek önem atfetmiştir. Özellikle, satisfaction\_level (İş Tatmini Düzeyi): %18,3 ile en yüksek ağırlığa sahip değişken olmuştur. Bu durum, çalışan memnuniyetinin işten ayrılma üzerinde belirleyici bir faktör olduğunu göstermektedir. number\_project, last\_evaluation, average\_monthly\_hours ve time\_spend\_company gibi performans odaklı değişkenler sırasıyla yüksek ağırlıklar almıştır.

Bu yöntemin çıktıları, tahminsel modelleme açısından değerlendirildiğinde anlamlıdır ve karar ağacı tabanlı modellerin ayırt edici gücünü yansıtmaktadır.

CRITIC yöntemi, özniteliklerin varyanslarını ve kendi aralarındaki korelasyonları dikkate alarak bilgi içeriğini ölçen bir tekniktir. low(salary) (%9), medium(salary) (%8.7) ve sales(department) (%8.1) gibi kategorik değişkenler yüksek ağırlıklar almıştır. satisfaction\_level, number\_project, average\_monthly\_hours gibi davranışsal değişkenler daha arka planda

kalmıştır. CRITIC yöntemi, veri setindeki bağımsız bilgi katkısına göre karar vermeyi hedeflediği için, korelasyonu az olan ama varyansı yüksek değişkenleri öne çıkarmaktadır.

Entropi temelli yöntem, öznelitliklerin bilgi kazanımına göre önemini belirler. Bu yöntemde, bilgi belirsizliğini azaltma kabiliyeti olan değişkenler öne çıkarılmıştır. promotion\_last\_5years (%10,9), management, hr, accounting, RandD gibi departman bilgileri ve IT (department) gibi birçok kategori bazlı değişken yüksek ağırlıklar almıştır. Buna karşın satisfaction\_level, number\_project, average\_monthly\_hours gibi sayısal performans değişkenleri düşük puan almıştır. Bu yöntem, özellikle nadir gözlemler veya sınıf dengesizliği içeren kategorik değişkenleri daha fazla ön plana çıkarma eğilimindedir.

Özetlenecek olursa, RAFOW yönteminin ayırım gücünün yüksek ve sınıflandırma başarısının kanıtlanmış değişkenleri öne çıkardığı belirlenmiştir. Eğer analizin amacını veri setindeki bilgi içeriğinin yüksek olması ve birbirinden bağımsız öznelitliklerin ön plana çıkarma olarak belirlenirse, CRITIC yöntemi daha faydalı olabilir. Veri setindeki kategorik değişkenlerin dağılımlarını anlamak, nadir durumları vurgulamak veya bilgi kazanımına odaklanmak isteniyorsa, Entropi yöntemi daha etkili olabilir.

Farklı teknikler ve yaklaşımlara göre değerlendirmeler yapıldıktan sonra, RAFOW yöntemi ile elde edilen ağırlıkların ARAS yöntemine girdi olarak verilerek personellerin sıralanması hedeflenmektedir. Python dili kullanılarak sentetik olarak bu adaylar türetilmiştir. Sarı renkle boyanmış alanlar maliyet yönlü kriterleri temsil etmektedir. Tablo 12’de normalize karar matrisi gösterilmektedir.

Tablo 12

*Beş Anahtar Personele İlişkin Normalize Karar Matrisi*

Personel ID	1	4	2	3	0
average_monthly_hours	1,000	0,674	0,653	0,853	0761
number_project	0,667	0,833	0,667	1,000	0,667
last_evaluation	0851	1,000	0545	0,883	0,589
time_spend_company	0,220	1,000	0,400	0,330	0,670
satisfaction_level	0,858	1,000	0,706	0,475	0,518
Work_accident	0,000	0,000	1,000	0,000	1,000
promotion_last_5years	1,000	0,000	0,000	0,000	0,000
product_mng	0,000	1,000	0,000	0,000	0,000
RandD	0,000	0,000	1,000	0,000	0,000

Tablo 12 (devam)

<b>hr</b>	0,000	0,000	0,000	1,000	0,000
<b>management</b>	0,000	0,000	0,000	0,000	1,000
<b>marketing</b>	0,000	0,000	0,000	0,000	0,000
<b>accounting</b>	0,000	0,000	0,000	0,000	0,000
<b>IT</b>	0,000	0,000	0,000	0,000	0,000
<b>sales</b>	0,000	0,000	0,000	0,000	0,000
<b>support</b>	0,000	0,000	0,000	0,000	0,000
<b>technical</b>	1,000	1,000	0,000	0,000	0,000
<b>low</b>	1,000	1,000	1,000	1,000	1,000
<b>medium</b>	0,000	0,000	1,000	1,000	0,000
<b>high</b>	0,000	0,000	0,000	0,000	1,000

ARAS işlemlerini yapabilmek için Spyder uygulaması kullanılmıştır.

Bu kısmın algoritması Şekil 6'da gösterilmiştir. Personellerin işten ayrılma tahminlerini yapabilmek üzere RAFOW yöntemiyle nesnel olarak elde edilmiş olan kriter ağırlıklar, Python yazılım diliyle sentetik olarak türetilmiş ve anahtar personel olduğu kabul edilmiş 5 personel üzerinde ARAS skorlarının hesaplanmasında kullanılmış olup, adaylar bu bütünlük puanına göre büyükten küçüğe sıralanmıştır. Çalışmanın çıktısı Tablo 13'te gösterilmektedir. Bu sonuçlara göre ID 4 ve ID 1 nolu personellerin şirkette kalmaya meyilli olduğu; ID 0, ID 2 ve ID 3 personelinin işten ayrılmaya yakın olduğu sonucu ortaya çıkmaktadır.

```

1  # -*- coding: utf-8 -*-
2  """
3  Created on Thu Nov  7 12:16:35 2024
4
5  @author: bilen
6  """
7
8  import pandas as pd
9
10 # Creating the weighted normalized matrix dataframe from the provided data
11 data = {
12     "Personel ID": [1, 4, 2, 3, 0],
13     "average_monthly_hours": [0.131032, 0.088183, 0.085547, 0.111703, 0.099714],
14     "number_project": [0.098945, 0.123681, 0.098945, 0.148418, 0.098945],
15     "last_evaluation": [0.103051, 0.121088, 0.066099, 0.107337, 0.071315],
16     "time_spent_company": [0.028182, 0.1281, 0.05124, 0.042273, 0.085827],
17     "satisfaction_level": [0.154836, 0.180698, 0.127561, 0.085092, 0.093898],
18     "Work_accident": [0.000000, 0.000000, 0.068037, 0.000000, 0.068037],
19     "promotion_last_5years": [0.018129, 0.000000, 0.000000, 0.000000, 0.000000],
20     "product_mng": [0.000000, 0.008894, 0.000000, 0.000000, 0.000000],
21     "Rand": [0.000, 0.000, 0.0141, 0.000, 0.000],
22     "hr": [0, 0, 0, 0.0071, 0],
23     "management": [0.000, 0.000, 0.000, 0.000, 0.00499],
24     "marketing": [0.000, 0.000, 0.000, 0.000, 0.000],
25     "accounting": [0.000, 0.000, 0.000, 0.000, 0.000],
26     "IT": [0.000, 0.000, 0.000, 0.000, 0.000],
27     "sales": [0.000, 0.000, 0.000, 0.000, 0.000],
28     "support": [0.000, 0.000, 0.000, 0.000, 0.000],
29     "technical": [0.009937, 0.009937, 0.000000, 0.000000, 0.000000],
30     "low": [5.13, 5.13, 0.0513, 0.0513, 0.0513],
31     "medium": [0.000, 0.000, 0.03399, 0.03399, 0.000],
32     "high": [0.000, 0.000, 0.000, 0.000, 0.03839]
33 }
34 df = pd.DataFrame(data)
35
36 # Determine the ideal solution (reference vector) as the maximum for each criterion
37 ideal_solution = df.drop(columns=["Personel ID"]).max()
38
39 # Calculate the utility degree for each employee by summing their weighted normalized values
40 df["utility_degree"] = df.drop(columns=["Personel ID"]).sum(axis=1)
41
42 # Calculate ARAS score for each employee by dividing by the ideal solution's utility degree
43 total_ideal = ideal_solution.sum()
44 df["ARAS_score"] = df["utility_degree"] / total_ideal

```

Şekil 6. ARAS skorunu hesaplayabilmek için yapılan geliştirmeler.

Tablo 13  
Personellerin ARAS Skorlarının Gösterimi

Personel ID	ARAS Skorları
4	0.958
1	0.938
0	0.101
2	0.098
3	0.097

## Bölüm 7

### RAFOW Destekli Dinamik ARAS

ARAS yöntemi ile makine öğrenmesi tekniklerinin birlikte kullanarak elde edebileceği kazanımlar aşağıdaki şekilde sınıflandırabilir:

- Şeffaflık ve karmaşık veri analizi kabiliyeti.
- Dinamik kriter ağırlıkları elde etme.
- Hibrit karar destek sistemlerinden gelen doğrulama ve güvenilirlik.
- Esneklik ve uyarlanabilirlik.
- Genel performans artışı.

Yazında makine öğrenmesi ve klasik ÇKKV yöntemlerinin birlikte kullanıldığı çalışmalara rastlanmaktadır. Navanitlal (2021), ARAS yöntemini makine öğrenimi ile birleştirerek öğretim üyesi sayısı, öğrenci kabulü ve kampüs alanı gibi çeşitli kriterlere dayalı olarak Hindistan'daki teknik kurumları değerlendirmiştir. Çalışma, karar verme süreçlerini optimize etmek için ağırlıklı toplam modelleri ve makine öğrenimi algoritmalarını kullanarak kapsamlı bir analiz sunmaktadır.

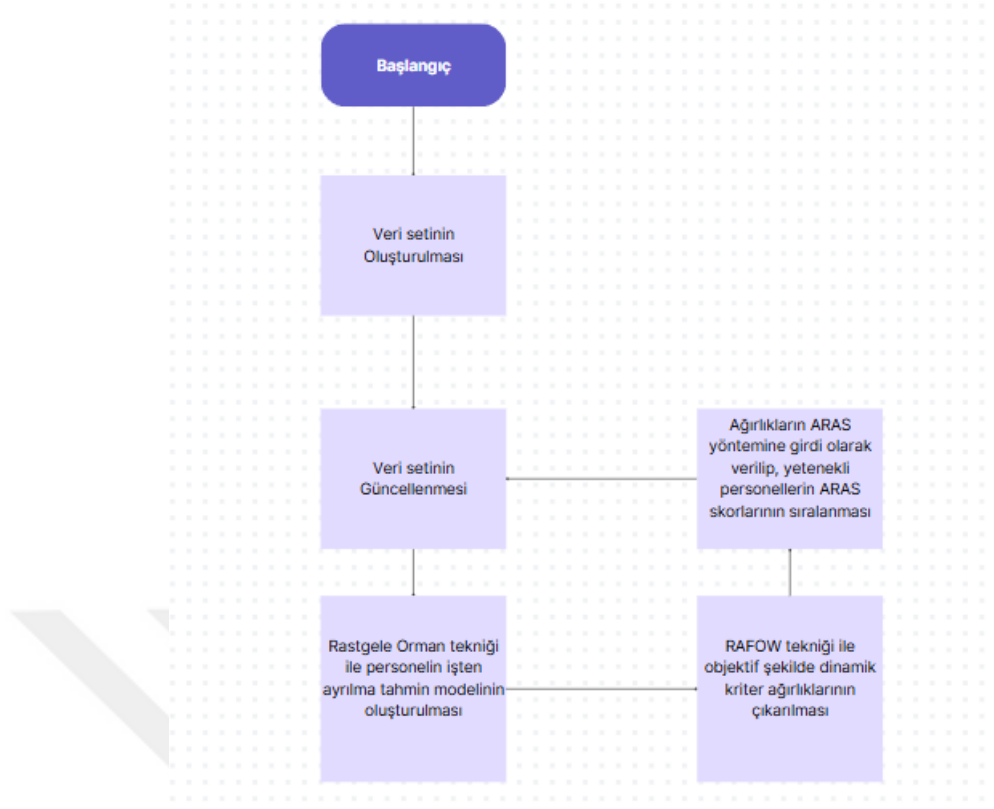
Chaudhary, Gaur, Jhanjhi, Masud ve Aljahdali (2022), geçmişte yapılmış çalışmaların çeşitli makine öğrenimi algoritmaları kullanarak çalışan churn analizine odaklandığını görüp, çalışanları başarılarına göre kategorilendirmenin göz ardı edildiğini farketmişler ve çalışmalarında, çalışanların önemini niceliklendirerek kategorize etmek için bir yaklaşım önermişlerdir. Çalışan başarılarına göre kriter ağırlıklarını atamak için CRITIC ve çalışanları üç kategoriye ayırmak için de MARCOS yöntemini kullanmışlardır. CatBoost, Destek Vektör Makinesi, Karar Ağacı, RO ve XGradient Boost algoritmaları, kategorize edilmiş ve kategorize edilmemiş veri kümeleri üzerinde doğruluk, keskinlik, geri çağırma gibi metrikler ve Matthew Korelasyon Katsayısı (MCC) ile analiz edilerek, kullanılan veri kümesi için en uygun algoritma belirlenmiştir. CatBoost algoritması, kategorize edilmiş çalışanlar için performans ölçütleri açısından en iyi sonuçları vermiştir. Ma ve Li (2024) ise çalışmalarında tedarikçi kalite değerlendirmesi için bir karar destek sistemi geliştirmişlerdir. Burada da, büyük veri çağında makine öğrenimi kavramlarına dayalı bütünleşik bir

ÇKKV modeli önerilmiştir.

Pourkhodabakhsh, Mamoudan ve Bozorgi-Amiri (2022), çalışan devir hızını etkileyen faktörleri belirlemek için kullanılmış olan ÇKKV tekniklerinin performansını izlemek ve çalışan devir hızını etkileyen en önemli faktörleri belirlemek amacıyla bazı makine öğrenimi algoritmalarını kullanmışlardır. Bir şirketin insan kaynakları veri kümesi üzerinde uygulamalar yapılmış ve sonuçlar Mutual Information algoritmasıyla belirlenen faktörlerin, çalışan devir hızını tahmin etmede daha iyi sonuçlar gösterdiğini ortaya koymuştur.

Literatürdeki bunlar gibi çalışmalar incelendiğinde, ÇKKV yöntemlerinin makine öğrenmesi ile birlikte kullanıldığı durumlarda, şeffaflık ve karmaşık veri analizi kabiliyeti, dinamik kriter ağırlıkları elde edilebilmesi, hibrit karar destek sistemlerinden gelen doğrulama ve güvenilirlik, esneklik ve uyarlanabilirlik, genel performans artışı sağlayabilme yetilerinin kazanılabileceği görülmektedir.

Bu çalışmada, RAFOW ile elde edilen öznelik ağırlıkları ARAS yöntemi ile birlikte karar destek sistemini oluşturmuştur. Personel veri setine personel eklenip, çıkarma olduğu takdirde güncel veri seti değişecek ve çok fazla ek işleme gerek kalmadan model dinamik şekilde güncellenebilecektir. Böylelikle RAFOW kriter ağırlıklarını dinamik şekilde değiştirilebilecek ve tekrar ARAS yöntemi ile sıralama yapılabilecektir. Böylelikle insan kaynakları alanında karar vericiler herhangi bir  $t$  anında analitik karar destek sistemini kullanarak güncel ağırlıklarla oluşturulmuş ARAS skorlarını kolaylıkla hesaplayabileceklerdir. Örneğin, A kişisi işten ayrılmaya B kişisinden daha yakinken 6 ay sonra yeni kriter ağırlıklarına göre bu durum tam tersi de şekilde de olabilir. Bu nedenle, ileri sürülen karar destek sistemi dinamik olarak ağırlık güncelleyen bir sistem sunmaktadır. Oluşturulan karar destek sisteminin iş akışı Şekil 7’de gösterilmiştir.



Şekil 7. Karar destek sisteminin modellenmesi.

## Bölüm 8

### Ek Analizler

RO yöntemi üzerine inşa edilen personelin işten ayrılmasının tahminlenmesi sürecinde bir takım deneysel çalışmalar da gerçekleştirilmiştir. Bunlara ilişkin ayrıntılı analiz sonuçları bu bölümün konusunu teşkil etmektedir.

#### 8.1 Sadece Ağırlığı En Yüksek Çıkan 4,5 ve 6 Adet Öznitelik ile Modelin Çalıştırılması

Modelin inşa edilme sürecinde tüm öznitelikler kullanılmış olup, bu ek çalışmada özniteliklerden ağırlığı en yüksek olan 4, 5 ve 6 adet en önemli öznitelik dikkate alınarak bir duyarlılık analizi gerçekleştirilmiştir.

Ağırlığı en yüksek çıkan 4 öznitelik şunlardır: `satisfaction_level`, `salary`, `number_project`, `time_spend_company`. Bu öznitelikler ile model tekrar inşa edildiğinde değerlendirme metrikleri aşağıdaki şekilde oluşmuştur.

$R^2$  değeri 0,819'tür. Bu sonuç, modelin bağımlı değişkendeki toplam değişimin yaklaşık %81,9'ünü açıklayabildiğini göstermektedir. Yani model, veriler üzerinde oldukça iyi bir uyum sağlamaktadır. Benzer şekilde Ayarlanmış  $R^2$  değeri de 0,819 olarak hesaplanmıştır. Bu değerın klasik  $R^2$  ile birebir aynı olması, modele eklenen bağımsız değişkenlerin modelin açıklayıcılığına olumlu katkı sağladığını ve gereksiz/etkisiz değişken içermediğini gösterir. Bu sonuç, modelin kurulumunda doğru öznitelik seçiminin yapıldığını işaret etmektedir. Ortalama mutlak hata değeri 0,102'dur. Bu metrik, modelin tahminlerinde ortalama olarak yaklaşık 0.10 birim sapma yaptığını ifade eder. Ortalama kare hatası ise 0,033'dir. Bu değer, modelin bazı noktalarda hata yapmasına rağmen, büyük sapmaların genelde düşük olduğunu ifade eder. Kareli hatalar uç değerlerden daha fazla etkilendiğinden, MSE'nin düşük çıkması modelin aşırı hatalar yapmadığını göstermektedir.

Ağırlığı en yüksek çıkan 5 öznitelik dikkate alındığından önem sıralaması şu şekilde bulunmaktadır: `satisfaction_level`, `last_evaluation`, `number_project`, `time_spend_company`, `salary`. Bu öznitelikler ile model

tekrar inşa edildiği takdirde değerlendirme metrikleri aşağıdaki şekilde oluşmuştur.

5 parametrelili modelin  $R^2$  değeri 0,881 olarak hesaplanmıştır. Bu, bağımlı değişkendirdeki toplam varyansın %88,1'i model tarafından açıklanıyor demektir. 4 parametrelili modelde bu değer 0,819'du. Buna göre, modele eklenen beşinci parametre, modelin açıklayıcılığını %6,2 artırmıştır. Ayarlanmış  $R^2$  değeri de 0,880 olup,  $R^2$ 'ye hemen hemen aynı bulunmuştur. Bu durum, modele eklenen beşinci değişkenin istatistiksel olarak anlamlı olduğunu ve modelin genellenebilirliğini bozmadığını göstermektedir. Ortalama mutlak hata değeri 0,072 bulunmuştur. Bu değer önceki modelde 0,102'ydi. Yani modelin tahminleri daha da hassas hale gelmiş ve ortalama hata yaklaşık %28 oranında azalmıştır. Ortalama kare hata değeri ise 0,022 olup 4 parametrelili modele göre (0,033) önemli ölçüde düşmüştür. Bu sonuç da büyük hataların daha da azaltıldığını işaret etmektedir.

Ağırlığı en yüksek 6 özneliğin sıralaması şu şekilde tespit edilmiştir: satisfaction\_level, last\_evaluation, number\_project, time\_spend\_company, salary, average\_monthly\_hours. Bu öznelikler ile model tekrar inşa edildiği takdirde değerlendirme metrikleri aşağıdaki şekilde oluşmaktadır.

$R^2$  değeri 0,920 bulunmuş olup modelin bağımlı değişkendirdeki toplam varyansı açıklayıcılığı %92 seviyesine çıkmıştır. Bu değer 4 ve 5 parametrelili modellerden daha iyi bir sonuçtur. Ayarlanmış  $R^2$  de 0,920 bulunmuş olup modele eklenen altıncı parametrenin aşırı uyuma neden olmadığını ve modelin genellenebilirliğini bozmadığını göstermektedir. Ortalama mutlak hata, 0,049 bulunmuş olup oldukça düşük bir değerdedir. Ayrıca ortalama kare hatası da 0,015 ile yine hataların küçük sapmalarla sınırlı kaldığını ve modelin istikrarlı tahminler yapmakta olduğunu ortaya koymaktadır.

Altı parametrelili son model, dört ve beş parametrelili modellerle kıyaslandığında hem hata oranları düşmüş hem de açıklayıcılık düzeyi artmıştır. Bu da modele dahil edilen altıncı parametrenin modelin başarısını artırdığını ve güçlü bir katkı sağladığını göstermektedir.

Bu analiz, parametre sayısı değişikçe modelin açıklanabilirliğinin nasıl değiştiğini gösterebilmek amacıyla yapılmıştır. 4, 5 ve 6 parametre kullanarak oluşturulan tahminleme modellerinin değerlendirme metrikleri Tablo 14'te karşılaştırmalı olarak gösterilmektedir.

Tablo 14

*Farklı Parametre Sayılarına Göre Elde Edilen Değerlendirme Metrikleri*

Parametre Sayısı	R <sup>2</sup>	Düzeltilmiş R <sup>2</sup>	Ortalama Mutlak Hata	Ortalama Kare Hata
4 Parametrelili	0,819	0,819	0,102	0,033
5 Parametrelili	0,881	0,880	0,072	0,022
6 Parametrelili	0,920	0,920	0,049	0,015

Sonuç olarak;

- 6 parametrelili model, tüm doğruluk ve hata kriterlerine göre en iyi performansı sergilemektedir.
- 5 parametrelili model ile 6 parametrelili model arasındaki iyileşme oranı, 4 parametreden 5 parametreye geçişe göre daha düşüktür. Bu durum, azalan getiriler ilkesini yansıtmaktadır.
- Eğer modelin sadelik, yorumlanabilirlik veya aşırı uyum riski göz önünde bulundurulacak olursa, 5 parametrelili model performans ve sadelik dengesi açısından tercih edilebilir.
- Fakat doğruluk en öncelikli kriterse, 6 parametrelili modelin seçilmesi daha uygundur.

Analiz, sınıflandırma problemi olarak ele alıp model tahminlemede kullanıldığında oluşan değerlendirme metrikleri Tablo 15’te verilmektedir.

Tablo 15

*Ek Analiz İçin Değerlendirme Metrikleri*

Parametre Sayısı	Sınıf (1/0)	Doğru Pozitif	Yanlış Pozitif	Doğru Negatif	Yanlış Negatif	Duyarlılık	Kesinlik	Hassaslık	Özgüllük
4 Parametrelili	1	980	59	3369	92	0,914	0,943	0,914	0,983
	0	3369	92	980	59	0,983	0,973	0,983	0,914
5 Parametrelili	1	982	20	3408	90	0,916	0,980	0,916	0,994
	0	3408	90	982	20	0,994	0,974	0,994	0,916
6 Parametrelili	1	1008	8	3420	64	0,940	0,992	0,940	0,998
	0	3420	64	1008	8	0,998	0,982	0,998	0,940

Modelde göz önüne alınan öznelik sayısı arttıkça genel performansın iyileştiği görülmektedir. 4 parametrelili modelde birinci sınıfa ait duyarlılık %91,4 iken bu oran 6 parametrelili modelde %94’e yükselmiş, kesinlik ise %94,3’ten %99,2’ye çıkmıştır. Benzer şekilde özgüllük ve hassaslık da

artarak 6 parametrelili modelde %99,8 ve %94'e ulaşmıştır. Yanlış pozitif ve yanlış negatif sayılarında da düzenli bir azalma olduğu dikkate alındığında, modelin parametre sayısındaki artış doğruluk ve denge açısından olumlu bir etki yaratmıştır. Doğruluk oranları ise sırasıyla 0,966; 0,976 ve 0,984 olarak bulunmuştur. Bu sonuçlar, modelin daha fazla öznitelikle daha doğru ve güvenilir sınıflandırmalar yaptığını göstermektedir.

Parametre sayısı arttıkça modelin doğruluk oranı artmakta ve hata değerleri azalmaktadır. Ancak artan karmaşıklığın getirdiği ek katkı azalmaktadır. Bu nedenle uygulama amacına göre 5 veya 6 parametrelili modelin kullanılması önerilebilir.

## 8.2 Personel Sayısının Değişmesi Durumu

Personel sayısı artıp, azaldığı takdirde özniteliklerin ağırlıkları değişebilecektir. Ağırlıkların değişebildiğini görebilmek amacıyla 14.999 verinin bulunduğu veri setinden rastgele 10.000 veri seçilerek model tekrar inşa edilmiştir ve özniteliklerin ağırlıklarının değişip değişmediği görülmek istenmiştir. Tablo 16'da öznitelik ağırlıklarının değişimi gösterilmektedir.

Tablo 16

### *Personel Sayısının Değişmesi Durumunda Ağırlıkların Değişimi*

Öznitelikler	14.999 Personel için Ağırlıklar	10.000 Personel için Ağırlıklar
satisfaction level	0.183	0.189
last_evaluation	0.121	0.103
number_project	0.160	0.143
average_monthly_hours	0.117	0.141
time_spend_company	0.131	0.121
Work_accident	0.075	0.053
promotion_last_5years	0.018	0.025
sales	0.003	0.009
accounting	0.001	0.009
hr	0.008	0.019
technical	0.004	0.008
support	0.001	0.012
management	0.002	0.011
IT	0.012	0.002
product_mng	0.004	0.009
marketing	0.005	0.013
RandD	0.017	0.008
low	0.052	0.039
medium	0.031	0.024
high	0.044	0.052

Bu sonuçlar, veri setinin değişmesi durumunda RAFOW ile elde edilen ağırlıkların dinamik şekilde değişeceğini göstermektedir. Böylelikle dinamik olarak ağırlıkların güncellendiği ve ARAS yöntemine sıralama yapması için girildi oluşturduğu bir karar destek sisteminin önerilebileceği fikri desteklenmektedir.

### 8.3 Yeni Bir Veri Setinde Yöntemin Uygulanması

Bu çalışmanın dışında açık kaynaklardan elde edilen başka bir şirketin personel verileri üzerinde de personelin işten ayrılma tahmini yapılmıştır. RAFOW yöntemi ile daha önce ileri sürülen ağırlıkların daha homojen dağılması ve saflılık ilkesine göre belirlenen ağırlıklardan daha dengeli bir yapı oluşup oluşmadığının tespit edilmesi hedeflenmiştir. Tablo 17’de hem saflılık temelli yaklaşım hem de RAFOW yaklaşımı ile belirlenen ölçüt önemleri gösterilmektedir.

Tablo 17

#### *Yeni Veri Seti İçin Bulunan Ağırlıkların Karşılaştırması*

Öznitelikler	Saflılık Tabanlı Özellik Önemi	RAFOW ile Özellik Önemi
PHD (Education)	0,005	0,023
Bachelors (Education)	0,024	0,101
EverBenched	0,029	0,056
Masters (Education)	0,051	0,097
Gender	0,052	0,129
ExperienceInCurrentDomain	0,114	0,066
PaymentTier	0,172	0,220
Age	0,175	0,079
Number Of CurrentFirm Experience	0,372	0,229

Buna göre, RAFOW yöntemi özellikler arasında daha homojen ve dengeli bir önem dağılımı sunmuştur. Gini tabanlı saflılık odaklı yaklaşımda “Number Of CurrentFirm Experience” özneliğinin ağırlığı %37,2 iken RAFOW bu ağırlığı %22,9 seviyesine çekmiş ve diğer özelliklere de anlamlı oranlar tanımıştır. Bu sayede modelin tek bir özelliğe bağımlı kalmadan karar verebilme kapasitesi artırılmış olmaktadır. Yine saflılık temelli yaklaşımda neredeyse ortaddan kaldırılması önerilen olan PHD (%0,5), RAFOW’da %2,3 gibi anlamlı bir seviyede çıkmıştır.

Benzer şekilde, temel yaklaşımda “Number Of Current Firm Experience” ve “Age” gibi iki özellik toplam önemin %54’ünden fazlasını temsil etmekte, yani model bu iki özelliğe aşırı bağımlı hale gelmektedir. RAFOW ağırlıkları daha dengeli dağıtarak özellikler arası çeşitliliği artırmış ve modelin genellenebilirliğini iyileştirmiştir. RAFOW yöntemi, karar verme sürecinde sadece birkaç dominant özelliğe değil, daha fazla sayıda özelliğe makul katkı payları tanımladığı için özellikle insan kaynakları gibi çok boyutlu sorunlarda daha açıklanabilir ve adil sonuçlar sağlayabilir.

#### **8.4 SHAP Yöntemi ile Ağırlıklandırma Analizi**

RO algoritması, çok sayıda karar ağacı oluşturarak veri kümesi üzerinde güçlü tahmin performansı sunan bir topluluk öğrenme yöntemidir. SHAP (SHapley Additive exPlanations) yöntemi, bir makine öğrenmesi modelinin çıktılarına katkıda bulunan özniteliklerin etkisini, oyun teorisine dayalı olarak sayısal şekilde ifade eder. Her bir özelliğin modele yaptığı katkı, Shapley değerleri üzerinden hesaplanarak açıklanabilirlik sağlanır (Lundberg ve Lee, 2017). Bu yöntem, özellikle "black-box" modellerin kararlarını yorumlamak için hem teorik olarak tutarlı hem de yerel doğruluk sağlayan bir açıklama çerçevesi sunar. Ancak bu güçlü tahmin gücüne rağmen modelin karar verme süreci genellikle bir “kara kutu” gibi anlaşılmaktadır. Bu durum, özellikle kararların nasıl ve ne gerekçe ile o şekilde alındığını anlamak isteyen uzmanlar ve yöneticiler için sınırlayıcı olarak değerlendirilmektedir. Bu noktada, SHapley Additive exPlanations (SHAP) yöntemi, modelin her bir tahmininde hangi özniteliklerin ne kadar etkili olduğunu açık ve nicel şekilde ortaya koyabilmektedir. SHAP’ın temel algoritması, katkıların hesaplanması sırasında tüm olası özellik sıralamalarını dikkate alır; bu da açıklamanın lokal doğruluğunu sağlar (Štrumbelj & Kononenko, 2014).

SHAP yöntemi, her bir özneliğin model çıktısına katkısını oyun teorisi temelli bir yaklaşımla hesaplar. Bu sayede, klasik önem belirleme yöntemlerinin aksine sadece genel önem sıralaması değil, her bir örnek için özniteliklerin tahmin üzerindeki bireysel katkılarını da gösterir. SHAP algoritması, kooperatif oyun teorisindeki Shapley değerlerinin tüm öznitelik

kombinasyonları üzerindeki marjinal katkılarının ortalamasını alarak çalışır; bu işlem her bir özneliğin modele olan katkısını adil ve tutarlı biçimde ölçer (Lundberg ve Lee, 2017). Rastgele orman gibi çok sayıda ağacın oluşturduğu karmaşık yapılar içinde SHAP, doğrusal olmayan etkileşimleri ve karşılıklı ilişkileri de hesaba katarak daha güvenilir ve yorumlanabilir sonuçlar sunabilir.

Modelin SHAP ile yorumlanabilir hale gelmesi sadece akademik ya da teknik açıdan değil, pratik karar alma süreçlerinde de büyük avantaj sağlayabilir. Örneğin, bir çalışanın işten ayrılma ihtimali üzerine kurulan bir RO modeli kullanılıyorsa SHAP sayesinde bu çalışanın neden bu şekilde değerlendirildiği (örneğin memnuniyet düzeyinin, maaşının ya da geçirdiği iş kazalarının katkısı) sayısal olarak ortaya konulabilir. Bu da hem insan kaynakları yöneticilerine hem de model geliştiricilere daha isabetli ve güvenilir kararlar alma imkânı sunabilir.

SHAP yöntemi, RO gibi güçlü fakat yorumlanması zor modelleri açıklanabilir hale getirerek veri bilimi uygulamalarının güvenilirliğini artırır. Sadece modelin ne yaptığı değil, neden yaptığı da açıklandığından dolayı karar destek sistemleri çok daha şeffaf ve hesap verebilir hale gelir. Bu durum özellikle yüksek riskli veya etik boyutları olan karar süreçlerinde model tabanlı analizlerin kabul edilebilirliğini büyük ölçüde artırma potansiyeli taşımaktadır. Tablo 18’de SHAP yöntemiyle oluşturulmuş ağırlıklar gösterilmektedir.

Tablo 18

*SHAP Yöntemiyle Elde Edilen Ağırlıklar*

Öznitelikler	SHAP Yöntemiyle Öznitelik Ağırlıkları
satisfaction level	0.429
time spend company	0.157
number project	0.152
last evaluation	0.105
average montly hours	0.099
Work accident	0.001
salary	0.009
department	0.048

Personelin işten ayrılma tahminlemede kullanılan özneliklerin ağırlıkları, RAFOW yöntemi, saflılık temelli yaklaşım ve SHAP yöntemiyle

karşılaştırıldığında bir takım benzerlikler ve farklılıklar ortaya çıkmaktadır. En dikkat çeken özniteliklerden biri olan çalışan memnuniyeti düzeyi (satisfaction\_level), üç yöntem tarafından da en önemli faktör olarak değerlendirilmiştir. Ancak bu değişkenin önemi, saflılık temelli yaklaşımda %50,0 ve SHAP yaklaşımında %42,9 gibi yüksek seviyelerdeyken, RAFOW yöntemi bu özneliğe daha dengeli bir ağırlık %18,3 vermiş ve diğer değişkenlerin katkısını önemsizleştirmemiştir. Bu durum, RAFOW'un çok boyutlu analiz anlayışını ve tek bir değişkene aşırı bağımlılığı azaltma amacını yansıtmaktadır.

Performans, iş yükü ve kıdemle ilgili değişkenler olan number\_project, last\_evaluation, average\_monthly\_hours ve time\_spend\_company, üç yöntem tarafından da orta-yüksek derecede önemli olarak değerlendirilmiştir. Özellikle SHAP ve RAFOW yöntemleri bu özniteliklere benzer ağırlıklar vererek çalışan davranışlarını açıklamada bu kriterlerin belirleyici olduğunu ortaya koymuştur. Örneğin, SHAP ve RAFOW'a göre number\_project değeri sırasıyla %15,2 ve %16,0 iken, saflılık temelli yaklaşım bu özneliğin ağırlığını görece olarak daha düşük bulmuştur (%10,5). Bu da Gini'nin bazı önemli değişkenlerin etkisini yeterince yansıtamayabileceğini göstermektedir.

Daha az belirleyici görülen ancak istisnai durumlarda etkili olabilecek work\_accident ve promotion\_last\_5years gibi değişkenlerde ise dikkat çekici bir fark vardır. RAFOW bu özniteliklere sırasıyla %7,5 ve %1,8 gibi belirli bir önem atfederken, Saflılık temelli yaklaşım sırasıyla %0,2 ve %0,01; SHAP ise ikisine de eşit şekilde %0,1 ağırlık vermiş ve bu kriterleri neredeyse ihmal edilebilir seviyede göstermiştir. Bu durum, RAFOW'un nadir ama anlamlı etkileri göz ardı etmemeye çalışan daha dengeli ve kapsamlı bir yaklaşımı benimsediğini ortaya koymaktadır.

Maaş düzeyi (salary) ve departmanlar (departments) değişkenlerinde de belirgin farklılıklar gözlemlenmiştir. RAFOW maaş değişkenine %12,7; Gini %0,67 ve SHAP ise %0,9 önem atfetmiştir. Burada, RAFOW insan kaynakları perspektifinden maaş gibi sosyoekonomik etmenleri daha ciddi bir belirleyici olarak değerlendirmiştir. Benzer şekilde, departman değişkeni de RAFOW'da %5,7 gibi anlamlı bir öneme sahip olarak bulunmuş iken diğer yöntemler bu faktörü çok daha az önemli olarak

değerlendirmektedirler.

Sonuç olarak, RAFOW yönteminin öznitelik ağırlıklarını belirlerken daha dengeli, çok boyutlu ve bağlamsal etmenleri dikkate alan bir yaklaşım sunduğu değerlendirilirken, SHAP yönteminin bireysel tahminler üzerindeki katkıyı hassas şekilde ölçerek özniteliklerin etkisini daha doğrudan ve matematiksel olarak açıklamakta olduğu anlaşılmaktadır. Safılık temelli ağırlıklandırma yaklaşımı ise bazı önemli değişkenleri abartılı biçimde ön plana çıkarabilirken, diğer bazı etkili değişkenleri de göz ardı edebilmektedir. Bu analiz aynı zamanda, karar vericilerin hangi yöntemi tercih edeceklerini uygulama bağlamına göre belirlemeleri gerektiğini açıkça göstermektedir.



## Bölüm 9

### Tartışma ve Sonuçlar

Bu çalışmada temel olarak RO makine öğrenmesi algoritması temelli dinamik ve nesnel bir ölçüt ağırlıklandırma yaklaşımı önerilmektedir. RAFOW ismi verilen bu yaklaşım RO, algoritmasının doğasında bulunan bir özneliği dallanma için aday gösterilmesi ve seçilmesi bileşenlerinin oranı üzerinden bir önem ağırlığı belirleme ilkesi oluşturmaktadır. Önerilen RAFOW yaklaşımı, açık kaynak bir insan kaynakları veri seti üzerinde uygulanmış ve başarı parametreleri yorumlanmıştır.

Temelde personelin işten ayrılma durumunun tahmin edilmeye çalışıldığı insan kaynakları analitiği probleminde üç farklı makine öğrenmesi algoritması (RO, Gradyan Artırma ve Ekstrem Gradyan Artırma) uygulanmış ve model performansları çeşitli değerlendirme metrikleri temel alınarak karşılaştırılmıştır. Burada RO yaklaşımının temel olarak  $R^2$ , düzeltilmiş  $R^2$  ve ortalama mutlak hata metrikleri açısından diğer iki algoritmadan daha iyi performans gösterdiği, sadece hata kareleri ortalamasının karekökü göstergesinde kısmen daha kötü performans gösterdiği gözlemlenmiştir. Burada hareketle RAFOW yaklaşımının RO üzerine temellendirilmesinin daha doğru olduğu değerlendirilmiştir. En yüksek doğruluğu ve en düşük tahmin hatasını sunan RO algoritması burada tercih edilmiş olmakla birlikte gelecek çalışmalarda daha dengeli bir hata dağılımı sunabilecek olan Ekstrem Gradyan Artırma gibi bir teknik üzerinde de benzer bir yaklaşım geliştirilmesi olanaklıdır.

Performans metriklerine göre yaklaşım tespit edildikten sonra yeni bir nesnel ağırlıklandırma yaklaşımı olan RAFOW tekniğinin çalışma prensipleri oluşturulmuştur. Daha sonra eldeki veri setine hem RAFOW hem saflılık temelli ağırlıklandırma yaklaşımı hem de SHAP algoritması uygulanmış ve üç farklı ölçüt ağırlığı seti oluşturulmuştur.

Saflılık temelli öznelik önemi belirleme, özellikle karar ağacı modellerinde sıkça kullanılan ve modelin doğruluğunu artıran özneliklerin önem düzeyini ölçen bir anlayışa sahiptir. Burada satisfaction\_level ile gösterilen çalışan memnuniyeti düzeyi %50 gibi uç bir ağırlık kazanarak ön

plana çıkmıştır. Buna göre, çalışanların memnuniyet düzeyinin işten ayrılma kararında en kritik ölçüt olduğu anlaşılmıştır. Benzer şekilde, time\_spend\_company (şirkette çalışma süresi) ve number\_project (yürütülen proje sayısı) öznitelikleri de sırasıyla %14,4 ve %10,5 ağırlıklarla işten ayrılma kararını etkileyen önemli faktörler olarak öne çıkmıştır. RAFOW yöntemiyle öznitelik önemleri analiz edildiğinde yine satisfaction\_level %18,3 ile en önemli kriter olarak başı çekmiştir. Arkasından gelen number\_project ölçütü %16; time\_spend\_company ise %13,1 ağırlık alarak en önemli diğer iki ölçüt olarak belirlenmiştir. Genel bir değerlendirme yapıldığında, RAFOW yaklaşımının kriter önemleri konusunda daha homojen bir dağılım önermesi, bu nedenle de kararların değerlendirilmesinde daha geniş bir perspektifin göz önüne alınması gerektiğini vurgulaması dikkat çekicidir.

Çalışmanın üçüncü aşamasında, RAFOW yönteminin ÇKKV problemlerinde alternatiflerin sıralanması ve incelenmesinde nasıl kullanılabileceğinin bir örneği olarak ARAS yaklaşımıyla bütünleşik kullanımı örneklendirilmiştir. Böylelikle, RAFOW-ARAS bütünleşik yaklaşımı oluşturulmuş olmaktadır. Bu yapılırken sadece statik bir durum değerlendirmesiyle yetinilmemiş, modelin ne şekilde dinamik bir yapıyla karar desteği sunacağı da vurgulanmıştır. Buna göre, kullanılan veri setinde bir değişiklik olduğunda, yani veri setine herhangi bir personel eklendiği veya personel işten ayrıldığı durumda ağırlıkların güncellenmesinin nasıl olacağı da çalışma içerisinde ayrıntılandırılmıştır. Güncellenen RAFOW ağırlık setinin ARAS sıralamasını da otomatik olarak güncelleyeceği şekilde bütünleşik bir model kurgusu bu tez kapsamında önerilmektedir. Burada ARAS bir bütünleştirme ve uygulama örneği olarak özellikle de matematiksel yapısının kuvvetli ancak basitçe anlaşılabilir olması, bu nedenle alan uzmanlığı olmayan karar analistleri tarafından da rahatça kullanılabileceği düşünülmüş ve seçilmiştir. Gelecek çalışmalarda TOPSIS, PROMETHEE, MARCOS gibi çok daha farklı ve komplike karar verme yaklaşımlarıyla da bütünleştirme işlemleri yapılabilir ve sonuçlar karşılaştırmalı şekilde özetlenebilir.

RAFOW-ARAS bütünleşik modelinin uygulaması Python yazılım diliyle oluşturulmuş beş sentetik personelin işten ayrılma eğilimlerinin

irdelenmesi üzerinde gösterilmiştir. Örneğin Personel ID 4, 0,958 ARAS skoruyla en başı çekmiştir. Burada, en yüksek ARAS skoruna sahip olan bu personelin, işten ayrılma olasılığı en düşük olan personel olduğu değerlendirilmektedir. Yüksek skor, çalışanın işine ve şirketine bağlılık düzeyinin genel çerçevede yüksek olduğunu ve işten ayrılma riskinin en düşük seviyede bulunduğu işaret etmektedir. Uygulamada, bu personelin işten ayrılma riski düşük olduğundan, şirkete olan bağlılığının devam etmesi için mevcut memnuniyet düzeyini korumaya yönelik önlemler alınması düşünülebilir. Bununla bağlantılı olarak, motivasyon ve bağlılığı artırıcı bir ödüllendirme ya da performans değerlendirme sistemleriyle bu durum sürdürülebilir hale getirilebilir. Maaşı düşük fakat memnuniyeti yüksek bir çalışanın işten ayrılma eğilimi göstermiyor oluşu aynı zamanda, iş tatmininin sadece maaş gibi finansal unsurlarla değil, daha geniş bir memnuniyet çerçevesiyle şekillendiğini de göstermektedir. Ayrılma eğilimi en yüksek olan personel ise ARAS skoru 0,097 olan personel ID 3'tür. Bu sonuçtan hareketle, çalışanın mevcut durumundan memnun olmadığı ve işten ayrılma eğiliminin bu nedenle güçlü olduğu anlamı çıkarılabilir. Burada çalışan ile ilgili faktörlerin çoğu çalışan tarafından kötü seviyede olabileceği gibi bu şekilde algolanıyor da olabilir. Eğer bu personel kilit bir önem taşıyorsa, şirket içerisinde kalması için hızlı bir aksiyon alınması gerekliliği bu durumun doğal bir sonucu olacaktır. Çalışanın işten ayrılma sebeplerini anlamak için bire bir görüşmeler yapılması ve çalışana şirkette tutmak için ne gibi düzenlemelerin yapılabileceğinin incelenmesi konusunda bir takım göstergeler bu analiz sonucunda elde edilebilir; maaşta iyileştirme, iş yükünü azaltma veya kariyer gelişimi fırsatlarının sağlanması gibi seçenekler bu kilit personelin korunması için uygulanabilir. Bu gibi değerlendirmeler, çalışanların işten ayrılma risklerini analiz etmek ve şirkete olan bağlılıklarını artırmak için ne gibi stratejik kararların alınmasının gerektiği konusunda faydalı öneriler getirebilir.

Kilit veya önemli yararlıklar sağlayan personellerin işten ayrılmasını önleyici tedbirler almak amacıyla karar destek sistemi (KDS) olarak nitelendirilebilecek, özellikle RAFOW gibi makine öğrenmesi temelli insan kaynakları analitiği çalışmaları şirketlerin insan kaynakları yönetiminde önemli avantajlar sağlayabilir.

1. Yüksek işten ayrılma oranları, şirkette kalan uzman personel üzerindeki iş yükü ve sorumlulukları arttırabilir. Bu da çalışan motivasyonunu ve verimliliğini olumsuz etkileyecektir. Çalışanların işten ayrılma nedenleri analiz edilip farklı durumlar için önleyici tedbirler almak, iş gücü verimliliğini arttırmaya yardımcı olacaktır.

2. Yeni çalışanların ayrılan personellerin yerine işe alınmasına bağlı olarak yükselen işgücü devir oranları, personelin eğitim ve adaptasyon süreçlerinde artan bir maliyet yaratmaktadır. Gereksiz işten ayrılmaların azalması, işe alım ve eğitim masraflarının da azalmasına neden olacaktır.

3. KDS, geçmiş veriler ve analitik modeller kullanarak işten ayrılma olasılığı yüksek olan çalışan gruplarını belirleyebilir. Bu sayede şirketler, bu çalışanlar için özel eğitimler, kariyer geliştirme fırsatları veya esnek çalışma düzenlemeleri gibi önlemler alarak işten ayrılma oranlarını düşürmeye çalışabilir.

4. KDS'ler, çalışan bağlılığını ve işten ayrılma risklerini etkileyen çeşitli faktörleri dikkate alarak karar vericilere daha sağlıklı ve nesnel bir perspektif sunar. Örneğin, çalışan memnuniyeti, kariyer gelişim fırsatları ve iş-yaşam dengesi gibi konularda veri tabanlı öngörüler sağlanarak insan kaynakları stratejilerinin iyileştirilmesine yardımcı olunabilir.

5. Çalışanların bağlılık düzeyini arttırmaya yönelik stratejiler belirlenebilir. Çalışanların kişisel ve profesyonel ihtiyaçlarına yönelik politikalar geliştirilebilir ve kilit personelin işletmeye bağlılığı arttırılabilir. Bu da uzun vadeli başarıyı destekleyecektir.

Teknik kısımda, RAFOW yönteminin daha dengeli bir veri seti sunduğunu gösterebilmek amacıyla ikinci bir veri setinde daha uygulama yapılmış ve RAFOW yöntemi ile daha dengeli ağırlık dağılımı, azımsanan özneliklerin öne çıkarılması, aşırı baskın etkilerin törpülenmesi, karar sürecinin açıklanabilirliğinin artması gibi avantajlar sunulduğu matematiksel olarak da ortaya konulmuştur.

Çalışmada ortaya koyulan RAFOW tabanlı ARAS karar destek sistemi gibi analitik süreçler insan kaynaklarında yetenek yönetimi yapmak isteyen yöneticilerin karar verme süreçlerini kolaylaştıracak; yetenekli görülen personellerin işten ayrılma durumları önceden tespit edilebilecek ve böylelikle stratejik insan kaynakları kararları alabilmenin önü açılmış

olacaktır. Ayrıca, RAFOW tekniđi ile yeni ve nesnel bir kriter ađırlıklandırma yöntemi olarak literatürde yerini alacaktır. ÇKKV'de kriterler arasında ilişki olup olmadığın belirlenmesi karar vericinin bir sorumluluđu olup bazı durumlarda bu tespiti yapmak oldukça zor olabilmektedir. RAFOW yaklaşımında bu gibi ön tespitlerin yapılması zorunlulukları karar vericinin üstünden kalkmakta, karar desteđi daha otomatize bir hale dönüştürülmüş olmaktadır.

Buna ek olarak, daha homojen ve dengeli bir ađırlıklandırma ile makine öğrenmesi modeli üzerinden ek bir prosedür uygulanmasına gerek kalmaksızın ađırlıkların çıkarımı sağlanmış olmaktadır. İnsan kaynakları alanında görev yapan karar vericilere, sadece kompleks makine öğrenmesi modellerinden elde edilen çıktılar önermek yerine, öznel ađırlıklarını makine öğrenmesinden tedarik edip uygulanması oldukça pratik ve anlaşılır olan ARAS yaklaşımına dayanan bir KDS sunulması, çalışmanın ortaya çıkardığı temel uygulama çıktısıdır.

## KAYNAKÇA

- Abdulsalam, S. O., Ajao, J. F., Balogun, B. F. ve Arowolo, M. O. (2022). A churn prediction system for telecommunication company using random forest and convolution neural network algorithms. *EAI Endorsed Transactions on Mobile Communications and Applications*, 7(21).
- Ada, M. ve Çakır, H. (2022). TOPSIS ve AHP çok kriterli karar verme yöntemlerinin personel seçim sürecine uygulanması. *International Journal of 3D Printing Technology and Digital Industry*, 6(2), 186-200.
- Agarwal, S., Bhardwaj, C., Gatkamani, G., Gururaj, R., Darapaneni, N. ve Paduri, A. R. (2023). AI based employee attrition prediction tool. In *Lecture Notes in Computer Science*, 580–588.
- Aguinis, H., Joo, H., ve Gottfredson, R. K. (2013). What monetary rewards can and cannot do: How to show employees the money. *Business Horizons*, 56(2), 241–249.
- Akçakaya, O. ve Akçakaya, E. D. U. (2019). Türkiye'deki büyükşehirlerin çevresel performanslarının Entropi temelli COPRAS ve ARAS yöntemleri ile değerlendirilmesi. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 11(18), 1437-1473.
- Akkermans, J. ve Tims, M. (2017). Crafting your career: How career competencies relate to career success via job crafting. *Applied Psychology*, 66(1), 168–195.
- Alam, M. M., Mohiuddin, K., Islam, M. K., Hassan, M., Hoque, M. A. ve Allayear, S. M. (2018). A machine learning approach to analyze and reduce features to a significant number for employee's turn over prediction model. *Advances in Intelligent Systems and Computing*. Vol:2, 142–159.
- Alao, D. O. ve Adeyemo, A. B. (2013). Analyzing employee attrition using decision tree algorithms. 4(1).
- Alao, A. A. ve Adeyemo, S. A. (2013). An empirical study of factors influencing employee turnover in IT firms. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, 7(12), 2874-2879.
- Alao, D. ve Adeyemo, A. B. (2013). Analyzing employee attrition using decision tree algorithms. *Computing, Information Systems, Development Informatics and Allied Research Journal*, 4(1), 17–28.
- Atef, M., Elzanfaly, D. ve Ouf, S. (2022). Early prediction of employee turnover using machine learning algorithms. *International Journal of Electrical and Computer Engineering Systems*, 13(2), 135–144.
- Alamsyah, A. ve Salma, N. (2018). Employee turnover prediction using machine learning algorithms. *International Journal of Advanced Computer Science and Applications*, 9(2), 1-10.
- Alduayj, S. S. ve Rajpoot, K. (2018). Predicting employee attrition using machine learning. 2018 International Conference on Innovations in Information Technology (IIT), 2018, 93–98.
- Allen, T. D., Golden, T. D. ve Shockley, K. M. (2015). How effective is telecommuting? Assessing the status of our scientific findings. *Psychological Science in the Public Interest*, 16(2), 40–68.

- Allen, T. D., Johnson, R. C., Kiburz, K. M. ve Shockley, K. M. (2012). Work–family conflict and flexible work arrangements: Deconstructing flexibility. *Personnel Psychology*, 66(2), 345–376.
- Alpaydin, E. (2020). *Introduction to machine learning*. Cambridge, MA:MIT Press.
- Altıntaş, F.F. (2023). G20 Grubu Ülkelerin İş Yapma Kolaylığı Performanslarının Analizi: Standart Sapma Tabanlı Aras Yöntemi ile Bir Uygulama. *Malatya Turgut Özal Üniversitesi İşletme ve Yönetim Bilimleri Dergisi*, 4(1), 1-21.
- Aras, G. ve Yıldırım, F. M. (2020). Sosyo-Ekonomik refah düzeyinin belirlenmesinde alternatif bir endeks çalışması: ARAS yöntemi ile G-20 ülkeleri uygulaması. *Business and Economics Research Journal*, 11(3), 735-751.
- Atağan, G. (2013). İşletme Kısıtlarının Aşılmasında Üretme-Satın Alma Karar Modeli Önerisi (Kalite Maliyetleri Yaklaşımı), *Altın Nokta Yayınevi*.
- Atıcı, K. ve Ulucan, A. (2009). Enerji Projelerinin Değerlendirilmesi Sürecinde Çok Kriterli Karar Verme Yaklaşımları ve Türkiye Uygulamaları. *Hacettepe Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 27(1), 161-186.
- Avey, J. B., Luthans, F., Smith, R. M. ve Palmer, N. F. (2010). Impact of positive psychological capital on employee well-being over time. *Journal of Occupational Health Psychology*, 15(1), 17–28.
- Aydoğdu, A. ve Gül, S. (2022). New entropy propositions for interval-valued spherical fuzzy sets and their usage in an extension of ARAS (ARAS-IVSFS). *Expert Systems*, 39(4), e12898.
- Baker, R. S. J. d. ve Inventado, P. S. (2014). Educational data mining and learning analytics. In *Learning analytics*, Springer, 61-75.
- Bakker, A. B. ve Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, 22(3), 273–285.
- Baltes, B. B., Briggs, T. E., Huff, J. W., Wright, J. A. ve Neuman, G. A. (1999). Flexible and compressed workweek schedules: A meta-analysis of their effects on work-related criteria. *Journal of Applied Psychology*, 84(4), 496–513.
- Belli, E. ve Vantini, S. (2020). Measure inducing classification and regression trees for functional data. *arXiv preprint arXiv:2011.00046*.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Boxall, P. ve Purcell, J. (2016). *Strategy and Human Resource Management* (4th ed.). Palgrave Macmillan.
- Bozuyula, M. (2021). AdaBoost Ensemble Learning on top of Naive Bayes Algorithm to Discriminate Fake and Genuine News from Social Media. *Avrupa Bilim ve Teknoloji Dergisi*, (23), 459–462.
- Böckerman, P. ve Ilmakunnas, P. (2009). Job disamenities, job satisfaction, quit intentions, and actual separations: Putting the pieces together. *Industrial Relations: A Journal of Economy and Society*, 48(1), 73-96.
- Braun, S., Peus, C., Weisweiler, S. ve Frey, D. (2013). Transformational leadership, job satisfaction, and team performance: A multilevel mediation model of trust. *The Leadership Quarterly*, 24(1), 270–283.
- Sparrow, P., Brewster, C. ve Chung, C. (2016). *Globalizing Human Resource Management*

(2nd ed.). Routledge.

- Brough, P., Timms, C., O'Driscoll, M. P., Kalliath, T., Siu, O. L., Sit, C. ve Lo, D. (2014). Work-life balance: A longitudinal evaluation of a new measure across Australia and New Zealand workers. *International Journal of Human Resource Management*, 25(19), 2724-2744.
- Brynjolfsson, E. ve McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton & Company.
- Burgess, S., Lane, J. ve Stevens, D. (2001). Job flows, worker flows, and churning. *Journal of Labor Economics*, 18(3), 482-500.
- Büyüközkan, G. ve Göçer, F. (2018). Digital supply chain: Literature review and a proposed framework for future research. *Computers in Industry*, 97, 157-177.
- Cascio, W. F. ve Montealegre, R. (2016). How technology is changing work and organizations. *Annual Review of Organizational Psychology and Organizational Behavior*, 3(1), 349-375.
- Chaudhary M., Gaur L., Jhanjhi NZ., Masud M. ve Aljahdali S. (2022). Envisaging Employee Churn Using MCDM and Machine Learning. *Intelligent Automation & Soft Computing*. 33(2), 1009-1024.
- Chen, L. ve Chen, Y. (2023). Research on Decision Tree Algorithm in Colleges and Universities Financial Management. In *Proceedings of the 2023 3rd International Conference on Education, Information Management and Service Science (EIMSS 2023)* (pp. 49-55). Atlantis Press.
- Chen, T. ve Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Chen, G., Ployhart, R. E., Thomas, H. C., Anderson, N. ve Bliese, P. D. (2017). The power of momentum: A new model of dynamic relationships between job satisfaction change and turnover intentions. *Academy of Management Journal*, 54(1), 159-181.
- Chen, Y., Lin, X. ve Zhan, K. (2023). Employee attrition classification model based on stacking algorithm. *Psychology Research*, 13(6), 279-285.
- Alamsyah, A. ve Salma, N. (2018). A comparative study of employee churn prediction model. *2018 4th International Conference on Science and Technology (ICST)*, 1-4.
- Christian, M. S., Garza, A. S. ve Slaughter, J. E. (2011). Work engagement: A quantitative review and test of its relations with task and contextual performance. *Personnel Psychology*, 64(1), 89-136.
- Clark, A. (2001). What really matters in a job? Hedonic measurement using quit data. *Labour Economics*, 8(2), 223-242.
- Cohen, J., Cohen, P., West, S. G. ve Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.
- Cottini, E., Kato, T. ve Westergaard-Nielsen, N. (2011). Adverse workplace conditions, high-involvement work practices and labor turnover: Evidence from Danish linked employer-employee data. *Labour Economics*, 18(6), 872-880.
- Dağdeviren, M., Akay, D. ve Kurt, M. (2013). İŞ DEĞERLENDİRME SÜRECİNDE ANALİTİK HİYERARŞİ PROSESİ VE UYGULAMASI. *Gazi Üniversitesi*

Mühendislik Mimarlık Fakültesi Dergisi, 19(2).

- Dale-Olsen, H. (2006). Wages, fringe benefits and worker turnover. *Labour Economics*, 13(1), 87-105.
- Dang, Y., Chen, Z., Li, H. ve Shu, H. (2022). A comparative study of non-deep learning, deep learning, and ensemble learning methods for sunspot number prediction. arXiv preprint arXiv:2203.05757
- Davis, S. J. ve Haltiwanger, J. (2014). Labor market fluidity and economic performance. NBER Working Paper No. 20479.
- De Simone, S., Planta, A. ve Cicotto, G. (2018). The role of job satisfaction, work engagement, self-efficacy and agentic capacities on nurses' turnover intention and patient satisfaction. *Applied Nursing Research*, 39, 130–140.
- De Vos, A., De Hauw, S. ve Van der Heijden, B. I. J. M. (2011). Competency development and career success: The mediating role of employability. *Journal of Vocational Behavior*, 79(2), 438–447.
- Deci, E. L., Olafsen, A. H. ve Ryan, R. M. (2017). Self-determination theory in work organizations: The state of a science. *Annual Review of Organizational Psychology and Organizational Behavior*, 4, 19–43.
- Delfgaauw, J. (2007). The effect of job satisfaction on job search: Not just whether, but also where. *Labour Economics*, 14(3), 299-317.
- Demerouti, E. ve Bakker, A. B. (2011). The Job Demands–Resources model: Challenges for future research. *SA Journal of Industrial Psychology*, 37(2), 01–09.
- DeRue, D. S., Nahrgang, J. D., Wellman, N. ve Humphrey, S. E. (2011). Trait and behavioral theories of leadership: An integration and meta-analytic test of their relative validity. *Personnel Psychology*, 64(1), 7–52.
- Díaz-Pérez, F. ve Bethencourt-Cejas, M. (2017). An Application of the CHAID Algorithm to Study the Environmental Impact of Visitors to the Teide National Park in Tenerife, Spain. *International Business Research*, 10(7), 168–177.
- Dinçer, S. (2019). Çok Kriterli Karar Alma. Ankara: Gece Akademi Yayıncılık.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78-87.
- Draper, N. R. ve Smith, H. (1998). *Applied regression analysis* (3rd ed.). John Wiley & Sons.
- Engin, U. ve Durer, S. (2023). Financial Distress Prediction from Time Series Data Using XGBoost: BIST100 of Borsa Istanbul. *Doğuş Üniversitesi Dergisi*, 24(2), 589–604.
- Erdogan, B., Bauer, T. N., Truxillo, D. M. ve Mansfield, L. R. (2012). Whistle while you work: A review of the life satisfaction literature. *Journal of Management*, 38(4), 1038–1083.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). SAGE Publications.
- Fisher, C. D. (2010). Happiness at work. *International Journal of Management Reviews*, 12(4), 384–412.
- Fletcher, L., Alfes, K. ve Robinson, D. (2016). The relationship between perceived training and development and employee retention: the mediating role of work attitudes

- The International Journal of Human Resource Management, 29(18). 2701-2728.
- Fountzoula, C. ve Aravossis, K. (2021). Analytic hierarchy process and its applications in the public sector: A review. *Academy of Accounting and Financial Studies Journal*, 25(6), 1-15.
- French, K. A., Dumani, S., Allen, T. D. ve Shockley, K. M. (2018). A meta-analysis of work–family conflict and social support. *Psychological Bulletin*, 144(3), 284–314.
- Gao, X., Wen, J. ve Zhang, C. (2019). An improved random forest algorithm for predicting employee turnover. *Mathematical Problems in Engineering*.
- Gajendran, R. S. ve Harrison, D. A. (2007). The good, the bad, and the unknown about telecommuting: Meta-analysis of psychological mediators and individual consequences. *Journal of Applied Psychology*, 92(6), 1524–1541.
- Gillet, N., Fouquereau, E., Forest, J., Brunault, P. ve Colombat, P. (2012). The impact of organizational factors on psychological needs and their relations with well-being. *Journal of Business and Psychology*, 27(4), 437–450.
- Golden, T. D. ve Veiga, J. F. (2005). The impact of extent of telecommuting on job satisfaction: Resolving inconsistent findings. *Journal of Management*, 31(2), 301–318.
- Greene, W. H. (2012). *Econometric analysis* (7th ed.). Pearson.
- Gülpinar V. (2013). Yapay Sınır Ağları ve Sosyal Ağ Analizi Yardımı ile Türk Telekomünikasyon Piyasasında Müşteri Kaybı Analizi, *Marmara Üniversitesi İ.İ.B. Dergisi*, Cilt: 34, Sayı:1, 331–350.
- Handayani, D., Karim, A. M., Siregar, D. ve Ladayya, F. (2023). The performance of decision tree and ensemble algorithms for classifying the graduation status of undergraduate students at Universitas Negeri Jakarta. In *Proceedings of the 5th International Conference on Statistics, Mathematics, Teaching, and Research 2023* (pp. 33-41). Atlantis Press.
- Harker Martin, B. ve MacDonnell, R. (2012). Is telework effective for organizations? A meta-analysis of empirical research on perceptions of telework and organizational outcomes. *Management Research Review*, 35(7), 602–616.
- Hobfoll, S. E., Halbesleben, J., Neveu, J. P. ve Westman, M. (2018). Conservation of resources in the organizational context: The reality of resources and their consequences. *Annual Review of Organizational Psychology and Organizational Behavior*, 5, 103–128.
- Hoch, J. E., Bommer, W. H., Dulebohn, J. H. ve Wu, D. (2016). Do ethical, authentic, and servant leadership explain variance above and beyond transformational leadership? A meta-analysis. *Journal of Management*, 44(2), 501–529.
- Hoerl, A. E. ve Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67.
- Hosmer, D. W., Lemeshow, S. ve Sturdivant, R. X. (2013). *Applied Logistic Regression* (3rd ed.). John Wiley & Sons.
- Hu, Q., Schaufeli, W. B. ve Taris, T. W. (2011). The job demands–resources model: An analysis of additive and joint effects of demands and resources. *Journal of Vocational Behavior*, 79(1), 181–190.
- Idaman, A., Amrullah ve Rolanda, V. (2024). Analysis of the Additive Ratio Assessment

- Method in the Selection of the Best Production Head. *Jurnal Informasi dan Teknologi*, 6(2), 172–181.
- Ilies, R., Aw, S. S. Y. ve Pluut, H. (2015). Intraindividual models of employee well-being: What have we learned and where do we go from here? *European Journal of Work and Organizational Psychology*, 24(6), 827–838.
- Inoue, T., Ichikawa, D., Ueno, T., Cheong, M., Inoue, T., Whetstone, W. D., Endo, T., Nizuma, K. ve Tominaga, T. (2020). XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury. *Neurotrauma Reports*, 1(1), 48–56.
- Jaehn, P., Fügemann, H., Gödde, K. ve Holmberg C. (2023). Using decision tree analysis to identify population groups at risk of subjective unmet need for assistance with activities of daily living. *BMC Geriatrics*, 23, 543.
- Jantan, H., Hamdan, A. R. ve Othman, Z. A. (2010). Human talent prediction in HR management using C4.5 classification algorithm. *International Journal on Computer Science and Engineering*, 2(8), 2526-2534.
- Jordan, M. I. ve Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.
- Joseph, R., Udupa, S., Jangale, S., Kotkar, K. ve Pawar, P. (2021). Employee attrition using machine learning and depression analysis. 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 1000–1005.
- Jun, M.-J. (2021). A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: the case of the Seoul metropolitan area. *International Journal of Geographical Information Science*, 35(11), 2149–2167.
- Karakaya, İ., Sata, M., Çorbacı, E. C. ve Çetin, B. (2018). An Investigation of the Factors That Affect High School Students' Attitudes Towards Social Media by CHAID Analysis. *Eurasian Journal of Educational Research*, 76, 21–40.
- Katsikea, E., Theodosiou, M., Perdikis N. ve Kehagias J. (2011). The effects of organizational structure and job characteristics on export sales managers' job satisfaction and organizational commitment. *Journal of World Business*, 46(2), 221-233.
- Kenger, M. D. ve Organ, A. (2017). Banka personel seçiminin çok kriterli karar verme yöntemlerinden entropi temelli ARAS yöntemi ile değerlendirilmesi. *Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 4(4), 152-170.
- Klusowski, J. M. (2020). Sparse learning with CART. arXiv preprint arXiv:2006.04266.
- Kool, M. ve Van Dierendonck, D. (2012). Servant leadership and commitment to change, the mediating role of justice and optimism. *Journal of Organizational Change Management*, 25(3), 422–433.
- Kossek, E. E., Ruderman, M. N., Braddy, P. W. ve Hannum, K. M. (2012). Work–nonwork boundary management profiles: A person-centered approach. *Journal of Vocational Behavior*, 81(1), 112–128.
- Kramar, R. (2014). Beyond strategic human resource management: Is sustainable human resource management the next approach? *The International Journal of Human Resource Management*, 25(8), 1069-1089.
- Krishnan, A. R., Kasim, M. M., Hamid, R. ve Ghazali, M. F. (2021). A modified CRITIC

- method to estimate the objective weights of decision criteria. *Symmetry*, 13(6), 973.
- Kumar, P., Gaikwad, S. B., Ramya, S. T., Tiwari, T., Tiwari, M. ve Kumar, B. (2023). Predicting employee turnover: A systematic machine learning approach for resource conservation and workforce stability. *Engineering Proceedings*, 59, 117.
- Kurşunmaden, F. İ. (2024). Measuring Corporate Governance Maturity Level with CRITIC-based ARAS Method. *Selçuk Üniversitesi Sosyal Bilimler Meslek Yüksekokulu Dergisi*, 27(2), 766–774.
- Kutner, M. H., Nachtsheim, C. J., Neter, J. ve Li, W. (2005). *Applied linear statistical models* (5th ed.). McGraw-Hill.
- Kuvaas, B. (2006). Performance appraisal satisfaction and employee outcomes: Mediating and moderating roles of work motivation. *International Journal of Human Resource Management*, 17(3), 504–522.
- Kuyumcu, Z. C., Aslan, H. ve Yurtay, N. (2024). Casualty Analysis of the Drivers in Traffic Accidents in Turkey: A CHAID Decision Tree Model. *Applied Sciences*, 14(24), 11693.
- Lavy, S. ve Littman-Ovadia, H. (2016). My better self: Using strengths at work and work productivity, organizational citizenship behavior, and satisfaction. *Journal of Career Development*, 44(2), 95–109.
- Leroy, H., Palanski, M. E. ve Simons, T. (2012). Authentic leadership and behavioral integrity as drivers of follower commitment and performance. *Journal of Business Ethics*, 107(3), 255–264.
- Leslie, L. M., Manchester, C. F., Park, T. Y. ve Mehng, S. A. (2012). Flexible work practices: A source of career premiums or penalties? *Academy of Management Journal*, 55(6), 1407–1428.
- Lévy-Garboua, L., Montmarquette, C. ve Simonnet, V. (2007). Job satisfaction and quits: Theory and evidence from the German socio-economic panel. *Labour Economics*, 14(2), 251–268.
- Li Z. ve Fox E. (2023) Prediction and optimization of employee turnover intentions in enterprises based on unbalanced data. *PLoS ONE* 18(8).
- Liu, N. ve Xu, Z. (2021). An overview of ARAS method: Theory development, application extension, and future challenge. *International Journal of Intelligent Systems*, 36(7), 3524–3565.
- Lu, L., Lu, A. C. C., Gursoy, D. ve Neale, N. R. (2016). Work engagement, job satisfaction, and turnover intentions: A comparison between supervisors and line-level employees. *International Journal of Contemporary Hospitality Management*, 28(4), 737–761.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- Luthans, F. ve Stajkovic, A. D. (1999). Reinforce for performance: The need to go beyond pay and even rewards. *Academy of Management Perspectives*, 13(2), 49–57.
- Ma Q. ve Li H. (2024). A decision support system for supplier quality evaluation based on MCDM-aggregation and machine learning. *Expert Systems with Applications*. Vol:242.
- Ma, Z. (2022). The Superiority of XGBoost Model in the Forecast of Medical Stock Price in

the Period of COVID-19. Proceedings of the 2022 7th International Conference on Social Sciences and Economic Development (ICSSSED 2022).

- Maddala, G. S. (2001). *Introduction to econometrics* (3rd ed.). John Wiley & Sons.
- Martínez-Gramage, J., Albiach, J. P., Moltó, I. N., Amer-Cuenca, J. J., Huesa Moreno, V. ve Segura-Ortí, E. (2020). A random forest machine learning framework to reduce running injuries in young triathletes. *Sensors*, 20(21), 6388.
- Maslach, C. ve Leiter, M. P. (2016). Understanding the burnout experience: Recent research and its implications for psychiatry. *World Psychiatry*, 15(2), 103–111.
- Mason, R. D. ve Lind, K. (1999). *Statistical tools for business and economics* (3rd ed.). Irwin/McGraw-Hill.
- Matthews, R. A. ve Barnes-Farrell, J. L. (2010). Development and initial evaluation of an enhanced measure of boundary flexibility for the work and family domains. *Journal of Occupational Health Psychology*, 15(3), 330–346.
- Menard, S. (2002). *Applied Logistic Regression Analysis* (2nd ed.). Sage Publications.
- Michel, J. S., Kotrba, L. M., Mitchelson, J. K., Clark, M. A. ve Baltes, B. B. (2011). Antecedents of work–family conflict: A meta-analytic review. *Journal of Organizational Behavior*, 32(5), 689–725.
- Montgomery, D. C., Peck, E. A. ve Vining, G. G. (2012). *Introduction to linear regression analysis* (5th ed.). Wiley.
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT press.
- Nadkarni, S. B., Vijay, G. S. ve Kamath, R. C. (2023). Comparative study of random forest and gradient boosting algorithms to predict airfoil self-noise. *Engineering Proceedings*, 59(1), 24.
- Navanilal, S. S. (2021). A quick review of machine learning algorithms using the ARAS Method. *Data Analytics and Artificial Intelligence*. Vol: 1(2).
- Noe, R. A., Hollenbeck, J. R., Gerhart, B., ve Wright, P. M. (2015). *Human Resource Management: Gaining a Competitive Advantage* (9th ed.). McGraw-Hill Education.
- Obermeyer, Z. ve Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. *The New England Journal of Medicine*, 375(13), 1216-1219.
- Ordovas, J. M., Rios Insua, D., Santos-Lozano, A., Lucia, A., Torres, A., Kosgodagan, A. ve Camacho, J. M. (2021). A Bayesian network model for predicting cardiovascular risk. *arXiv preprint arXiv:2112.13963*.
- Öneren M., Arar T. ve Çelebioğlu E. S. (2017). Akademik Temelini Güçlü Kılmak: Araştırma Görevlisi Alımındaki Faktörlerin AHP ile Belirlenmesi. *Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*. 9(1), 39-50.
- Özgül Ö. (2006). Bir işletme için topsis ve AHP yöntemleri ile ERP yazılımının seçimi. *Sakarya Üniversitesi, Endüstri Mühendisliği Ana Bilim Dalı (Yüksek Lisans Tezi)*.
- Palaz, H. ve Kovancı, A. (2008). Türk Deniz Kuvvetleri denizaltılarının seçiminin AHP ile değerlendirilmesi. *Havacılık ve Uzay Teknolojileri Dergisi*, 3(3), 53-60.
- Stojčić, M., Zavadskas, E. K., Pamučar, D., Stević, Ž., & Mardani, A. (2019). Application of MCDM Methods in Sustainability Engineering: A Literature Review 2008–2018. *Symmetry*, 11(3), 350.

- Panaccio, A. ve Vandenberghe, C. (2011). The relationships of role clarity and organization-based self-esteem to commitment to supervisors and organizations, and turnover intentions. *Journal of Applied Social Psychology*, 41(6), 1455–1485.
- Parry, E. (2016). *Generational diversity at work: New research perspectives*. Routledge.
- Prediction of employee attrition using machine learning and ensemble methods. *International Journal of Machine Learning*, 11(2).
- Dutta, S. ve Bandyopadhyay, S. K. (2020). Employee attrition prediction using neural network cross validation method. *International Journal of Commerce and Management*, 6(3), 80–85.
- Poornappriya, T. S. ve Gopinath, R. (2021). Employee attrition in human resource using machine learning techniques. *Webology*, 18(6).
- Pourkhodabakhsh N., Mamoudan M. M. ve Bozorgi-Amiri A. (2022). Effective machine learning, Meta-heuristic algorithms and multi-criteria decision making to minimizing human resource turnover. *Applied Intelligence*. Vol:53, 16309–16331.
- Punnoose, R. ve Ajit, P. (2016). Prediction of employee turnover in organizations using machine learning algorithms. *International Journal of Advanced Research in Artificial Intelligence*, 5(9), 22-26.
- Rathore, H., Sahay, S. K., Thukral, S. ve Sewak, M. (2021). Detection of malicious Android applications: Classical machine learning vs. deep neural network integrated with clustering. arXiv preprint arXiv:2103.00637.
- Reşitoğlu S. (2011). “Yetkinlik Bazlı performans değerlendirme ve çalışan memnuniyeti- Bir Uygulama”. Yüksek lisans tezi. Dokuz Eylül Üniversitesi, İzmir.
- Reza, S. ve Majid, A. (2013). Ranking financial institutions based on of trust in online banking using ARAS and ANP method. *International Research Journal of Applied and Basic Sciences*, 6(4), 415-423.
- Reddy, S. S., Rajanikanth, J., Sivaramaraju, V. V. ve Murthy, K. V. S. S. R. (2018). Prediction of employee attrition using data mining. 2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCA), 1–8.
- Qutub, A., Al-Mehmadi, A., Al-Hssan, M., Aljohani, R. ve Alghamdi, H. S. (2021).
- Rhoades, L. ve Eisenberger, R. (2002). Perceived organizational support: A review of the literature. *Journal of Applied Psychology*, 87(4), 698–714.
- Ricci, F., Rokach, L. ve Shapira, B. (2015). *Recommender systems handbook*. Springer.
- Rich, B. L., Lepine, J. A. ve Crawford, E. R. (2017). Job engagement: Antecedents and effects on job performance. *Academy of Management Journal*, 53(3), 617–635.
- Rizkallah, L. W. (2025). Enhancing the performance of gradient boosting trees on regression problems. *Journal of Big Data*, 12(1), 35.
- Russell, S. ve Norvig, P. (2016). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.
- Rynes, S. L., Gerhart, B. ve Parks, L. (2005). Personnel psychology: Performance evaluation and pay for performance. *Annual Review of Psychology*, 56, 571–600.
- Saaty, T. L. (2004). Decision making- the analytic hierarchy and network processes (AHP/ANP). *Journal of Systems Science and Systems Engineering*, 13(1), 1–35.
- Saradhi, V. V. ve Palshikar, G. K. (2011). Employee churn prediction. *Expert Systems with Applications*, 38(3), 1999-2006.

- Sari, S. F. ve Lhaksana, K. M. (2022). Employee attrition prediction using feature selection with information gain and random forest classification. *Journal of Computer System and Informatics (JoSYC)*.
- Schaufeli, W. B. ve Taris, T. W. (2013). A critical review of the Job Demands-Resources Model: Implications for improving work and health. In G. F. Bauer & O. Hämmig (Eds.), *Bridging occupational, organizational and public health* (pp. 43–68). Springer.
- Sedghpour, A. S. ve Sedghpour, M. R. S. (2020). Web Document Categorization Using Naive Bayes Classifier and Latent Semantic Analysis. *arXiv preprint arXiv:2006.01715*.
- Shankar, R. S., Rajanikanth, J., Sivaramaraju, V. V. ve Murthy, K. V. S. S. R. (2018). Prediction of employee attrition using data mining. *Proceedings of the IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India.
- Shen, J., Tang, N. ve D'Netto, B. (2010). Effects of human resource diversity management on organizational citizen behaviour in the Chinese context. *The International Journal of Human Resource Management*, 25(12), 1662-1682.
- Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R. ve Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: Concepts, techniques, and applications in R*. Wiley.
- Shockley, K. M. ve Allen, T. D. (2007). When flexibility helps: Another look at the availability of flexible work arrangements and work–family conflict. *Journal of Vocational Behavior*, 71(3), 479–493.
- Shockley, K. M. ve Singla, N. (2011). Reconsidering work–family interactions and satisfaction: A meta-analysis. *Journal of Management*, 37(3), 861–886.
- Sivam S.P.S.S. ve Rajendran R. (2022). A study on deep drawn cups and the selection of optimal settings deploying ANN training and architectural parameters using the Taguchi ARAS approach. *Proceedings of the Institution of Mechanical Engineers, Part C*. 238(8):3287-3298.
- Soba, M. ve Eren, K. (2011). TOPSIS YÖNTEMİNİ KULLANARAK FİNANSAL VE FİNANSAL OLMAYAN ORANLARA GÖRE PERFORMANS DEĞERLENDİRİLMESİ, ŞEHİRLERARASI OTOBÜS SEKTÖRÜNDE BİR UYGULAMA. *Sosyal Ekonomik Araştırmalar Dergisi*, 11(21), 23-40.
- Sonnentag, S. ve Fritz, C. (2015). Recovery from job stress: The stressor-detachment model as an integrative framework. *Journal of Organizational Behavior*, 36(S1), S72–S103.
- Stamolampros, P., Korfiatis, N., Chalvatzis, K. ve Buhalis, D. (2019). Job satisfaction and employee turnover determinants in high contact services: Insights from employees' online reviews. *Tourism Management*, 75, 130-147.
- Stone, D. L. ve Deadrick, D. L. (2015). Challenges and opportunities affecting the future of human resource management. *Human Resource Management Review*, 25(2), 139-145.
- SVM Nasıl Çalışır (2023). <https://www.ibm.com/docs/tr/spss-modeler/18.4.0?topic=models-how-svm-works>. Erişim tarihi: 12.09.2024.
- Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3), 647–665.
- Şimşek, H. K. (2022). Makine öğrenmesi dersleri 5b: Random forest (regresyon). *Medium*. <https://medium.com/data-science-tr/makine-ogrenmesi-dersleri-5a-random-forest-regresyon->

- Subhashini, M. ve Gopinath, R. (2020). Employee attrition prediction in industry using machine learning techniques. *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 11(12).
- Thomas, N. S. ve Kaliraj, S. (2024). An improved and optimized random forest based approach to predict the software faults. *SN Computer Science*, 5, 530.
- Tims, M., Bakker, A. B. ve Derks, D. (2013). The impact of job crafting on job demands, job resources, and well-being. *Journal of Occupational Health Psychology*, 18(2), 230–240.
- Turskis, Z. ve Zavadskas, E. K. (2010). A novel method for multiple criteria analysis: Grey additive ratio assessment (ARAS-G) method. *Informatika*, 21(4), 597- 610.
- Turskis, Z., Zavadskas, E. K. ve Peldschus, F. (2009). Multi-criteria optimization system for decision making in construction design and management.
- Van der Heijden, B. I. J. M., ve Bakker, A. B. (2011). Toward a mediation model of employability enhancement: A study of employee-supervisor pairs in the building sector. *The Career Development Quarterly*, 59(3), 232–248.
- Wang, G., Oh, I. S., Courtright, S. H. ve Colbert, A. E. (2011). Transformational leadership and performance across criteria and levels: A meta-analytic review of 25 years of research. *Group & Organization Management*, 36(2), 223–270.
- Wang, H. ve Howell, J. M. (2010). Exploring the dual-level effects of transformational leadership on followers. *Journal of Applied Psychology*, 95(6), 1134–1144.
- Warr, P. ve Inceoglu, I. (2012). Job engagement, job satisfaction, and contrasting associations with person–job fit. *Journal of Occupational Health Psychology*, 17(2), 129–138.
- Wayne, J. H., Lemmon, G., Hoobler, J. M., Cheung, G. W. ve Wilson, M. S. (2017). The ripple effect: A spillover model of the detrimental impact of work–family conflict on job success. *Journal of Organizational Behavior*, 38(6), 876–894.
- White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4), 817-838.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data* (2nd ed.). MIT Press.
- Yalçın C., (2019). Müşteri Kayıp Analizi (Customer Churn Analysis), YBS Ansiklopedi, Cilt:7.
- Yang, J. T., Wan, C. S. ve Fu, Y. J. (2012). Qualitative examination of employee turnover and retention strategies in international tourist hotels in Taiwan. *International Journal of Hospitality Management*, 31(3), 837-848.
- Yapar, D., Yapar, A., Tokgöz, M. A. ve Bilge, U. (2023). Decision tree-based data mining approach for the evaluation of survival in primary malignant bone tumors: A surveillance, epidemiology and end results database study. *Journal of Orthopaedic Surgery*, 31(2).
- Yıldız, A. Y. ve Kalaycı, A. (2025). Gradient boosting decision trees on medical diagnosis over tabular data. *arXiv preprint arXiv:2410.03705*
- Ying, Z., Xu, Z., Li, Z., Wang, W., & Meng, C. (2022). MT-GBM: A multi-task gradient boosting machine with shared decision trees. *arXiv preprint arXiv:2201.06239*

- Yuan J. (2021). "Research on Employee Turnover Prediction Based on Machine Learning Algorithms," 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China., pp. 114-120.
- Zavadskas, E. K., Turskis, Z. ve Kildienė, S. (2014). State of art surveys of overviews on MCDM/MADM methods. *Technological and Economic Development of Economy*, 20(1), 165-179.
- Zhang, Z. (2023). Entropy-based statistics and their applications. *Entropy*, 25(6), 936.

