

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



SÜRÜ ZEKÂSI TABANLI YÖNTEMLER
KULLANILARAK ÇEVİRİMİÇİ SOSYAL AĞLARDA
SAHTE HABER TESPİTİ

Feyza ALTUNBEY ÖZBAY

Doktora Tezi

YAZILIM MÜHENDİSLİĞİ ANABİLİM DALI

OCAK 2020

T.C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

Doktora Tezi

**SÜRÜ ZEKÂSI TABANLI YÖNTEMLER KULLANILARAK ÇEVİRİMİÇİ SOSYAL
AĞLARDA SAHTE HABER TESPİTİ**

Tez Yazarı

Feyza ALTUNBEY ÖZBAY
(142137205)

Danışman

Prof. Dr. Bilal ALATAŞ

OCAK 2020

ELAZIĞ

T. C.
FIRAT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

Yazılım Mühendisliği Anabilim Dalı

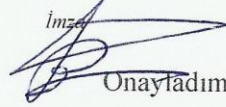
Doktora Tezi

Başlığı: Sürü Zekâsı Tabanlı Yöntemler Kullanılarak Çevrimiçi Sosyal Ağlarda Sahte Haber Tespiti
Yazarı: Feyza ALTUNBEY ÖZBAY
İlk Teslim Tarihi: 24.12.2019
Savunma Tarihi: 24.01.2020

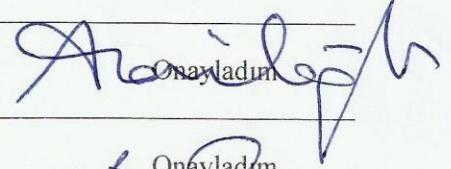
TEZ ONAYI

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına göre hazırlanan bu tez aşağıda imzaları bulunan jüri üyeleri tarafından değerlendirilmiş ve akademik dinleyicilere açık yapılan savunma sonucunda OYBİRLİĞİ ile kabul edilmiştir.

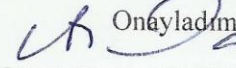
Danışman: Prof. Dr. Bilal ALATAŞ
Fırat Üniversitesi, Mühendislik Fakültesi

imza

Onayladım

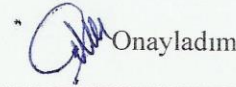
Başkan: Prof. Dr. Abdulsamet HAŞILOĞLU
Atatürk Üniversitesi, Mühendislik Fakültesi


Onayladım

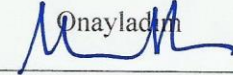
Üye: Prof. Dr. Asaf VAROL
Fırat Üniversitesi, Teknoloji Fakültesi


Onayladım

Üye: Prof. Dr. Sabri KOÇER
Necmettin Erbakan Üniversitesi, Müh. ve Mim. Fakültesi


Onayladım

Üye: Prof. Dr. Mehmet KAYA
Fırat Üniversitesi, Mühendislik Fakültesi


Onayladım

Bu tez, Enstitü Yönetim Kurulunun/...../20..... tarihli toplantısında tescillenmiştir.

imza
Prof. Dr. Soner ÖZGEN
Enstitü Müdürü

BEYAN

Fırat Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygun olarak hazırladığım “Sürü Zekâsı Tabanlı Yöntemler Kullanılarak Çevrimiçi Sosyal Ağlarda Sahte Haber Tespiti” Başlıklı Doktora Tezimin içindeki bütün bilgilerin doğru olduğunu, bilgilerin üretilmesi ve sunulmasında bilimsel etik kurallarına uygun davrandığımı, kullandığım bütün kaynakları atıf yaparak belirttiğimi, maddi ve manevi desteği olan tüm kurum/kuruluş ve kişileri belirttiğimi, burada sunduğum veri ve bilgileri unvan almak amacıyla daha önce hiçbir şekilde kullanmadığımı beyan ederim.

24/12/2019

Feyza ALTUNBEY ÖZBAY

ÖNSÖZ

20. yüzyılda internetin bulunması ve günümüzdeki bilgi teknolojilerinin hızlı gelişiminin bir sonucu olarak ortaya çıkan sosyal medya kavramı sosyal, kültürel, ekonomik ve daha birçok alanda yaşamın önemli bir parçası haline gelmiştir. Sosyal medya kavramının ortaya çıkışı ile toplum üzerinde bir takım değişiklikler meydana gelmiştir. Özellikle sosyal medyanın düşük maliyetli, hızlı ve kolay erişilebilir olması insanların haber kaynaklarını değiştirmesine neden olmuştur. Sosyal medyanın sağladığı birçok avantaj yanında getirdiği bir takım olumsuzluklar bulunmaktadır. Sosyal medyadaki haberlerin kalitesi geleneksel haber kaynakları ile kıyaslandığında oldukça düşüktür. Çevrimiçi olarak haberler sunmak ucuz ve sosyal medya aracılığı ile yaymak daha hızlı ve kolay olduğu için, sosyal medyada finansal, politik veya şahsi çıkarlar sağlamak amacıyla kasıtlı olarak yanlış bilgiler içeren büyük miktarda sahte haber üretilmektedir. Sosyal medyadaki sahte haberler, bireylere ve toplumlara büyük ölçüde zarar vermektedir. Bundan dolayı sahte haberlerin tespit edilmesi problemi ortaya çıkmıştır. Sahte haber tespiti güncel ve aktif bir araştırma alanı olmasına rağmen probleme daha önce optimizasyon bakış açısı ile yaklaşan çalışma bulunmamaktadır. Bu tez çalışmasında, sahte haber tespiti problemi ilk defa bir optimizasyon problemi olarak ele alınmıştır. Ayrıca iki yeni adaptif optimizasyon algoritması da ilk defa bu tez kapsamında önerilmiş ve sahte haber tespiti problemine uyarlanmıştır.

Doktora çalışmam süresince, değerli görüş ve katkılarıyla beni yönlendiren, her konuda önerileri ile desteğini esirgemeyen, bilimsel çalışma disiplinini bize kazandıran ve kıymetli tecrübelerinden faydalandığım tez danışmanım Sayın Prof. Dr. Bilal ALATAŞ'a teşekkürü borç bilirim.

Son olarak sadece tez döneminde değil, her zaman yanımda olup bana destek olan eşim Dr. Erdal ÖZBAY'a ve aileme teşekkür ederim.

Feyza ALTUNBEY ÖZBAY

ELAZIĞ-2020

İÇİNDEKİLER

	Sayfa
ÖNSÖZ.....	iv
İÇİNDEKİLER	v
ÖZET	vii
ABSTRACT.....	viii
ŞEKİLLER LİSTESİ	ix
TABLOLAR LİSTESİ	xi
SİMGELER	xii
KISALTMALAR	xiii
1. GİRİŞ.....	1
1.1. Tezin Amacı	3
1.2. Tezin Kapsamı	4
2. VERİ MADENCİLİĞİ.....	5
2.1. Veri Madenciliği Süreci	5
2.1.1. Veri Seçimi	6
2.1.2. Ön İşleme ve Veri Temizleme	6
2.1.3. Veri Dönüştürme	6
2.1.4. Veri Madenciliği Algoritmasının Uygulanması.....	6
2.1.5. Sonuçların Yorumlanması / Değerlendirilmesi	7
2.2. Veri Madenciliği Modelleri	7
2.2.1. Tahmin Edici Modeller.....	7
2.2.2. Tanımlayıcı Modeller	8
3. OPTİMİZASYON	10
3.1. Optimizasyona Genel Bakış	10
3.2. Metasezgisel Algoritmalar	11
4. SAHTE HABER TESPİTİ.....	15
4.1. Sahte Haber Nedir?	15
4.1.1. Sahte Haberin Tanımı	15
4.1.2. Sahte Haber İle İlişkili Kavramlar	16
4.1.3. Sahte Haber Kavramında Anahtar Roller	17
4.2. Sahte Haberlerin Toplum Üzerindeki Etkileri.....	17
4.3. Sahte Haberleri Tespit Etmenin Önemi.....	18
4.4. Sahte Haber Belirlemenin Zorlukları	19
4.5. Sahte Haber Tespiti ile İlgili Literatür Taraması	20
4.5.1. İçerik Temelli Sahte Haber Tespiti Yaklaşımları	20
4.5.2. Bağlam Temelli Sahte Haber Tespiti Yaklaşımları	25
5. UYGULAMADA KULLANILAN ALGORİTMALAR, YÖNTEM VE VERİ KÜMELERİ.....	28
5.1. Gri Kurt Optimizasyon Algoritması	28
5.1.1. Algoritmanın İlham Kaynağı	28
5.1.2. Matematiksel Model ve Algoritma	29
5.2. Salp Sürüsü Optimizasyon Algoritması	35
5.2.1. Algoritmanın İlham Kaynağı	35
5.2.2. Salp Zincirlerinin Hareketi için Önerilen Matematiksel Model	35

5.3. Adaptif Salp Sürüsü Optimizasyon Algoritmaları.....	38
5.3.1. Doğrusal Olmayan Şekilde Azalan Atalet Ağırlıklı Salp Sürüsü Optimizasyon Algoritması	39
5.3.2. Osilasyon (Salınım) Atalet Ağırlıklı Salp Sürüsü Optimizasyon Algoritması	40
5.4. Denetimli Yapay Zekâ Algoritmaları	40
5.5. Veri Ön İşleme Adımları	45
5.6. Sahte Haber Veri Kümeleri	47
5.6.1. BuzzFeed Politik Haber Veri Kümesi	47
5.6.2. Rastgele Politik Sahte Haber Veri Kümesi	48
5.6.3. LIAR Sahte Haber Veri Kümesi	49
5.6.4. ISOT Sahte Haber Veri Kümesi	49
5.7. Performans Değerlendirme Ölçütleri	50
5.8. Optimizasyon Algoritmalarının Sahte Haber Tespiti için Modellenmesi.....	51
5.9. Friedman Testi	52
5.10. Kutu Grafiği	53
6. DENEYSEL BULGULAR VE TARTIŞMA	55
6.1. BuzzFeed Politik Haberler Veri Kümesinden Elde Edilen Sonuçlar	55
6.2. Rastgele Politik Haberler Veri Kümesinden Elde Edilen Sonuçlar	62
6.3. LIAR Veri Kümesinden Elde Edilen Sonuçlar.....	70
6.4. ISOT Veri Kümesinden Elde Edilen Sonuçlar	77
7. SONUÇLAR	84
KAYNAKLAR	86
ÖZGEÇMİŞ	

ÖZET

Sürü Zekâsı Tabanlı Yöntemler Kullanılarak Çevrimiçi Sosyal Ağlarda Sahte Haber Tespiti

Feyza ALTUNBEY ÖZBAY

Doktora Tezi

FIRAT ÜNİVERSİTESİ
Fen Bilimleri Enstitüsü
Yazılım Mühendisliği Anabilim Dalı
Ocak 2020, Sayfa: xiii + 95

Son yıllarda, sosyal medyanın hızlı gelişimi insanların bilgiye erişim biçimlerini değiştirmiştir. Kullanıcılar, dünyadaki sosyal, ekonomik, politik ve bilimsel olaylar hakkındaki bilgilere sosyal medya aracılığıyla erişmektedir. Sosyal medyadaki haberlerin video ve resimler içermesi, televizyon ve gazete gibi geleneksel haber platformlarının önemini yitirmesine neden olmaktadır. Ayrıca, çevrimiçi sosyal medya, bilgiye kolay erişim, düşük maliyet ve bilgilerin hızla yayılması gibi avantajlar da sağlar. Sosyal medya birçok avantaja sahip olsa da, maalesef, sosyal medyadaki haberlerin çoğu kötü niyetli kişiler tarafından değiştirilebilir ve bu nedenle bu haberler güvenilir olmayabilir. Bu tür haberler sosyal medya üzerinden hızlı yayılır ve sosyal medya okuyucuları ve kullanıcıları üzerinde olumsuz etkiye neden olur. Bu nedenle, sahte haberlerin neden olduğu olumsuz etkiyi azaltmak için çevrim içi sosyal medyadaki sahte haberlerin tespit edilmesi gerekmektedir. Sahte haber tespiti yeni bir araştırma alanı olmasına rağmen çok dikkat çekmektedir.

Bu tez çalışmasında, oldukça popüler ve ilgi çekici çevrimiçi sosyal medya problemlerinden biri olan Sahte Haber Tespiti problemi bir optimizasyon problemi olarak ele alınmıştır. Optimizasyon çalışması sırasında ikisi ilk defa bu tezde önerilen yeni adaptif optimizasyon algoritmaları olmak üzere dört adet metasezgisel optimizasyon algoritması kullanılmış ve literatürde kullanılan otuz adet denetimli yapay zekâ algoritması ile sahte haber tespiti sonuçları kıyaslanmıştır. Önerilen yöntem çok yeni olmasına rağmen, umut verici sonuçlar elde edilmiştir.

Anahtar Kelimeler: Çevrimiçi sosyal medya, Sahte haber tespiti, Optimizasyon yöntemleri.

ABSTRACT

Fake News Detection in Online Social Networks Using Swarm Intelligence Based Methods

Feyza ALTUNBEY ÖZBAY

Ph. D. Thesis

FIRAT UNIVERSITY
Graduate School of Natural and Applied Sciences

Department of Software Engineering
December 2019, Pages: xiii + 95

In recent years, the rapid development of social media has changed the way people access information. Users can access information about social, economic, political and scientific events in the world through social media. The fact that the news on social media includes videos and pictures causes the loss of importance of traditional news platforms, such as television and newspapers. In addition, online social media provides advantages such as easy access to information, low cost and rapid dissemination of information. Although social media has many advantages, unfortunately, most of the news on social media may be changed by malicious people and therefore, it may not be reliable. Such news spreads quickly through social media and causes a negative impact on social media readers and users. Therefore, to reduce the negative effects of fake news, fake news on social media needs to be detected. Although fake news detection is a new area of research, it has attracted much attention.

In this thesis, Fake News Detection problem, one of the most popular and interesting online social media problems, has been considered as an optimization problem. During the optimization study, four heuristic optimization algorithms have been used, two of them have been proposed for the first time in this thesis, which are the new adaptive optimization algorithms and thirty supervised artificial intelligence algorithms used in the literature and fake news detection results have been compared. Although the proposed method is very new, promising results have been achieved.

Keywords: Online social media, Fake news detection, Optimization methods.

ŞEKİLLER LİSTESİ

Sayfa

Şekil 1.1.	Sosyal medya ve kapsamı	2
Şekil 2.1.	Veri madenciliği adımları	5
Şekil 2.2.	Veri madenciliği modelleri	7
Şekil 3.1.	Optimizasyon için matematiksel modeller	11
Şekil 3.2.	Metasezgisel yöntemler	14
Şekil 4.1.	Pınar Barış Harekâtı sırasında paylaşılan sahte habere ait resimler. (a) Barış Harekâtı sırasında çekildiği iddia edilen resim. (b) 2018 yılında Yemen’de çekilen resim	18
Şekil 4.2.	Sahte haber hacmi çeşitliliği ve hızı	18
Şekil 5.1.	Gri kurt hiyerarşisi (Baskınlık yukarıdan aşağıya doğru azalır)	28
Şekil 5.2.	Gri kurtların avlanma davranışı: (A) Avı takip etme, kovalama ve yaklaşma. (B-D) Avın peşinde olma, etrafını kuşatma ve taciz etme. (E) Sabit konum alma ve ava saldırı	30
Şekil 5.3.	2B ve 3B konum vektörleri ve onların bir sonraki olası konumları	31
Şekil 5.4.	GKO’da konum güncelleme	32
Şekil 5.5.	Ava saldırı ve av arayışının karşılaştırılması	33
Şekil 5.6.	GKO algoritmasının sözde kodu	35
Şekil 5.7.	(a) Bireysel salp. (b) Salp sürüsü (Salp zinciri)	36
Şekil 5.8.	SSO algoritmasının sözde kodu	37
Şekil 5.9.	Doğrusal olmayan atalet ağırlığı değerleri	39
Şekil 5.10.	Osilasyon atalet ağırlığı değerleri	40
Şekil 5.11.	Veri üzerinde yapılan ön işlem adımları	46
Şekil 5.12.	Doküman Terim Matrisi	47
Şekil 5.13.	BuzzFeed politik haberler veri kümesinden bir kesit	48
Şekil 5.14.	Rastgele sahte haber veri kümesinden bir kesit	48
Şekil 5.15.	LIAR sahte haber veri kümesinden bir kesit	49
Şekil 5.16.	ISOT sahte haber veri kümesinden bir kesit	50
Şekil 5.17.	Sahte haber tespiti için önerilen yöntemin yapısı	52
Şekil 5.18.	5 sayılı özetleme tablosu	54
Şekil 6.1.	SSO, GKO, ASSO1 ve ASSO2 algoritmaları ile BuzzFeed veri kümesi için elde edilen sonuçların kutu grafiği	56
Şekil 6.2.	BuzzFeed politik haberler veri kümesi için elde edilen doğruluk değerleri	59
Şekil 6.3.	BuzzFeed politik haberler veri kümesi için elde edilen kesinlik değerleri	60
Şekil 6.4.	BuzzFeed politik haberler veri kümesi için elde edilen duyarlılık değerleri	61
Şekil 6.5.	BuzzFeed politik haberler veri kümesi için elde edilen F-ölçütü değerleri	62
Şekil 6.6.	SSO, GKO, ASSO1 ve ASSO2 algoritmaları ile rastgele politik veri kümesi için elde edilen sonuçların kutu grafiği	63
Şekil 6.7.	Rastgele politik haberler veri kümesi için elde edilen doğruluk değerleri	66
Şekil 6.8.	Rastgele politik haberler veri kümesi için elde edilen kesinlik değerleri	67
Şekil 6.9.	Rastgele politik haberler veri kümesi için elde edilen duyarlılık değerleri	68
Şekil 6.10.	Rastgele politik haberler veri kümesi için elde edilen F-ölçütü değerleri	69
Şekil 6.11.	SSO, GKO, ASSO1 ve ASSO2 algoritmaları ile LIAR veri kümesi için elde edilen sonuçların kutu grafiği	71
Şekil 6.12.	LIAR veri kümesi için elde edilen doğruluk değerleri	73

Şekil 6.13. LIAR veri kümesi için elde edilen kesinlik değerleri	74
Şekil 6.14. LIAR veri kümesi için elde edilen duyarlılık değerleri.....	75
Şekil 6.15. LIAR veri kümesi için elde edilen F-ölçütü değerleri.....	75
Şekil 6.16. SSO, GKO, ASSO1 ve ASSO2 algoritmaları ile ISOT veri kümesi için elde edilen sonuçların kutu grafiği.....	78
Şekil 6.17. ISOT veri kümesi için elde edilen doğruluk değerleri	80
Şekil 6.18. ISOT veri kümesi için elde edilen kesinlik değerleri	81
Şekil 6.19. ISOT veri kümesi için elde edilen duyarlılık değerleri	82
Şekil 6.20. ISOT veri kümesi için elde edilen F-ölçütü değerleri	82



TABLolar LİSTESİ

Sayfa

Tablo 1.1.	Sosyal medyadaki aktif kullanıcı sayıları.....	2
Tablo 5.1.	ASSO algoritmasının sembol ve anlamları	39
Tablo 5.2.	Rastgele sahte haber veri kümesinin haber kaynakları.....	49
Tablo 5.3.	Sahte haber için karmaşıklık matrisi	50
Tablo 5.4.	Kullanılan performans değerlendirme ölçütleri.....	51
Tablo 5.5.	(5.18) denklemindeki semboller ve anlamları	53
Tablo 6.1.	Optimizasyon algoritmalarının parametreleri ve değerleri.....	55
Tablo 6.2.	Optimizasyon algoritmalarının BuzzFeed politik haberler veri kümesine uygulanması sonucunda elde edilen sonuçlar	56
Tablo 6.3.	BuzzFeed politik veri kümesi için ortalama fonksiyon değerlerine göre Friedman testi ile elde edilen algoritma ortalama sıralamaları	57
Tablo 6.4.	BuzzFeed politik veri kümesi için Friedman testi sonuçları	57
Tablo 6.5.	BuzzFeed politik haberler veri kümesi için denetimli yapay zekâ algoritması ile elde edilen sonuçlar	58
Tablo 6.6.	Optimizasyon algoritmalarının rastgele politik haberler veri kümesine uygulanması sonucunda elde edilen sonuçlar	63
Tablo 6.7.	Rastgele veri kümesi için Friedman testi ile elde edilen sıralamalar	64
Tablo 6.8.	Rastgele veri kümesi için Friedman testi ile elde edilen test istatistikleri	64
Tablo 6.9.	Random veri kümesi için denetimli yapay zekâ algoritması ile elde edilen sonuçlar	65
Tablo 6.10.	Optimizasyon algoritmalarının LIAR veri kümesine uygulanması sonucunda elde edilen sonuçlar	70
Tablo 6.11.	LIAR veri kümesi için Friedman testi ile elde edilen sıralamalar	71
Tablo 6.12.	LIAR veri kümesi için Friedman testi ile elde edilen test istatistikleri	71
Tablo 6.13.	LIAR veri kümesi için denetimli yapay zekâ algoritması ile elde edilen sonuçlar	72
Tablo 6.14.	Optimizasyon algoritmalarının ISOT veri kümesine uygulanması sonucunda elde edilen sonuçlar	77
Tablo 6.15.	ISOT veri kümesi için Friedman testi ile elde edilen sıralamalar	78
Tablo 6.16.	ISOT veri kümesi için Friedman testi ile elde edilen test istatistikleri	78
Tablo 6.17.	ISOT veri kümesi için denetimli yapay zekâ algoritması ile elde edilen sonuçlar	79

SİMGELER

\bar{a}	: Yinelemeler boyunca 2'den 0'a düşer.
\vec{A}	: Katsayı vektörü
\vec{C}	: Katsayı vektörü
c_1	: Keşif ve sömürü arasındaki denge parametresi
c_2 ve c_3	: Rastgele sayılar
df	: Serbestlik derecesi
d_i	: i . dokümandaki tüm terimlerin toplam sayısı
F_j	: j . boyutta yiyecek kaynağının konumu
H_0	: Sıfır hipotezi
H_1	: Karşıt hipotez
k	: Grup sayısı
lb_j	: j . boyutun alt sınırı
n	: Satır sayısı
n_{ij}	: i . dokümanda j . kelimenin sayısı
Q_1	: Birinci dörttebirlik
Q_3	: Üçüncü dörttebirlik
R_j	: Her bir gruba ilişkin sıra sayıları toplamı
t	: Mevcut yineleme
ub_j	: j . boyutun üst sınırı
w	: Atalet ağırlığı
w_{max}	: Maksimum atalet ağırlığı
w_{min}	: Minimum atalet ağırlığı
\vec{X}_p	: Avın konum vektörü
X_R^2	: İstatistiği seçilen yanılma düzeyinde $k-1$ serbestlik dereceli ki-kare dağılımı
x_j^i	: j . boyuttaki i . takipçi salpın konumu
x_j^1	: j . boyutta ilk salpın konumu
X_{maks}	: En büyük uygunluk fonksiyonu değeri
X_{med}	: Medyan (ortanca)
X_{min}	: En küçük uygunluk değeri

KISALTMALAR

2B	: İki-boyutlu
3B	: Üç-boyutlu
ABD	: Amerika Birleşik Devletleri
API	: Application Programming Interface
ARPANET	: Advanced Research Project Agency Network
ASC	: Attribute Selected Paper
ASSO1	: Adaptif Salp Sürü Optimizasyonu 1
ASSO2	: Adaptif Salp Sürü Optimizasyonu 2
CvC	: Classification via Clustering
CVPS	: Cross Validation Parameter Selection
ÇSM	: Çevrimiçi Sosyal Medya
DT	: Decision Tree
DTM	: Doküman Terim Matrisi
GA	: Genetik Algoritma
GKO	: Gri Kurt Optimizasyonu
KLR	: Kernel Logistic Regression
k-NN	: k-Nearest Neighbors
LMT	: Logistic Model Tree
LR	: Linear Regression
LSVM	: Linear Support Vector Machine
LWL	: Locally Weighted Learning
MLP	: Multi-Layer Perceptron
NLR	: Nonlinear Regression
OLM	: Ordinal Learning Model
PSO	: Parçacık Sürüsü Optimizasyonu
RFC	: Randomizable Filtered Classifier
SGD	: Stochastic Gradient Descent
SMO	: Sequential Minimal Optimization
SSO	: Salp Sürüsü Optimizasyonu
SVM	: Support Vector Machine
TF	: Term Frequency
TF-IDF	: Term Frequency-Inverse Document Frequency
VUM	: Vektör Uzay Modeli
WIHW	: Weighted Instances Handler Wrapper

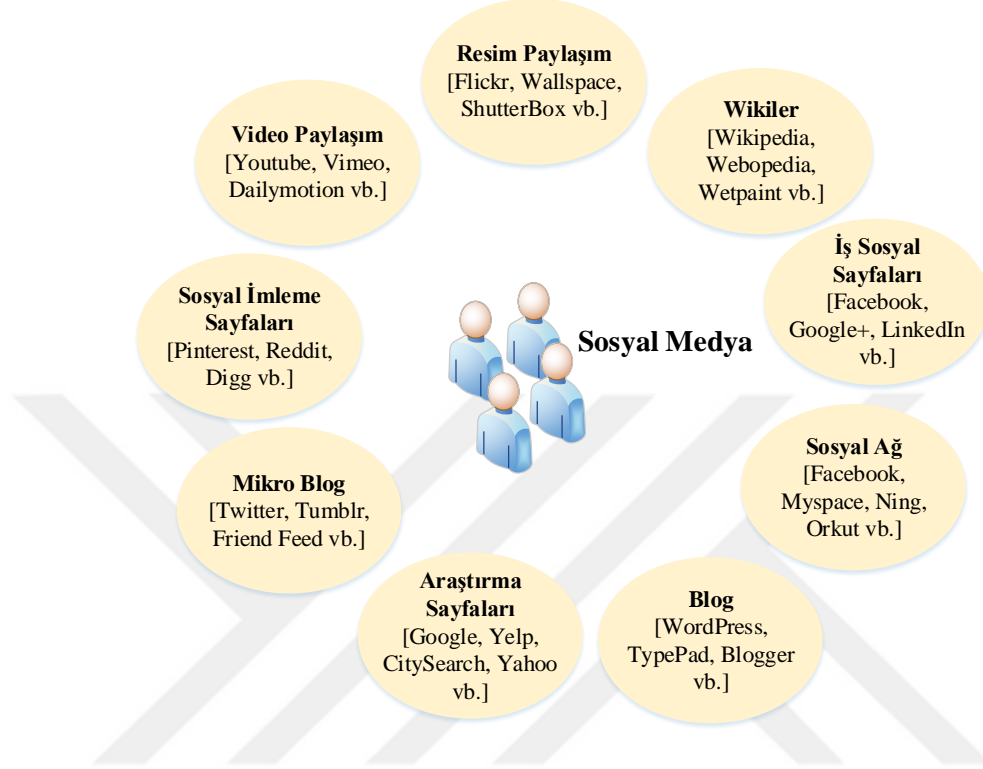
1. GİRİŞ

Altmışlı yılların başında, Amerika Birleşik Devletleri'nde (ABD) olası bir nükleer savaş sırasında askerlerin güvenli bir şekilde iletişim kurabilmesi için RAND Corporation adlı kuruluş tarafından geliştirilen bir bilgisayar ağı internetin gelişiminde ilk adım olmuştur. Sonrasında 1969 yılında, ABD Savunma Bakanlığı tarafından bilgisayarlar arasında ağ iletişimi gerçekleştirmek için başlatılan ARPANET (Advanced Research Project Agency Network) ile gelişimini sürdürmüştür [1]. Zaman içerisinde bilim adamları ve araştırmacılar tarafından birbirleriyle daha kolay iletişim kurabilmek için geliştirilen internet, 1990'lu yıllar itibariyle yazılım ve diğer unsurlar tarafından da tamamlanarak piyasaya sürülmüş ve ticarileştirilmiştir. Günümüzde ise, artık insanların herhangi bir kurum veya kuruluşa bağlı olmadan istedikleri an erişebildikleri bir ortam haline gelmiştir. İnternetin büyük gelişimi, insanların hayatını ekonomik, kültürel, siyasi ve daha birçok açıdan önemli ölçüde değiştirmiştir. Bu yeni teknolojik gelişim ve değişiklikler sayesinde, insanlar sosyal medya olarak adlandırılan bilgi paylaşım sayfaları (blog), sosyal ağlar ve birçok yeni kavram ile tanışmıştır. Hayatımıza giren bu yeni kavramlar ile insanlar, mekân ve zaman kavramı olmaksızın daha kolay, düşük maliyetli ve etkili iletişim kurmuştur.

Sosyal ağ ve sosyal medya kavramları genellikle birbirinin yerine kullanılsa da, aralarında bazı farklılıklar vardır. Sosyal ağ, sosyal medyanın genel bir kullanım alanıdır. Sosyal ağların kullanımı ile insanlar birbiri ile iletişime geçerler, resim, video paylaşırlar ve birbirlerine çevrimiçi iletiler gönderirler. İletişim kurabilmek amacı ile her geçen gün yeni kullanıcıların eklendiği sosyal ağlar giderek büyüyen bir platforma dönüşmektedir. Sosyal medya ise, insanların içerik oluşturduğu, paylaştığı, yer imi ve ağ oluşturdukları daha geniş kapsamlı bir iletişim aracıdır. Kaplan and Haenlein [2] sosyal medyayı "Web 2.0 ideoloji ve teknolojiye göre oluşturulan ve kullanıcılar tarafından içerik oluşturulmasına ve değiştirilmesine izin veren internet temelli uygulamaların bir grubu" olarak tanımlamıştır [2]. Kapsam ve işlevsellikleri çok çeşitli ve zengin bir sosyal medya ekolojisi vardır. Sosyal medya, sosyal ağlar (Facebook, Myspace, Ning, Orkut vb.), video paylaşım platformları (Youtube, Vimeo, Dailymotion vb.), resim paylaşım platformları (Flickr, Wallspace, ShutterBox vb.), Wiki (Wikipedia, Webopedia, Wetpaint vb.), iş sosyal sayfaları (Facebook, Google+, LinkedIn vb.), blog sayfaları (WordPress, TypePad, Blogger vb.), araştırma sayfaları (Google, Yelp, CitySearch, Yahoo vb.), mikro blog (Twitter, Tumblr, Friend Feed vb.), sosyal imleme sayfaları (Pinterest, Reddit, Digg vb.) olarak farklı kategorilere ayrılabilir. Sosyal medya ve sosyal medyanın kapsadığı ağ yapıları Şekil 1.1'de verilmiştir.

Sosyal medyanın günlük hayatımızın vazgeçilmez bir parçası oluşu ile araştırmacıları da sosyal medya da yönlendirmiştir. Bilimsel bir araştırma alanı olarak, sosyal medya kavramı literatürde 2000'li yıllarda ortaya çıkmış ve o tarihten beri hızla gelişimini sürdürmektedir.

Araştırmacılar, sosyal medyanın toplum üzerindeki olumlu ve olumsuz etkilerini, gelişen teknoloji ile ortaya çıkan sosyal medya problemlerini incelemek için her geçen gün çalışmaktadır.



Şekil 1.1. Sosyal medya ve kapsamı

Sosyal medya, hızlı olması, kullanım kolaylığı sağlaması, bilginin hızlı bir şekilde yayılımını sağlaması nedeniyle kültür, politika, eğitim, sağlık ve akademik gibi birçok alanda akımları belirleyen bir rehber konumundadır. Günümüzde, sosyal medyaya katılmanın faydaları yalnızca basit bir sosyal iletişim ağı oluşturmakla sınırlı değildir, aynı zamanda itibar ve kariyer fırsatları elde etmek veya doğrudan ya da dolaylı olarak maddi gelir elde etmek anlamında da önemli fırsatlar sunmaktadır [3]. Ayrıca, iş yerlerinin birçoğu, sosyal ağları bir tanıtım aracı olarak kullanarak ürün veya hizmetlerini tanıtmaktadır. Farklı amaçlar için kullanılan sosyal medyanın önde gelen bazılarının Temmuz 2019 tarihindeki aktif kullanıcı sayısı Tablo 1.1’de verilmiştir [4].

Tablo 1.1. Sosyal medyadaki aktif kullanıcı sayıları

Sosyal medya	Aktif kullanıcı sayısı (milyon)
Facebook	2,375
YouTube	2,000
Twitter	330
LinkedIn	310
Instagram	1,000
Pinterest	265

Sosyal medyadaki kullanıcı faaliyetlerinin devam eden büyümesine paralel olarak kullanıcılar tarafından oluşturulan verinin hacmi artmaktadır. Hacmi büyüyen verinin kontrol edilmesi zorlaştığı için sosyal medya, kullanıcıları olumsuz etkileyip, temelsiz ve çelişkili ifadeler ile kafalarını karıştırmaktadır. Bilgi çağında, büyük bilgilerin dikkatli bir biçimde analiz edilmeden ve değerlendirilmeden yayılması bilgiyi bir yük haline getirmiştir. Yapılan bir araştırmada birçok insanın Twitter'daki yanlış bilgileri retweet ettiğini bildirmiştir [5]. Yanlış bilgi, tüm bilgi kategorilerindeki gerçeklerden daha hızlı, uzak, derin ve geniş bir alana yayılır. Özellikle, sosyal medyanın artan popülerliği insanların haber tükettikleri platformları değişmesi ile televizyon, radyo, gazete gibi geleneksel haber ortamları yerini Çevrimiçi Sosyal Medya (ÇSM) ortamlarına bırakmıştır. Haberlere sosyal medya üzerinde erişilmesi, geleneksel ortamlara göre hız, maliyet ve erişim bakımından birçok avantaja sahip olmasına rağmen, sosyal medyadaki haberlerin kalitesi geleneksel ortamlara göre oldukça düşüktür. Gittikçe daha fazla insanın sahte haberleri okuması ve gerçekmiş gibi kabul etmesi politika, finans, askeri, bilim, eğitim alanlarını önemli ölçüde etkileyerek topluma büyük zararlar verebilir.

1.1. Tezin Amacı

Bu tez çalışmasında öncelikle bilgisayar teknolojileri ile birlikte yaşamımıza giren sosyal medya ve sosyal ağ kavramları, aralarındaki ilişkiler ve farklı kullanım alanları anlatılmaya çalışılmıştır. Daha sonra tez konusunun temel konusu olan sahte haber tespiti problemi, problem ile ilişkili kavramlar, özellikleri, tespit edilmesi sırasında yaşanan güçlükler ve probleme çözüm getirmek için literatürde yer alan farklı bakış açıları incelenerek derin bir analiz yapılmıştır. Literatürde yapılan çalışmalar, sahte haber tespiti problemi için içerik ve bağlam özelliklerini göz önüne alınarak makine öğrenmesi ve derin öğrenme tabanlı yöntemler kullanmışlardır. Bu çalışmada ise popüler ve çok yeni olan bir ÇSM problemi olan sahte haber tespiti problemi ilk defa bir optimizasyon problemi olarak ele alınmıştır. Çalışmada Gri Kurt Optimizasyonu (GKO) ve Salp Sürüsü Optimizasyonu (SSO) algoritmaları sahte haber tespiti problemine uyarlanmış ve dört farklı veri kümesi üzerinde yöntemlerin etkinliği test edilmiştir. Ayrıca bu tez kapsamında yeni adaptif optimizasyon algoritmalarının geliştirilmesi amaçlanmıştır. Bu amaçla Salp Sürüsü Optimizasyon algoritmasına atalet ağırlığı adında yeni bir kontrol parametresi eklenerek iki farklı adaptif yöntem önerilmiştir. Önerilen yeni yöntem de dört veri kümesi üzerinde test edilerek, yöntemlerin başarıları farklı metrikler üzerinden karşılaştırılmıştır.

Bu tezin literatüre katkıları genel olarak şu şekilde özetlenebilir:

- Sosyal medya ve çevrimiçi sosyal medya ile ilgili temel bilgiler sunmaktadır.

- Bu çalışmada sosyal medyada sahte haber tespiti ile ilgili literatürde yer alan çalışmalar ve önerilen yöntemler bir araya getirilerek farklı kategorilere ayrılmıştır. Problem ile ilgili detaylı bir analiz yapılmıştır.
- Optimizasyon, metasezgisel yöntemler, veri madenciliği ve sosyal medyada sahte haber tespiti problemi hakkında temel bilgiler sunulmaktadır.
- Bu çalışmada çevrimiçi sosyal medyada sahte haber tespiti probleminin sınıflandırılmasında daha önce hiç kullanılmayan optimizasyon yöntemi kullanılmıştır.
- Bu tez kapsamında ayrıca iki yeni adaptif optimizasyon algoritması önerilmiştir.
- Tez kapsamında geliştirilen yöntem daha çok yeni olmasına rağmen umut verici sonuçlar elde edilmiştir.

1.2. Tezin Kapsamı

Bu tez çalışması yedi temel bölümden oluşmaktadır. Birinci bölümde, internetin hızlı gelişimine paralel olarak ortaya çıkan sosyal medya ve sosyal ağ kavramları hakkında temel bilgiler verilmiştir ve tezin amacı açıklanmıştır.

İkinci bölümde, veri madenciliği hakkında temel bilgiler verilmiştir.

Üçüncü bölümde, optimizasyon ve metasezgisel algoritmalar hakkında bilgiler verilmiştir.

Dördüncü bölümde, sahte haber tespiti problemi, problem ile ilişkili kavramlar, özellikleri, tespit edilmesi sırasında yaşanan güçlükler ve probleme çözüm getirmek için literatürde yer alan farklı bakış açıları incelenerek sahte haber tespiti problemi hakkında detaylı bir analiz yapılmıştır.

Beşinci bölümde, kullanılan optimizasyon algoritmaları, tez kapsamında önerilen adaptif optimizasyon algoritmaları, denetimli yapay zekâ algoritmaları, veri ön işlem adımları, sınıflandırma için kullanılan sahte haber tespiti veri kümeleri ve elde edilen deneysel sonuçları karşılaştırmak için kullanılan istatistiksel test yöntemi hakkında ayrıntılı bilgiler verilmiştir.

Altıncı bölümde, 2 tanesi bu tez çalışması kapsamında önerilen yeni adaptif optimizasyon algoritması olmak üzere 4 farklı optimizasyon algoritması ile 30 tane denetimli yapay zekâ algoritmasının sahte haber tespiti için kullanılan veri kümelerinde elde ettiği test sonuçları karşılaştırmalı verilmiştir. Elde edilen sonuçların istatistiksel olarak değerlendirilmesi için Friedman testi kullanılmış ve test sonuçları da bu bölümde verilmiştir. Ayrıca kullanılan optimizasyon yöntemleri için kutu grafiği gösterimleri ve tüm yöntemlerden elde edilen sonuçların dört farklı performans değerlendirme kriterine göre değerlendirildiği grafiksel gösterimler de bu bölümde yer almaktadır.

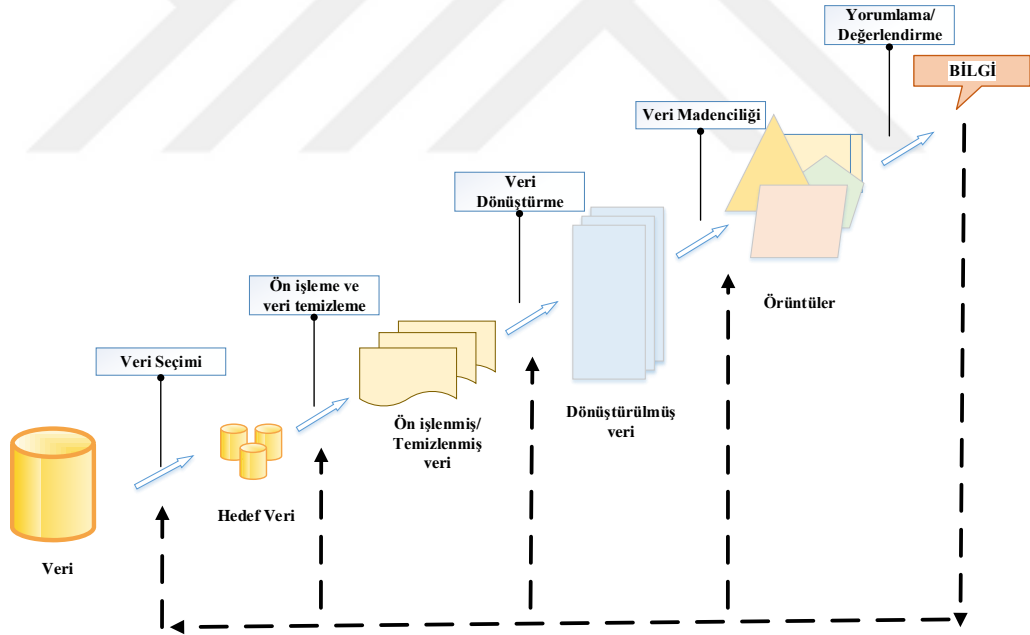
Yedinci bölümde, tezde yapılan çalışmalar ve elde edilen sonuçlar değerlendirilmiş, ileriye dönük çalışmalar için önerilerde bulunulmuştur.

2. VERİ MADENCİLİĞİ

Teknolojinin muazzam gelişimi ile birlikte elektronik ortamlarda depolanan veri miktarı hızla artmaktadır. İnsan analizi ve soyutlama, hem büyük boyutlu özelliklere sahip hem de çok sayıda örneğe sahip olan bu büyük veriler için uygun değildir. Ayrıca verilerin büyüme hızı, el ile yapılan analiz hızının çok üzerindedir. Bu durum, insanlara yararlı bilgileri toplamak, büyük veri kümelerini otomatik olarak ve akıllıca analiz etmeye yardımcı olmak için yeni tekniklere ve araçlara ihtiyaç duyulmasına neden olmuştur. Artan ihtiyaç, veri madenciliği kavramını ortaya çıkarmıştır.

Veri madenciliği ile ilgili birçok tanım yapılabilir. Fakat en genel ifade ile veri madenciliği, büyük bir veri kümesinden anlamlı ifadelerin çıkarılması, keşfi ve analiz edilmesi sürecidir [6].

Veri madenciliği, büyük veri kümelerinde, bilgi edinmenin en güçlü yollarından biridir ya da makine öğrenmesi ve yapay zekâ teknikleri yardımıyla veriler arasındaki temel ilişkileri saptamanın en iyi yaklaşımlarından biridir.



Şekil 2.1. Veri madenciliği adımları

2.1. Veri Madenciliği Süreci

Veri madenciliği birbirini takip eden adımlardan oluşan bir süreçtir ve bu süreç 5 adımda incelenebilir:

Bu adımlar Şekil 2.1'de gösterilmiştir.

- Veri seçimi
- Ön işleme ve veri temizleme
- Veri dönüştürme
- Veri madenciliği algoritmasının uygulanması
- Sonuçların yorumlanması / değerlendirilmesi

2.1.1. Veri Seçimi

Büyük miktarda verinin tek bir veri kümesinde veya veri ambarında birleştirilmesi veri madenciliği uygulamaları için önemlidir. Bu adımda, veri madenciliği yapılacak uygulama alanıyla ilgili bir veri kümesi oluşturmak için birçok farklı veri kaynağından veri seçilir. Bu adım veri madenciliği uygulamalarında en fazla zaman alan adımdır.

2.1.2. Ön İşleme ve Veri Temizleme

Gerçek dünyada, bazı veriler genellikle ham, gürültülü, eksik ve tutarsızdır. Verinin kalitesinin yüksek olması, veri madenciliği algoritmasının güvenilir sonuçlar üretebilmesi için gereklidir. Ön işleme ve veri temizleme adımı, gürültü olarak adlandırılan eksik verileri temizleme işlemleri yapılır. Veri temizleme işlemleri sayesinde; eksik değerler doldurulur, gürültüler düzeltilir, aykırı değerler belirlenerek, silinir ve tutarsızlıklar giderilir [7].

2.1.3. Veri Dönüştürme

Dönüşüm süreci verileri değiştirir ve birleştirerek bilgiyi keşfetmeye uygun yapısal biçime dönüştürür. Dönüşüm süreci bir takım alt süreçler kullanarak bir sonraki adım için veriyi uygun hale getirir. Bu alt süreçler aşağıda sıralanmıştır [8]:

- Aykırı verileri veriden kaldırır.
- Düşük seviyeli verilerin üst düzey kavramlarla değiştirerek verileri genelleştirir.
- Öznitelik verilerini ölçeklendirerek ve belirli bir aralıkta ayarlayarak veriyi normalleştirir.
- Yeni özellikler bulmak ve belirli nitelik gruplarına ekleyerek bilgi keşfi sürecini kolaylaştırır.

2.1.4. Veri Madenciliği Algoritmasının Uygulanması

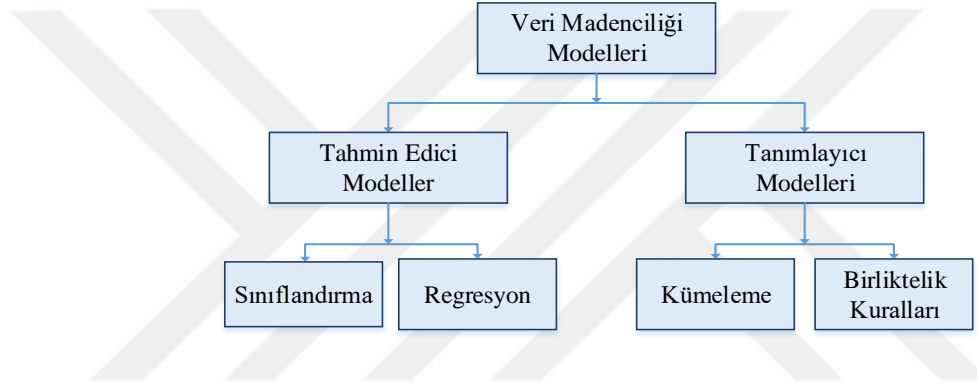
Önceki adımlar kullanılarak hazır hale getirilen veriye çalışmanın amacına göre belirlenen bir ya da daha fazla veri madenciliği algoritması uygulanarak sonuçlar elde edilir. Gerekli durumlarda farklı yöntemlerde birleştirilerek veriye uygulanabilir.

2.1.5. Sonuçların Yorumlanması / Değerlendirilmesi

Veri madenciliği algoritması, veriler üzerinde uygulandıktan sonra sonuçlar elde edilir ve bu sonuçlar değerlendirilir, yorumlanır ve çalışmanın doğru bir şekilde sonuçlandırılıp sonuçlandırılmadığı araştırılır. Bu aşamada birden fazla yöntem kullanılmışsa, bu yöntemlerin karşılaştırması yapılır.

2.2. Veri Madenciliği Modelleri

Veri madenciliğinde kullanılan modeller Şekil 2.2’de görüldüğü gibi tahmin edici modeller ve tanımlayıcı modeller olmak üzere iki ana başlık altında toplanabilir.



Şekil 2.2. Veri madenciliği modelleri

2.2.1. Tahmin Edici Modeller

Tahmin edici modeller belirli bir özelliğin açık değerini tahmin eder. Model sınıf üyeliğine göre tahminde bulunuyorsa sınıflandırma modeli veya basitçe sınıflandırıcı denir. Model çok çeşitli olası değerlerden bir sayı öngörüyorsa, o zaman regresyon modeli olarak adlandırılır.

Sınıflandırma

Sınıflandırma, farklı durumlar arasındaki ilişkiyi ve farklı nesnelerin özelliklerini anlamak için kullanılan en popüler veri madenciliği modellerinden biridir. Sınıfı bilinmeyen nesnelerin sınıfını tahmin etmek için sınıfları tanımlayacak bir modele ihtiyaç duyulmaktadır. Bu amaçla veri madenciliği bir dizi eğitim verisini analiz ederek bir model bulma sürecini kapsar. Sınıflandırmayı gerçekleştirmek için, hedef veri kümesi rastgele olarak eğitim seti ve test seti olarak adlandırılan iki özel ve kapsamlı kümeye bölünmüştür. Eğitim verilerinden, koşullu özellikler ve karar nitelikleri arasındaki ilişkiyi inceleyerek, kavramları tanımlamak için bir model (veya bir işlev) türetilmiştir. Elde edilen model, tahmin (eğer... öyleyse) kuralları, karar ağaçları, matematiksel formüller veya sinir ağları gibi çeşitli şekillerde sunulabilir. Model, test veri kümesindeki her veri

örneğinin sınıfını tahmin etmek için kullanılır. Sınıflandırma görevinin temel amacı, sınıflandırma doğruluğu oranını (doğru tahmin sayısının test veri kümesindeki toplam tahmin sayısına oranı) en üst düzeye çıkarmaktır [9].

Regresyon

Regresyon, çok sayıda veri öğelerinin, gerçek değerli tahmin değişkenlerine eşlendiği bir öğrenme yöntemidir. Farklı regresyon uygulamalarından birinde, bir ormanda mevcut olan biokütle miktarının tahmin edilmesi, bir diğerinde hastanın tanı testlerine ait veri setleri ile hayatta kalıp kalamayacağına öngörülmesi bir başkasında ise yeni bir ürünün tüketici talebini karşılmasına yönelik tahmin etme ihtimallerini ortaya koymaktadır. Tüm bu uygulamalarda kullanılan model sürekli bir hedefi tahmin etmek için eğitilmektedir. Regresyon görevleri genellikle nicel sınıf etiketine sahip sınıflandırma görevleri olarak değerlendirilmektedir. Tahmin yöntemleri, Doğrusal Olmayan Regresyon (Nonlinear Regression-NLR) ve Doğrusal Regresyondur (Linear Regression-LR) [10].

2.2.2. Tanımlayıcı Modeller

Bu modeller mevcut verilerdeki örnekleri tanımlar ve temel olarak anlamlı alt gruplar oluşturmak için kullanılır. Benzer örneklerin bir araya toplanarak veri karmaşıklığının azaltılması kümelemeyi oluşturur. İki ya da daha fazla örneğin ne kadar sıklıkla birleştiği yani benzerlik tespitlerinin yapılması ise birliktelik kuralı olarak bilinir.

Kümeleme

Kümeleme, bilinen bir sınıf etiketine danışmadan veri örneklerinin birlikte gruplandırıldığı sınıflandırmaya benzer bir veri madenciliği görevidir. Kümelemede, veri gruplamaları önceden tanımlanmamıştır, bunun yerine, gerçek verilerde bulunan özelliklere göre veri nesneleri arasındaki benzerlikleri bularak üretilirler. Bu benzerliğe dayanarak, veri kümesi farklı grup veya kümelere ayrılır. Gruplar içerisindeki nesnelere benzerliği grup dışındaki nesnelere benzerliğinden daha fazladır. Başka bir ifadeyle, iyi bir kümeleme algoritması küme içi benzerliği en üst düzeye çıkarmalı ve küme içi benzerliği en aza indirmelidir [11].

Birliktelik Kuralları

Kural keşfi, büyük veri kümelerindeki veri maddeleri arasında ilginç ilişkileri keşfeden bir süreçtir. Bu ilişkiler her zaman verilen veri kümesinde yer alan bilgileri temsil eder. Bu bilgi önerme mantığına göre (eğer... öyleyse) olarak ifade edilir. Bu kurallara genellikle birliktelik kuralları denir. Bir birliktelik kuralında, önceki bölüm, verilen veri kümesinde bulunan bir öğeyi

veya ögelerin bir kombinasyonunu temsil eder ve sonraki bölüm, önceki ögeyle birlikte bulunan bir ögeyi temsil eder [11].



3. OPTİMİZASYON

3.1. Optimizasyona Genel Bakış

Optimizasyon, en iyileme anlamına gelmektedir. Bir diğer tanımı ile belirli şartları sağlayacak şekilde mümkün olan en iyi çözümü elde etme işlemidir [12]. Belirli sınırlar içinde, bilinmeyen parametre değerlerinin bulunmasını içeren bir problem ise optimizasyon problemi olarak tanımlanır.

Tek amaçlı, doğrusal olmayan, kısıtlı bir minimizasyon problemi matematiksel olarak Denklem 3.1-4'deki gibi ifade edilebilir.

$$\text{Minimize: } f(x) \quad (3.1)$$

$$\text{Kısıtlar: } g_i(x_1, x_2, x_3, \dots, x_{n-1}, x_n) \geq 0, \quad i = 1, 2, \dots, m \quad (3.2)$$

$$h_i(x_1, x_2, x_3, \dots, x_{n-1}, x_n) = 0, \quad i = 1, 2, \dots, p \quad (3.3)$$

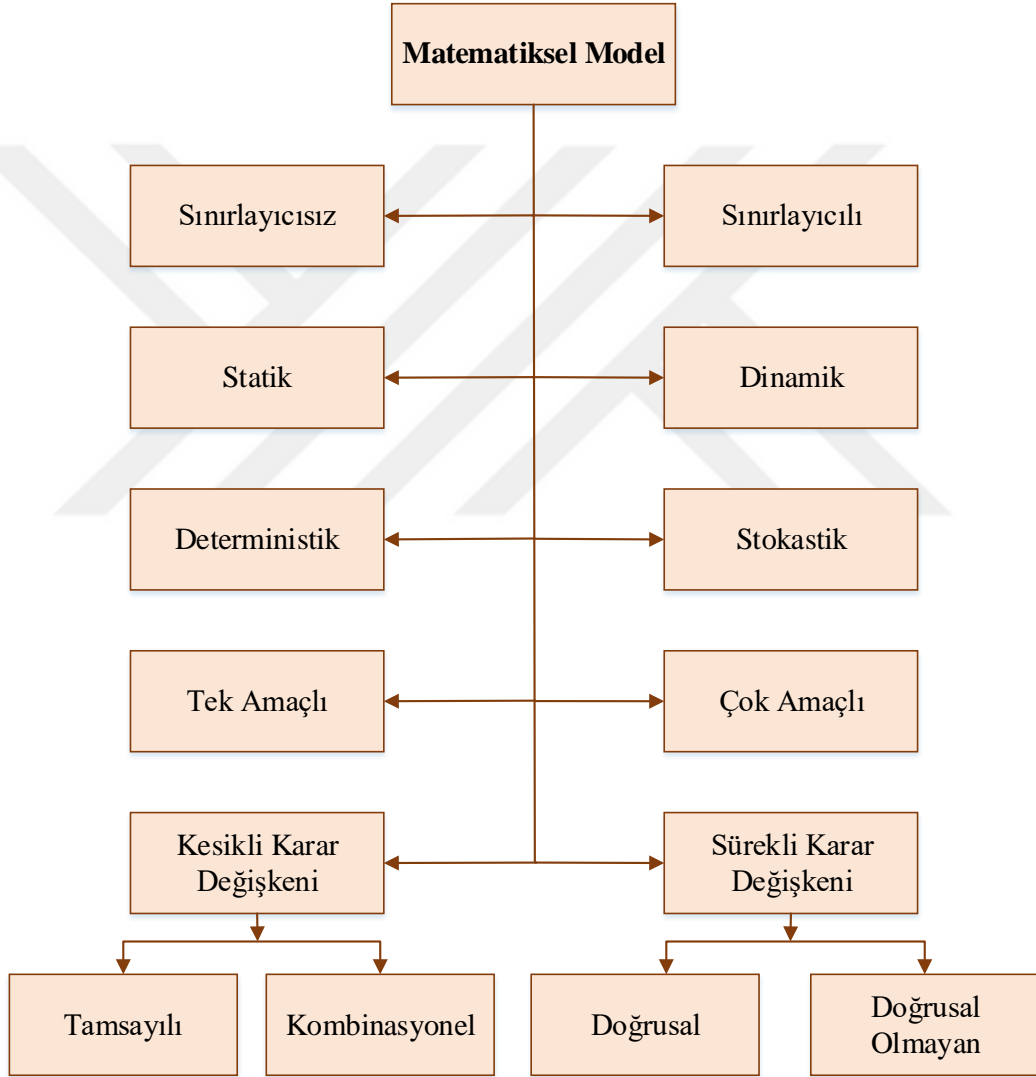
$$lb_i \leq x_i \leq ub_i, \quad i = 1, 2, \dots, n \quad (3.4)$$

$f(x)$, problemin amaç fonksiyonudur. Verilen denklemlerde n değişkenlerin sayısını ifade eder. m eşitsizlik kısıtlayıcılarını, p eşitlik kısıtlayıcılarını, lb_i i . değişkenin alt sınırını ve ub_i ise i . değişkenin üst sınırını gösterir.

Optimizasyon probleminin karar değişkenleri üzerinde hiçbir sınırlama yoksa sınırlayıcısız, en az bir sınırlama olması halinde ise matematiksel model sınırlayıcı olur. Eğer problem bir tek dönem için çözülecekse statik model, birden fazla dönem göz önüne alınarak çözülecekse dinamik model kullanılır. Modelin algoritmada işletilmesi sırasında, kesin parametre veya girdiler kullanılıyorsa model deterministik, olasılık özelliği varsa stokastiktir Modelin birden fazla amacı varsa, çok amaçlı problem olarak adlandırılır. Sürekli optimizasyon problemlerinde, karar değişkenlerinin tamamı pozitif reel değerler alırken kesikli optimizasyon problemlerinde bütün karar değişkenlerinin tamsayı değerler almasına gerek yoktur. Karar değişkenlerinin kombinasyonel biçimde ise kombinasyonel optimizasyon problemleri ortaya çıkar [13]. Şekil 3.1'de optimizasyon problemlerinin matematiksel modelleri ifade edilmiştir.

Optimizasyon yöntemlerinin en iyi performansı verebilmesi için amaç ve modelin en doğru şekilde belirlenmesi gerekir. Bunun için matematiksel modellemeye ihtiyaç duyulur. Birçok karmaşık sistem için, matematiksel modelin oluşturulması zordur. Model kurulmuş olsa bile çözüm zamanının uzun oluşu nedeniyle kullanılamamaktadır. Bu tip büyük ölçekli doğrusal olmayan problemler için klasik optimizasyon algoritmaları yetersiz kalmaktadır. Klasik optimizasyon yöntemlerinin çözüm stratejileri amaç fonksiyonun tipine, sınırlayıcı fonksiyonlara ve modellemede kullanılan değişkenlere bağlıdır. Ayrıca, bu yöntemlerin etkinliği, problemin

çözüm alanına, karar değişkenlerinin ve sınırlayıcıların sayısına bağlıdır. Bu yöntemlerin bir diğer eksikliği de, farklı karar değişkenleri, amaç fonksiyonları ve sınırlayıcılar varsa, klasik optimizasyon yöntemleri problem için genel bir çözüm veremez. Ancak birçok gerçek dünya problemi farklı karar değişkenleri, amaç ve sınırlayıcı fonksiyonları gerektirir. Klasik optimizasyonun birçok yetersizliğinden dolayı, metasezgisel optimizasyon algoritmaları geliştirilmiştir. Hesaplama güçlerinin iyi olması ve kolay uyarlanabilir olması, son yıllarda metasezgisel optimizasyon algoritmalarını birçok alanda popüler yapmıştır.



Şekil 3.1. Optimizasyon için matematiksel modeller [13]

3.2. Metasezgisel Algoritmalar

Gerçek dünya problemlerinde çözüm uzayı tüm çözümlerin değerlendirilemeyeceği kadar büyüktür. Aynı zamanda matematiksel modellerinin çıkarılması da zordur. Klasik optimizasyon

yöntemleri ile bu tip problemlere çözüm bulmak çok zaman alır. Çünkü problemin karmaşıklığı, olası her çözümü veya kombinasyonunu aramayı imkânsız hale getirir. Kabul edilebilir bir sürede çözüm uzayının değerlendirilerek iyi bir çözümün bulunması gerekir. Böyle durumlarda, tüm çözüm uzayındaki bazı çözümler değerlendirilir. Metasezgisel optimizasyon algoritmaları kabul edilebilir bir süre içinde, karmaşık bir problem için en uygun çözümleri üretmek için deneme yanılma yapan yöntemlerdir. Metasezgisel yöntemler en iyi çözümü bulmayı garanti etmezler. Bu yöntemlerin temel fikri, makul bir sürede işe yarayacak ve kaliteli çözümler sunmaktır [14].

Metasezgisel yöntemlere ihtiyaç duyulmasının sebepleri aşağıdaki gibidir:

- a) Optimizasyon problemi kesin çözümü bulma işleminin tanımlanamadığı bir yapıya sahip olabilir.
- b) Anlaşılabilirlik açısından metasezgisel algoritmalar karar verici açısından çok daha basit olabilir.
- c) Metasezgisel algoritmalar, öğrenme amaçlı ve kesin çözümü bulma işleminin bir parçası olarak kullanılabilir.
- d) Matematik formülleriyle yapılan tanımlamalarda genellikle gerçek dünya problemlerinin en zor tarafları (hangi amaçlar ve hangi sınırlamalar kullanılmalı, hangi alternatifler test edilmeli, problem verisi nasıl toplanmalı) ihmal edilir. Model parametrelerini belirleme aşamasında kullanılan verinin hatalı olması, metasezgisel yaklaşımın üretebileceği alt en uygun çözümden daha büyük hatalara sebep olabilir [15].

Metasezgisel algoritmalar aşağıdaki kriterler göz önüne alınarak değerlendirilebilir:

- *Çözüm Kalitesi ve Hesaplama Zamanı:* Herhangi bir algoritmanın etkinliğinin değerlendirilmesi için önemli bir kriterlerdir. Bir algoritma, ayarlanabilir parametreler kümesine sahip olmalı ve bu parametreler kullanıcıya önemlilik açısından hesaplama maliyeti ile çözüm kalitesi arasında bir vurgulamanın yapılabilmesine imkân sağlamalıdır. Başka bir ifade ile çözüm kalitesi ile hesap zamanı arasındaki ilişki kontrol edilebilmelidir.
- *Kod Basitliği ve Gerçeklenebilirlik:* Algoritmanın prensipleri basit olmalı ve genel olarak uygulanabilir olmalıdır. Böylece, algoritma uzman deneyimi olmadan ve problemin nasıl çözüleceği bilinmeden yeni problemlere kolaylıkla uygulanabilir.
- *Esneklik:* Algoritmalar modelde, sınırlamalarda ve amaç fonksiyonlarında yapılacak değişiklikleri kolayca karşılayabilmelidir.
- *Dinçlik:* Algoritma, başlangıç çözümüne bağlı olmaksızın her zaman kaliteli ve kabul edilebilir çözümler sunabilme yeteneğine sahip olmalıdır.
- *Basitlik ve Analiz Edilebilirlik:* Karmaşık algoritmalar, esneklik ve çözüm kalitesi açısından basit algoritmalarından daha zor analiz edilebilmektedir. Algoritma kolayca analiz edilebilir olmalıdır.

- *Etkileşimli Hesaplama ve Teknoloji Değişimleri:* Algoritma içinde insan-makine etkileşimini kullanma fikri çoğu sistemde yaygın olarak gerçekleştirilmektedir. Bilindiği gibi iyi bir kullanıcı ara yüzü herhangi bir bilgisayar sistemini veya algoritmayı daha çekici yapmaktadır. Bunun en önemli avantajı çözümlerin grafiksel olarak sergilenebilmesidir [15].

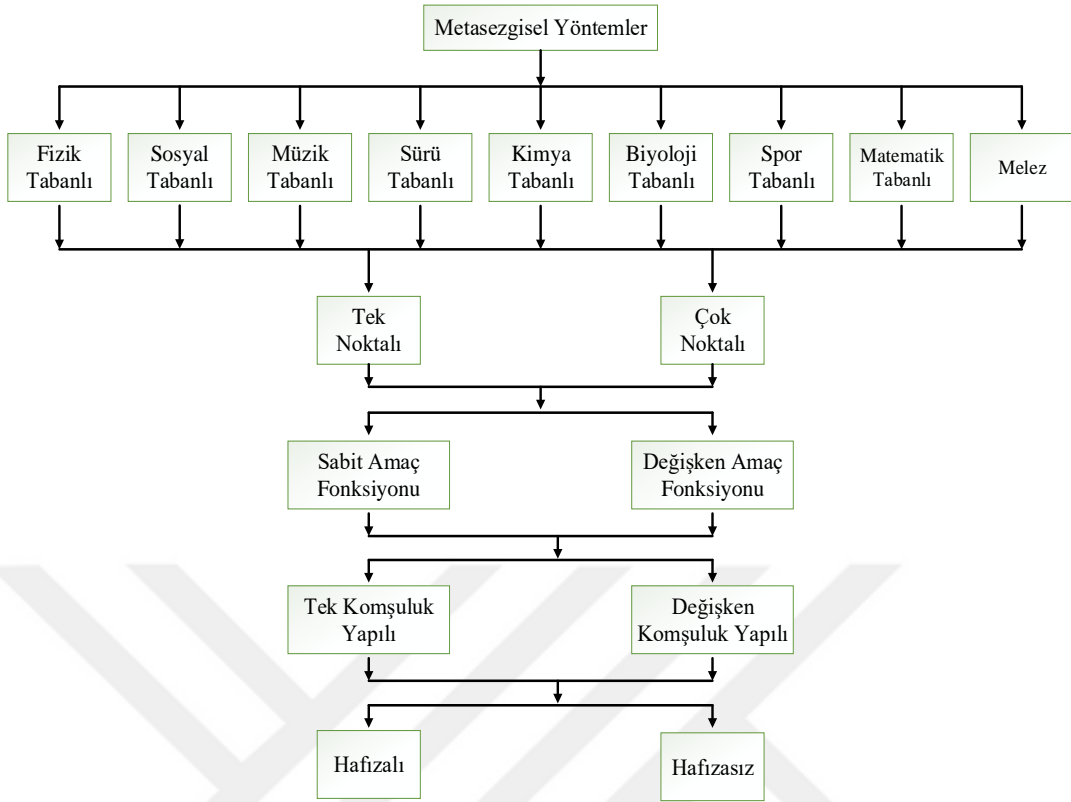
Optimizasyon algoritmaları, tek bir çözümden başlayarak bunu operatörlerle iletiyorsa tek noktali yöntemler olarak bilinir. Isıl işlem, tabu arama gibi yerel arama tabanlı metasezgisel algoritmalar bu kategoriye girer. Çok noktadan yani bir popülasyon üzerinden çözüme başlayarak farklı noktalarla optimizasyon yapıyorsa çok noktali ya da popülasyon tabanlı yöntemler denir. Genetik Algoritma (GA) ve Parçacık Sürüsü Optimizasyonu (PSO) bu yöntemlere örnek verilebilir.

Metasezgisel yöntemlerin bir kısmı problemi gösterirken amaç fonksiyonunu sabit tutar ve bu yöntemlere sabit amaç fonksiyonlu denir. Bazıları ise, yerel minimumdan kurtulmak için amaç fonksiyonlarını değiştirir ve değişen amaç fonksiyonlu olarak adlandırılır [13].

Metasezgisel algoritmaların birçoğu tek bir komşuluk yapısında çalışır ve tek komşuluk yapılı olarak sınıflandırılabilir. Bazıları ise arama işlemini sistematik bir şekilde değiştirerek birden fazla yerel arama yöntemiyle diğer çözüm alanlarına ulaşmaya çalışır ve değişken komşuluk yapılı şeklinde sınıflandırılabilir [13].

Metasezgisel yöntemler çalışırken daha önceki durumlar ya da en iyi durumlar hatırlanıyorsa hafızalı, hatırlanmıyorsa hafızasız şeklinde sınıflandırılır. Örneğin, tabu arama ve PSO hafızalı, GA hafızasızdır [13].

Metasezgisel yöntemlerin çoğu doğadan esinlenerek geliştirilmiştir. Metasezgisel yöntemler doğadan esinlendikleri alana göre fizik tabanlı (Parçacık Çarpışma Algoritması [16], Büyük Çöküş Algoritması [17], Elektromanyetizma Benzeri Algoritması [18]), sosyal tabanlı (Emperyalist Yarışmacı Algoritması [19], Parlamenter Optimizasyon Algoritması [20]), müzik tabanlı (Melodi Arama Algoritması [21], Müzikal Kompozisyon Algoritması [22]), sürü tabanlı (Parçacık Sürüsü Optimizasyonu [23], Balina Optimizasyonu [24]), kimya tabanlı (Kimyasal Reaksiyon Optimizasyonu [25], Yapay Kimyasal Reaksiyon Optimizasyon Algoritması [26]), biyoloji tabanlı (Genetik Algoritma [27]), spor tabanlı (Lig Şampiyonası Algoritması [28], Futbol Ligi Optimizasyonu [29]), matematik tabanlı (Sinüs Kosinüs Algoritması [30]) olmak üzere 8 farklı gruba ayrılabilir. Bunların dışında birden fazla algoritmanın bir arada kullanılmasıyla elde edilen melez yöntemler de bu kategorilere dâhil edilebilir. Tüm bu yöntemler Şekil 3.2’de verilmiştir.



Şekil 3.2. Metasezgisel yöntemler [13]

4. SAHTE HABER TESPİTİ

ÇSM'nin önem kazanması ile toplumumuzda haber ve bilgiye erişim ve tüketim biçimini değiştirmiştir. Mobil cihazların kullanımının artışı ve Wi-Fi ağlarına kolay erişim, Facebook, Instagram, YouTube ve Twitter gibi sosyal medya platformlarını haber ve eğlence sağlamak için kullanılan güçlü ortamlara dönüştürdü. Öyle ki cumhurbaşkanı ve başbakanlar, seçmenleri ve destekçileri ile iletişim kurabilmek için Twitter hesapları kullanıyor durumuna geldi. İnsanlar dünyada meydana gelen sosyal, ekonomik, politik ve bilimsel her türlü gelişmelere ilişkin haberlere sosyal medya aracılığıyla erişmektedir. Sosyal medyadaki haberlerin görüntü ve video bakımından zengin oluşu okuyucuların daha fazla ilgisini çektiği için nispeten daha güvenli olan televizyon ve gazete gibi geleneksel haber kaynaklarının önemini yitirmesine neden olmuştur. Ayrıca, sosyal medyada haberlere kolay erişilmesi, düşük maliyetli olması ve bilginin sosyal medyada daha hızlı yayılımı kullanıcılar arasında sosyal medyayı popüler hale getiren avantajlardır. Sosyal medya haber oluşturmak ve yaymak için avantajlar sağladığı gibi yanlış ve sahte haberler çoğalması için verimli bir alan haline geldi.

4.1. Sahte Haber Nedir?

Sahte haber kavramı onlarca yıldır varlığını sürdürmesine rağmen sosyal medyadaki önemli gelişim ile günümüzde sosyal ağ problemlerinin en dikkat çeken konularından biri olmuştur. Gazetecilikte bile sahte haberlerin evrensel bir tanımı olmamıştır. Net ve doğru bir tanım, sahte haber analizi ve ilgili çalışmaların değerlendirilmesi için sağlam bir temel oluşturmaya yardımcı olur. En genel ifade ile sahte haber “doğrulanabilir ve kasıtlı olarak yanlış olduklarından, okuyucular için potansiyel veya kasıtlı olarak yanıltıcı olan haberler” olarak tanımlanmaktadır [31, 32]. Biçim olarak geleneksel haber içeriğini taklit eden, ancak niyeti insanları yanıltmak olan üretilmiş bilgiyi temsil edebilirler.

4.1.1. Sahte Haberin Tanımı

#Tanım 1 (Sahte haberin geniş tanımı): Sahte haberler, kamuya açık şahıslar ve kuruluşlarla ilgili diğer bilgi türleri arasında geniş çapta alınacak malumatları, açıklamaları, konuşmaları, yayınları içeren, gerçeği yansıtmayan haberlerdir. Bu tanım, mevcut gerçek denetleyici web siteleri tarafından sağlanan çoğu sahte haberlerle ilgili araştırmayı ve veri kümesini desteklemektedir [33].

#Tanım 2 (Sahte haberin dar tanımı): Sahte haberler, bir haber kuruluşu tarafından yayınlanan kasıtlı ve doğrulanabilir şekilde yanlış aktarılmış olan bir haberdir [33].

Bu dar tanım, özellikle 2016 ABD başkanlık seçimlerinden sonra, halkın sahte haber algılarına hitap etmektedir. Kamuoyunu bilinçli bir şekilde daha iyi yanıltmak için gerçekmiş gibi gösterme çabasıyla oluşturulan aldatıcı haberler, dikkatsizce yapılan sahte haberlerden daha zararlı ve daha az ayırt edici olduğunu unutmamak gerekir. Dar tanım, hem haber gerçekliğini hem de niyetlerini vurgulamaktadır.

4.1.2. Sahte Haber İle İlişkili Kavramlar

Var olan çalışmalar sahte haberi çoğu zaman dezenformasyon, yanlış bilgi, aldatmaca, propaganda, hiciv, söylenti, tıklama yemi ve önemsiz haberler gibi kavramlarla birlikte kullanmaktadır. Aşağıda bu kavramlarla ilgili kısa bilgiler verilmiştir:

Yanlış bilgilendirme (Misinformation): Yanlış veya yanıltıcı olan bilgiler olarak tanımlanır [5]. Dürüst raporlama hataları veya yanlış yorumlar nedeniyle istemeden yayılabilir [34].

Dezenformasyon (Disinformation): İnsanları yanıltmak için kasıtlı olarak oluşturulan ve yasadışı konuşma biçimleri, şiddete teşvik gibi zararlı içeriklerdir [34].

Aldatmaca (Hoax): Bir gerçek veya yanlış habere dayanmadan yayınlanan haberlerdir [35]. Genel olarak, gerçekleşmemiş bir olaya inandırmayı ve insanları kandırmak için aldatmacalar içerirler.

Hiciv (Satirical): Hicivli haberler, haberleri taklit eden ve sosyal, politika ve suç dâhil olmak üzere çeşitli konuları kapsayan bir komedi türüdür [36]. Okuyucuları eğlendirmek veya eleştirmek amacıyla yazılan bu haberler bağlam dışında paylaşıldıklarında aldatmacalara benzer şekilde zarar verebilirler.

Propaganda: Aldatıcı ve tek taraflı mesajlar yoluyla hedef kitlelerin duygularını, görüşlerini ve eylemlerini etkilemeye çalışan bilgiler olarak tanımlanmaktadır [37].

Tıklama yemi (Click-bait): Tıklanma trafiği çekmeye ve reklam geliri yoluyla para kazanmaya yönelik düşük kaliteli gazetecilik olarak tanımlanır [38].

Önemsiz haberler (Junk news): Bu içerik çeşitli propaganda ve ideolojik olarak aşırı partizan veya komplocu politik haber ve bilgi biçimlerini içerir. Bu içeriğin çoğu kasıtlı olarak yanlış raporlama üretilir. Okuyucuları, kurumların, nedenlerin veya insanların ahlaki erdemleri veya başarısızlıkları konusunda ikna etmeye çalışır ve yorumu bir haber ürünü olarak sunar. Bu içerik, profesyonel gazeteci istihdam etmeyen kuruluşlar tarafından üretilir ve içerik, dikkat çekici teknikler, çok sayıda resim, hareketli görüntüler, aşırı büyük harf kullanımı, duygusal olarak yüklenen kelimeler ve resimler, güvenli olmayan genellemeler ve diğer mantıksal hatalar kullanır [39].

4.1.3. Sahte Haber Kavramında Anahtar Roller

Aldatma niyetine sahip sahte haberler incelenirken farklı varlıkların (bireyler veya kuruluşlar) oynadığı farklı roller bulunmaktadır. Bu rollere anahtar oyuncular da denilebilir. Sahte haberler için önemli olan anahtar roller aşağıdaki gibidir:

Muhalf Kışiler: Politik ya da sosyal gündemi olan kötü niyetli bireyler ve kuruluşlar genellikle sosyal botlar veya gerçek hesaplar kullanan sıradan sosyal medya kullanıcıları olarak görünerek, sahte haberleri desteklemenin yanı sıra sahte haber kaynağı da olabilirler [40].

Gerçek Denetleyiciler: Artan miktarda yanlış bilgi ile mücadele etmek amacıyla, Snopes ve PolitiFact gibi çeşitli gerçekleri denetleyen kuruluşlar, haber hikâyelerini ifşa etmek veya onaylamak için başlatılmıştır. Bu kuruluşlar, insan doğrulamasına dayanan “gerçeği kontrol eden gazeteciliği” temel alırken, işletmelere ve tüketicilere sahte haber tespiti çözümleri sunar. Aynı zamanda yapay zekâ kullanan web içeriklerine güvenilirlik puanları vermeyi amaçlayan Factmata gibi teknolojik şirketler tarafından daha fazla istenen otomatik teknolojik çözümler önerir [40].

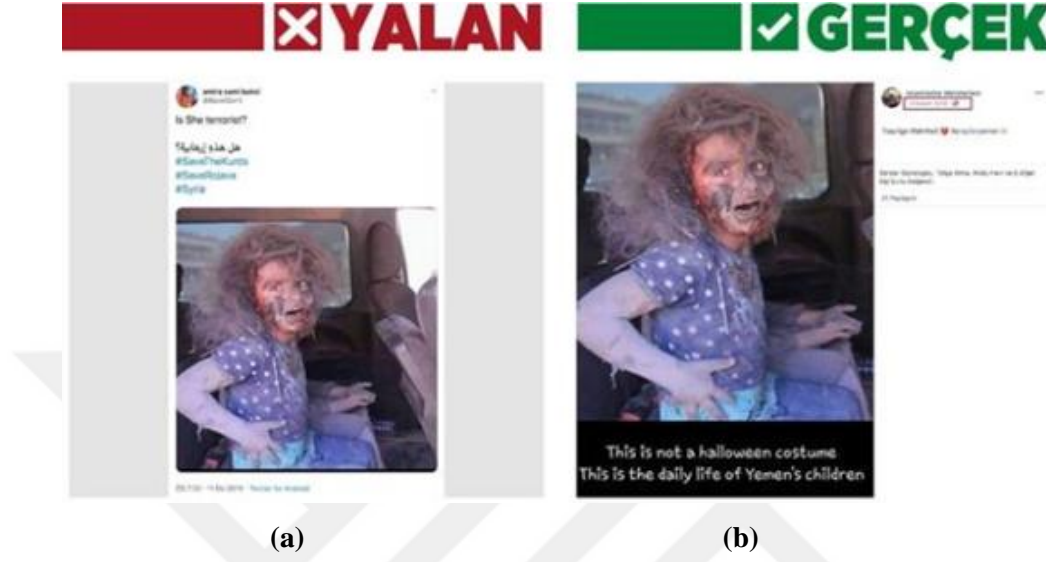
Duyarlılık: Sahte haberler, toplum üzerinde hareket ettirici etkisi bulunan çeşitli kişi ve kuruluşları da etkilemektedir. Örneğin; 2016 ABD başkanlık seçimleri sürecine tanık olanların genel fikirlerini etkilemeye yönelik sahte haber saldırılarına karşı saygın kurum ve kuruluşlar duyarlı olabilirler. Aksi durumda, sahte haberler tüm dünya için artan bir risk durumuna gelebilir. Örneğin; küresel ısınmanın ciddiyetini geçersiz kılan yanlış bilgiler, insanların iklim değişikliğini algılarını etkileyerek dünya için büyük risk oluşturmaktadır [41]

4.2. Sahte Haberlerin Toplum Üzerindeki Etkileri

İnsanlar, fikir ve kararlarını edindikleri bilgileri dikkate alarak şekillendirirler. Bu nedenle, elde ettiğimiz bilgi kaynaklarının güvenilir olması gerekir. İnsanların bilgi edinme biçiminin teknolojik gelişmeler ile birlikte sosyal medyaya kayması toplumsal hayatta bazı değişikliklere neden olmuştur. Ne yazık ki, sosyal medyadaki haberlerin içeriklerini değiştirerek okuyucu yanıltıcı içerikler oluşturan kötü niyetli kişiler nedeniyle sosyal medyadaki haberler güvenilemez duruma gelmiştir [42].

Sahte haberlerin sosyal medya ortamında hızlı bir şekilde yayılması insanlar ve toplumlar üzerinde ciddi olumsuz etkilere neden olmaktadır. Sahte haberler tüketicileri yanlış bilgilendirerek bir ürün hakkında önyargılı ya da yanlış düşünceleri kabul ettirebilirler. Örneğin, Rusya’da sahte öyküler yaymak için sahte hesaplar ve sosyal botlar oluşturulmuştur [43]. Bir diğer örnekte ise sosyal medya kullanıcılarını yanıltmak için kötü niyetli kişiler görsel ifadeler kullanmıştır. Türkiye’nin güney sınırındaki terörü yok etmek için yapılan Barış Pınarı Harekâtı aleyhine, terör örgütü yanlıları tarafından tarihte sembolleşmiş görüntüler paylaşılarak sosyal medyadaki kullanıcıları yanıltmaya çalışmışlardır. Tarihteki başka olay ve başka zamanlarda çekilmiş resimler

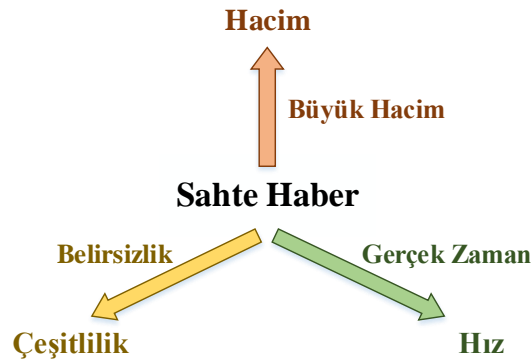
Barış Pınar Harekâtında çekilmiş gibi gösterilerek sosyal medya manipüle edilmeye çalışılmıştır. Şekil 4.1’de verilen görseldeki fotoğraf aslında 2018 yılında Yemen’de Suudi Arabistan liderliğinde koalisyon müdahalesi sırasında çekilmiştir [44].



Şekil 4.1. Pınar Barış Harekâtı sırasında paylaşılan sahte habere ait resimler. (a) Barış Harekâtı sırasında çekildiği iddia edilen resim. (b) 2018 yılında Yemen’de çekilen resim [44]

4.3. Sahte Haberleri Tespit Etmenin Önemi

Son yıllarda, sosyal medyanın hızlı ve dikkat çekici gelişimi, sahte haber sayısındaki geniş büyümeye tanık olmuştur. Günümüzde sahte haberler sinir bozucu, rahatsız edici, dikkat dağıtıcıdır. Gazeteciler, sivil toplum kuruluşları, toplumun bireyleri, hatta politikacıların da içinde bulunduğu geniş bir kesimi ilgilendiren sahte haber tespiti önemli bir sosyal ağ problemi haline gelmiştir. Bu yüzden sahte haber tanımlama için etkili bir tespit sistemi oluşturmak için önemlidir. Basit şekilde sahte haberlerin özellikleri Şekil 4.2’de gösterildiği gibi özetlenebilir.



Şekil 4.2. Sahte haber hacmi çeşitliliği ve hızı

• *Sahte haberlerin hacmi:* Herhangi bir doğrulama süreci olmadan, herkes kolayca internet üzerinden sahte haberler yazabilir [45]. Sahte haberler ve hikâyeler yayınlamak için bilerek yapılmış çok sayıda web sayfası vardır. Bu web siteleri genellikle meşru haber kuruluşlarına benzemektedir ve kasıtlı olarak, sahte, propaganda ve dezenformasyonu sık sık finansal veya politik kazanç için dağıtmak amacıyla oluşturulmuştur. Bu nedenle, kullanıcılara farkına bile varmadan İnternet üzerinden çok miktarda sahte içerik dağıtılmaktadır.

• *Sahte haberlerin çeşitliliği:* Söylentiler, yanlış bilgilendirmeler, politikacıların sahte ifadeleri gibi insan yaşamının her yönünü etkileyen ve sahte haberlere yakın anlamlı birçok tanımlama bulunması sahte haberlerin çeşitliliğini göstermektedir. Bu çeşitlilik ile sahte haberler insanların görüşlerini, çıkarlarını ve kararlarına hükmetmektedir. Ayrıca, insanların gerçek haberlere bakış açılarını değiştirmektedir. Bazı sahte haberler kasıtlı olarak okuyucuların özellikle gençlerin öz koruma bilincinden mahrum yaşlıların kafasını karıştırmak için oluşturulmaktadır. Örneğin, 2010'da Şili'deki deprem sonrasında Twitter'da paylaşılan söylentiler halkın paniğini artırmıştır.

• *Sahte haberlerin hızı:* Sahte haber yaratıcıları kısa ömürlü olma eğilimindedir [31]. Örneğin, 2016 ABD seçimleri sırasında birçok aktif sahte haber web sayfası kampanyadan sonra hemen yok olmuştur. Son yıllarda sahte haberlere daha fazla dikkat edildiğinden, sahte haber üreticileri tespit sistemleri tarafından tespit edilmekten kaçınmak için sürekli değildir. Ayrıca, sosyal medyadaki sahte haberlerin çoğu, çevrimiçi kullanıcılara daha fazla dikkat çekmek için güncel olaylara ve sıcak olaylara odaklanmaktadır. Sahte haberlerin sosyal medyada gerçek zamanlı olması çevrimiçi sahte haberlerin belirlenmesini daha da zorlaştırmaktadır.

4.4. Sahte Haber Belirlemenin Zorlukları

Sosyal medyadaki sahte haber tespiti problemi bir takım zorluklar sunar. Bu zorluklar aşağıdaki gibi özetlenebilir:

- Sosyal medyada birden fazla teknolojik platformun ve işletmenin bulunması ve her geçen gün hesaplama ve tasarım karmaşıklığının artması kaliteli bilgi hacminin azalmasına neden olmaktadır.
- Okuyucuları aldatmak ve geleneksel haber kanallarını taklit etmek için yanlış haberler kasıtlı olarak yaratılmakta ve sonuçta gerçek haber makalelerini yanlış haberlerden ayırmanın çok zor olduğu olumsuz bir senaryo ortaya çıkmaktadır [46].
- Yanlış haberlerin üretilme hızı ve miktarları, haberlerin dikkatli bir şekilde kontrol etme ve doğrulama olasılığının çok üzerindedir.

- Sahte haberlerin dinamik bir şekilde yayılması sosyal medyayı daha karmaşık hale getirmektedir. Yanlış bilgiler çok kısa bir sürede kullanıcılar tarafından erişilebilir ve kolaylıkla insanları etkileyebilir durumdadır.

4.5. Sahte Haber Tespiti ile İlgili Literatür Taraması

Sosyal medya problemleri arasında yeni bir araştırma alanı olan sahte haber tespiti problemi son yıllarda çok dikkat çekmiştir. Bu bölümde, bu çalışmanın da konusu olan “Sahte Haber Tespiti” ile ilgili geniş bir literatür taraması yapılmıştır. Bu çalışmada, literatürde sahte haber tespiti problemi için önerilen yöntemler, içerik temelli ve kapsam temelli olmak üzere iki farklı kategoriye ayrılarak incelenmiştir.

Sahte haberler konusu, yalnızca halkın ilgisini çekmemiştir, aynı zamanda akademik topluluktan da artan bir ilgi görmektedir. Bu bağlamda, sahte haber tespiti ile ilgili literatürü inceleme ve özetleme girişimleri olmuştur. Bu alandaki ilk çalışmalardan birinde, sahte haberler ile sosyal ve psikolojik teoriler arasında ilginç bir bağlantı olduğunu savunulmuştur. İnsanlar genellikle, ideolojik inançlarına uygun bilgileri araştırmaya, tüketmeye ve inanmaya eğilimlidirler. Bu fikir, çoğu zaman benzer düşüncelere sahip topluluklarda yanlış bilgilerin algılanması ve paylaşılmasıyla sonuçlanır [32].

Literatürü incelemek için yapılan bir diğer çalışmada ise, web üzerinde daha geniş bir yanlış bilgi kapsamı ele alınmıştır [47]. Araştırmacılar, yaptıkları bu inceleme çalışmasında, sosyal medyadaki sahte haberlerin yanı sıra, anket, e-ticaret platformlarındaki sahte incelemeleri ve Wikipedia gibi işbirlikçi platformlardaki aldatmacaları tartışmışlardır.

Sahte haber tespiti kavramının algılanması ve gelişmesinde etkili olan yeni teknolojilerin araştırıldığı bir çalışmada, aldatma içerikli haberlerin temel hedeflerini mevcut çalışmalar üzerinden göstererek, araştırmacılara yeni yöntemler için bir harita çıkarmak amaçlanmıştır. Sahte haber tespiti yöntemleri, dilbilimsel yaklaşımlar ve ağ yaklaşımları olmak üzere iki temel kategoriye ayrılmıştır. Dilbilimsel yaklaşımda, haber metinlerinin içerikleri çıkarılır ve dil kalıplarını aldatma ile ilişkilendirmek için analiz edilirken, ağ yaklaşımında, haberin üst verileri (metadata) veya yapılandırılmış bilgi ağı sorguları gibi ağ bilgileri, sahte haber tespiti yapmak için kullanılır.

4.5.1. İçerik Temelli Sahte Haber Tespiti Yaklaşımları

Haberlerden çıkarılan özellikler, haberin içeriğine bağlı olarak farklı özellikler barındırır. En genel anlamda, haberin başlığı, gövde metni, içerdiği görsel ve haber kaynağı haberin içeriğini oluşturur [49]. Manşet, haber makalesinin konusunu özetleyen kısa bir başlıktır. Haber ile ilgili detaylar ana metinde sunulur. İçerik tabanlı özellikler sahte haberlerin tespit edilmesi için önemli

ipuçlarıdır, bu nedenle sahte haberlerin temsili ve tespiti için sıklıkla kullanılır. Haber makalelerinin başlığı, gövde metni, içerdiği görsel ve kaynağını inceleyerek sahte haber tespiti yapılması içerik tabanlı sahte haber tespiti kavramı kapsamına girmektedir. Literatürde haber içerikleri dikkate alınarak geliştirilen yöntemler bu bölümde incelenmiştir. İçerik tabanlı yaklaşımlar ele aldıkları özelliklere göre metin tabanlı yaklaşımlar ve görüntü tabanlı yaklaşımlar olarak ayrılmıştır.

Metin Tabanlı Yaklaşımlar

Sosyal medyada yayınlanan haberlerde bilginin büyük bir kısmı metinsel öğeler ile doldurulur. Sahte haberleri tespit etmeye yönelik yaklaşımların büyük kısmı, sahte haberlerin metinsel verilerindeki kalıpları karakterize etmeye ve analiz etmeye dayanmaktadır. Araştırmacılar metinsel verileri göz önünde bulundurarak sahte haber tespiti yapmak için farklı yöntemler kullanmışlardır. Bu bölümün devamında, sahte haber tespiti için makine öğrenmesi, derin öğrenme ve diğer yöntemleri kullanan çalışmalar detaylıca verilmiştir.

• Makine Öğrenmesi

Perez-Rosas vd. [50] iki adımlı bir bakış açısından oluşan otomatik sahte haber tespiti modeli geliştirmişlerdir. Birinci adımda, ilki altı farklı haber alanından elde edilen diğeri ise ünlüleri kapsayan ağdan elde edilen iki farklı veri kümesi oluşturmuşlardır. İkinci adımda ise, N -gram, noktalama işaretleri, psikodilbilimsel özellikler, okunabilirlik ve sözdizimi özellikleri temsil eden özelliklerin birleşimine dayanan bir öğrenme sistemi geliştirmişlerdir. Geliştirdikleri sistemi test etmek için Support Vektor Machine (SVM) sınıflandırıcısını kullanmışlardır. Yazarlar, çalışmalarında otomatik ve elle sahte haber tespitinin ilişkili analizini de göstermişlerdir.

Potthast vd. [51] hiper-partizan ve sahte haberleri tespit etmek için stilometri tabanlı bir yöntem önermişlerdir. Araştırmacılar, konu tabanlı bir sınıflandırma için kelime çantası yöntemini kullanmışlardır. Karakterlerin N -gramları, durak kelimeler ve okunabilirlik puanının yöntemlerine dâhil etmişlerdir. Önerdikleri yöntemi Random Forest (RF) algoritmasını kullanarak test etmişlerdir.

Rubin vd. [52] yanıltıcı haberleri tespit edebilmek için hiciv ipuçlarını kullanarak sahte haber tespit problemini farklı bir bakış açısı ile almışlardır. Genel olarak çalışmalarında, hiciv ve mizahın temel kavramlarını ele alan ve daha sonra gazetecilik raporlarının tarzını ve biçimini taklit eden sahte haberleri ele almışlardır. Saçmalıklar, mizah, dilbilgisi, olumsuz duygular ve noktama işaretleri gibi 5 farklı özellikleri kullanarak SVM yöntemleri ile önerdikleri modeli test etmişlerdir.

Çevrimiçi sosyal ağ sitelerinde, yanlış bilgi yayılmasının saptanması ve önlenmesi için yapılan bir diğer çalışmada CURB adında etkili ve ölçeklenebilir çevrimiçi bir algoritma önerilmiştir. Stokastik diferansiyel denklemlerin stokastik çevrimiçi en iyi kontrolü ile atlamalar, hayatta kalma analizi ve Bayes çıkarımı arasında keşfedilmemiş bir bağlantı kurulmuştur. Önerilen

yöntem, Twitter ve Weibo'dan toplanan iki gerçek dünya veri kümesinde denenmiş ve algoritmanın yanlış bilgi yayılımını etkili bir şekilde azaltabileceği gösterilmiştir [53].

Son zamanlarda Facebook tarafından kullanıcıların sahte haberleri işaretlemelerini sağlayan ve bayrak olarak da isimlendirilen araçları kullanılarak yapılan bir diğer çalışmada, kullanıcıların bayrakları bir araya getirilerek, her gün haberlerin bir bölümü seçilir ve bir uzmana gönderilerek sahte olup olmadığı onaylanır. Uzman tarafından sahte olduğu belirlenen haberlerin yayılımının durdurulması ya da en aza indirgenmesi çalışmanın en temel amacıdır. Sahte haber tespit etmek için Bayes çıkarımı yapan bu yöntem aynı zamanda kullanıcı bayraklarını dikkate alarak kalabalık sinyallerini kullanır [54].

Sahte haber tespiti için yapılan bir diğer çalışmada metin içerikli haberlerden Term Frequency (TF) ve Term Frequency-Inverse Document Frequency (TF-IDF) kullanılarak özellik seçimi yapılmıştır. Ardından Stochastic Gradient Descent (SGD), SVM, Linear Support Vector Machine (LSVM), k -Nearest Neighbors (k -NN) ve Decision Tree (DT) kullanılarak bir öğrenme modeli oluşturulmuş ve önerilen model test verilerini tahmin etmek için kullanılmıştır. Önerilen yöntemi test etmek ve etkinliğini ölçmek için açık erişimli bir veri kümesi kullanılmıştır [55].

• **Derin Öğrenme**

Derin öğrenme, yapay sinir ağlarının çoklu katmanlardan oluşan ağları kullanarak öğrenme görevlerine uygulanmasıdır [56]. Makine öğrenmesinin yeni bir alanı olan derin öğrenme yöntemleri özellikle görüntü sınıflandırma, sosyal ağ analizi, metin madenciliği, bilgisayarlı görme, konuşma analizi ve doğal dil işleme gibi birçok farklı alanda dikkat çekmiştir. Derin öğrenme yöntemleri, karmaşık ve dinamik sosyal medya problemlerinde kullanmak için etkilidir. Literatürde, sosyal medyadaki sahte haberleri tespit etmek için derin öğrenme yöntemlerinin kullanıldığı birçok çalışma bulunmaktadır.

Bhatt vd. [57] sahte haber tespiti problemine etkili bir çözüm bulmak için sinirsel, istatistiksel ve dış özellikleri birleştiren yeni bir yöntem önermişlerdir. Çalışmada, sinirsel özellikleri derin tekrarlı modelden, istatistiksel özellikleri ağırlıklandırılmış N -gram kelime çantası modelinden ve diğer dış özellikleri ise mühendisliğin sezgiselliği ile elde etmişlerdir. Bütün özellikleri derin sinir katmanı ile birleştirerek, derin sinir ağları temelli sahte haber tespiti modeli geliştirmişlerdir.

Bir diğer çalışmada, 3 seviyeli hiyerarşik ilgi ağı aracılığıyla derin öğrenme temelli otomatik bir sahte haber tespit edici önerilmiştir. Önerilen yöntem, kelime, cümle ve başlığın her biri için bir seviyeye sahiptir ve haber makalesini etkin bir biçimde temsil edecek bir haber vektörü oluşturur. Vektörü oluşturmak için haber makalesi aşağıdan yukarıya hiyerarşik bir biçimde işlenir. Haberin

başlık kısmı az kelime içerdiği için kelime ve cümlelere göre sahte haberin tespit edilmesi için daha az ayırt edici özelliğe sahiptir [58].

Fang vd. [59] sahte haber tespiti yapmak için konvolüsyonel sinir ağları ve kendi kendine çok-kafalı dikkat mekanizmasının avantajlarını bir araya getiren hiyerarşik sinir ağı modeli geliştirmişlerdir. Kendi kendine çok-kafalı dikkat mekanizması modele kelimelerin uzaysal ilişkilerini öğrenmesinde yardımcı olur. Daha sonra dikkat mekanizmasını konvolüsyonel sinir ağları ile birleştirmişlerdir. Önerdikleri yöntemin geçerliliğini kanıtlamak için halka açık bir veri seti üzerinde deneyler yapmış ve sonuçları var olan çalışmaların yöntemleri ile kıyaslamışlardır.

Wang [60], sahte haber tespiti için melez konvolüsyonel sinir ağı modeli önermiştir. Bu modelde, ifade metni ve konuşmacının üst veri bilgileri olmak üzere iki giriş parametresi alınır. Metin verisi, cümledeki her bir kelime için sürekli düşük boyutlu temsil biçimleri elde etmek kelime gömme katmanı tarafından işlenir. Katmanın çıktısı özellik gösterimini üretmek için konvolüsyonel ve maksimum havuzlama katmanı tarafından işlenir. Benzer şekilde, konuşmacı üst verileri de, son özellik gösterim biçimlerini elde etmek için farklı bir konvolüsyonel katman ve çift yönlü bir uzun kısa süreli bellek katmanı ile işlenir. İki gösterim birleştirilir ve diğer katmanlarla uçtan uca eğitilmiş sınıflandırıcıyı besler. Derin öğrenme temelli konvolüsyonel sinir ağları, SVM ve Lojistik Regresyon algoritmaları ile karşılaştırılarak etkinliği test edilmiştir.

Bir diğer çalışmada, iki seviyeli konvolüsyonel sinir ağları kullanıcı yanıt üretici ile birlikte kullanılarak sahte haberlerin tespit edilmesi amaçlanmıştır. İki seviyeli konvolüsyonel sinir ağı, cümle ve kelime seviyesinde gösterilen haber makalelerinden anlamsal bilgiler toplar. Kullanıcı yanıt üretici ise, sahte haberlerin tespit edilmesini kolaylaştırmak için, geçmiş kullanıcı yanıtlarını kullanarak haber metinlerine kullanıcı yanıtı üreten bir model öğrenir. Önerdikleri yöntemin etkinliğini test etmek için gerçek haber makalelerini içeren Weibo erişime açık veri kümesini ve ilişkili kullanıcı yanıtlarını kullanmışlardır [61].

Girgis vd. [62] çevrimiçi metinlerde sahte haberleri tespit etmek için derin öğrenme bakış açısını temel alarak tekrarlayan sinir ağları ve uzun kısa süreli belleği kullanarak yeni bir model önermişlerdir. Önerilen modelde, ilk olarak metin üzerinde ön işlemler yapılarak veri modelde işlenmeye uygun biçime dönüştürülür. Çıkarılan özellikler, tekrarlayan sinir ağları ve uzun kısa süreli bellek modeline girdi olarak kullanılır.

Farklı bir çalışmada, graf yapılı veriler üzerinde çalışmak amacıyla tasarlanan yeni bir derin öğrenme sınıfı olan geometrik derin öğrenme yöntemi ile sahte haberlerin tespiti için yeni bir yöntem önerilmiştir. Geometrik derin öğrenme, farklı verileri (kullanıcıların aktiviteleri, sosyal ağ yapısı, haber yayılımı ve içeriği gibi) inceler, böylece içerik, sosyal bağlam ve yayılma tabanlı yaklaşımlar için birleştirici bir yapı oluşturur. Bu model 2013-2018 yıllarında Twitter ortamında yayılan gerçek ve sahte hikâyeler üzerinde denetimli bir şekilde eğitilmiştir. Önerilen yöntem ile yüksek performans elde edildiği deney sonuçları ile gösterilmiştir [63].

Ruchansky vd. [64] derin sinir ağıları kullanarak farklı alanlardan bilgi çıkarabilen ve haber makaleleri ile kullanıcı etkileşiminde geçici bağımlılıkları yakalayabilen, Capture (Yakalama), Score (Puanlama) ve Integrate (Birleştirme) aşamalarından oluşan yeni bir yöntem önermiştir. Yakalama adımı, tekrarlayan sinir ağıları kullanılarak, makalenin geçici gösterim biçimi çıkarılır. Geçici bağımlılıklar bir vektör olarak temsil edilir ve tekrarlayan sinir ağılarını girdi olarak besler. Kullanıcı özellikleri puanı hesaplamak için kullanılır ve en son adımda yakalama ve puanlama adımlarının sonuçları birleştirilerek bir vektör olarak sınıflandırıcıya verilir.

Bir diğer çalışmada, takviyeli öğrenmeyi bir nokta süreç ağı ile birleştirerek, sosyal ağlarda sahte haberlerle mücadele eden ilk çok aşamalı müdahale çerçevesi önerilmiştir. Sahte haberlerin ağ içinde yayılması, ek dışsal kontrol terimleri içeren çok değişkenli Hawkes süreci ile modellenmiştir. Metodlarının etkinliğini test etmek için Twitter verisi ve yapay veri kullanmışlardır ve umut verici sonuçlar elde etmişlerdir [65].

Dong vd. [66] hem içeriğin hem de yan bilgilerin ipuçlarını ele almışlardır ve bu ipuçlarından yararlanmak için bir hibrid dikkat modeli önermişlerdir. Haber içeriklerinden özellikler çıkarmak için çift yönlü Geçitli Tekrarlayan Ünite ve yan bilginin gizli özelliklerini çıkarmak için derin bir model kullanmışlardır. İki yöntemin çıktılarında iki gizli vektörü bir dikkat matrisinde birleştirmişlerdir ve vektörler üzerinde dikkat dağılımını öğrenmişlerdir. Son olarak, dağıtımı daha iyi sahte haber tespitini kolaylaştırmak için kullanmışlardır. İki gerçek dünya veri kümesindeki deneysel sonuçlar, yaklaşımlarının sahte haberleri algılama doğruluğunun çoklu taban hatlarından daha iyi performans verdiğini göstermiştir.

• Diğer Yöntemler

Sosyal medyadaki haber yayılımı sürecinde haberler, haber yayıncıları ve kullanıcılar arasındaki ilişkinin bilinmesi sahte haberlerin tespit edilmesini kolaylaştırmaktadır. Bu üç faktör göz önüne alınarak önerilen bir çalışmada sosyal bağlamdan yararlanmışlardır. Eş zamanlı olarak yayıncı-haber ve kullanıcı-haber ilişkilerini modelleyerek sahte haberleri tespit etmişlerdir. Önerilen yöntemin performansını değerlendirmek için BuzzFeed ve PolitiFact gerçek dünya verileri kullanılmıştır [67].

Fairbanks vd. [68] GDEL'T'ten toplanan makalelerden oluşan bir koleksiyonu kullanarak yanlış ve gerçek haberleri sınıflandırmayı amaçlamışlardır, etiketleri gerçekleri kontrol eden bir web sitesinden elle taramışlardır. Geleneksel metin özellikleri üzerinde bir sınıflandırıcı kullanan içerik tabanlı bir yöntem ve haber makalelerinin bağlantı yapısından oluşturulan bir graf üzerinde döngüsel inanç yayılımı uygulayan yapısal bir yöntem olmak üzere iki farklı yöntemi karşılaştırmışlardır.

Hosseini vd. [69] sadece haber içeriğine dayanarak, farklı yanlış haber kategorilerini (hicivden önemsiz haberlere kadar) ayırt etme problemini ele almışlardır. Altı farklı etikete sahip

Kaggle veri kümesini kullanmışlardır. Önerdikleri yaklaşımlar makaleler ve terimler arasındaki gizli ilişkileri ve terimler arasındaki uzamsal / bağlamsal ilişkileri yakalamayı amaçlayan dokümanların tensör belgelerinin ayrıştırılmasını içerir [69].

Görüntü Tabanlı Yaklaşımlar

Canlı ve kolay anlaşılabilir görüntüler, çoğu zaman metin ifadelerinden daha fazla dikkat çeker ve sosyal medya platformlarında insanlar arasında hızla yayılırlar. Bu nedenle haberlerin daha fazla okuyucu kitlesine ulaşması için, haber makalelerine görsel ifadeler eklenerek içerikleri dikkat çekici hale getirilir. Sahte haberlerin, sosyal medya kullanıcıları arasında yayılmasında görsel unsurlar çok önemlidir. Sahte haber oluşturucuları, okuyucuların ilgisini çekebilmek için haber içeriklerine genellikle sahte görüntü ve videolar eklemektedirler. Ancak, sosyal medyada yayılan bu tip görüntü ve videolar okuyuculara yanlış bilgi sunup, onları yanıltmaktadır. Bu tip görsel ipuçları ile sahte haberlerin farklı karakteristik özellikleri bulunabilir.

Literatürde, sosyal medyadaki sahte haberleri tespit etmek için görüntüyü temel alarak önerilen farklı yöntemler yer almaktadır.

Jin vd. [70] yaptıkları çalışmada bir takım görsel özellikler önererek, haberde geçen olayın görüntülerindeki gizli özellikleri ortaya koymuş ve bu şekilde sahte haberleri tespit etmeye çalışmışlardır. Aynı zamanda bir takım istatistiksel özellikler de kullanmışlardır. İstatistiksel özellikler ile haber olayının özellik bilgilerini ve görüntü istatistiklerini özetlemişlerdir. Böylelikle görüntü dağılım modellerini nicel olarak yakalamış ve performanslarını artırmışlardır [70].

Bir diğer çalışmada, gerçek fotoğraflardan oluşan büyük bir veri kümesi kullanılarak eğitilmiş görsel görüntüdeki değişiklik/bozuklukları tespit etmek için bir öğrenme algoritması önerilmiştir. Algoritma, otomatik olarak kaydedilen fotoğraf EXIF üst verilerini, bir görüntünün tutarlı olup olmadığını yani içeriğinin tek bir görüntüleme hattı ile üretilip üretilmeyeceğini belirlemek için bir modeli eğitmek amacıyla bir denetim sinyali olarak kullanır. Yazarlar, önerilen yöntemin eğitimde değişikliğe uğramış herhangi bir görüntü görmemesine rağmen, adli tıp kıyaslamaları için en gelişmiş performansı elde ettiğini ifade etmişlerdir [71].

4.5.2. Bağlam Temelli Sahte Haber Tespiti Yaklaşımları

Haber bağlamı genellikle haberin yazarını, yayınlanma zamanını, sayfa görüntülemelerini, popülerlik derecesini, aktarım oranlarını ve haber makalesine ilişkin yorumları içerir. Sosyal ağlarda yayılan haber makalesinin kaynağı ve durumu analiz edilerek, haberin gerçekliği çıkarılabilir. Literatürde haberlerin bağlam özellikleri dikkate alınarak önerilen çalışmalar kullandıkları yöntemlere göre makine öğrenmesi ve derin öğrenme olmak üzere iki kategoriye ayrılarak incelenmiştir.

• *Makine Öğrenmesi*

Tacchini vd. [72] sahte haberleri tespit etmek için Facebook'ta bunları beğenen kullanıcıları temel olarak bir yöntem önermişlerdir. İletilerin ve hem komplo teorilerinin hem de bilimsel sayfaların kullanıcılarından bir küme oluştururlar ve her bir özellik vektörünün bir sayfayı beğenen kullanıcıları gösteren bir veri kümesi oluşturmuşlardır. Logistic Regression algoritması ile harmonik algoritmayı karşılaştırarak, eğitim verilerinin küçük bir bölümü ile yüksek doğruluk elde ettiklerini göstermişlerdir.

Yang vd. [73] sahte haberleri denetimsiz öğrenme yöntemleri ile çözmek için araştırma yapmışlardır. Haberin gerçekliğini ve kullanıcıların güvenilirliğini gizli rastgele değişkenler olarak ele almış ve kullanıcıların haberin gerçekliğine yönelik görüşlerini tanımlamak için kullanıcıların sosyal medyadaki ilişkilerinden yararlanmışlardır. Haberin gerçekliği, kullanıcıların görüşleri ve kullanıcıların güvenilirliği arasındaki koşullu bağımlılıkları yakalamak için bir Bayesian ağ modelini kurmuşlardır. Haberlerin gerçekliğini ve kullanıcıların güvenilirliğini etiketli veriler olmadan ortaya çıkarmak için etkin bir çökmüş Gibbs örnekleme yaklaşımı önermişlerdir. Önerdikleri yöntemi iki veri kümesinde üzerinde test ederek farklı denetimsiz yöntemlerle karşılaştırmışlardır.

• *Derin Öğrenme*

Volkova vd. [74] sahte haberleri tespit etmek için sahte haberleri hiciv, aldatmaca, tıklamayemi ve propaganda adı altında dört ayrı alt tür ile sorunu ele almaktadır. Bunun için öncelikle Twitter'daki güvenilir ve sahte haber hesaplarının el ile oluşturulmuş bir listesini kullanmayla başlanmıştır. Bu bakımdan 2016'daki Brüksel'deki bombalama olayına ait bir dizi tweet toplamışlardır. Bu tweet metinleri önyargı, öznellik ve ahlaki temeller gibi çeşitli dil ipuçlarını içermektedir. Bunlar kullanıcı etkileşimleri kaynaşmış bir sinir ağı modelinde eğitilerek karşılaştırılmıştır. Aynı özellikler ile geliştirilen modelin performansları gözlemlenerek, farklı haber kategorilerinin özelliklerini niteliksel olarak analiz edilmektedir.

Wang vd. [75] sahte haber unsurlarını tespit etmek için Twitter ve Weibo görüşmelerinden elde edilmiş hem metin hem de görsel özellikleri ayıran çok durumlu bir sinir ağı modeli önermektedirler. Olumsuz ayarlardan esinlenerek, olaya özgü özellikleri kaldırabildiklerini ve olay sayısının bir parametre olarak belirtildiği görünmeyen senaryolara genelleştirilebildiklerini iddia ettikleri bir olay ayırıcı ile birleştirmektedirler. Modeli sahte haber tespiti için iki özel veri kümesi üzerinde değerlendirmektedirler.

Yu vd. [76] sahte haber öğelerinin, çok değişkenli zaman serilerini kullanarak yayılma yollarını modellemek için evrişimli ve Geçitli Tekrarlayan Üniteler (GRU)'in bir kombinasyonunu kullanan, retweetlenen haberin her noktasının kullanıcıların özelliklerine tekabül ettiği ve sahte haberin erken tespitini gerçekleştirdiği bir yöntem önermektedirler. Bu yöntem, orijinal olarak

söylenti çözünürlüğü için tasarlanmış olan diğer son teknolojilerden daha iyi performans gösteren kaskadları paylaşan iki gerçek dünya veri kümesinde değerlendirilmiştir.



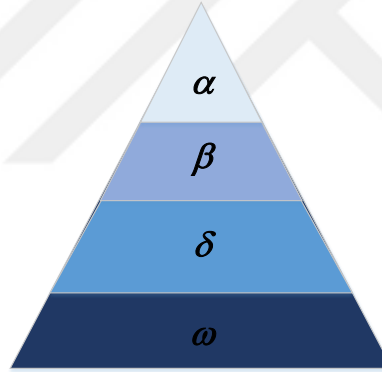
5. UYGULAMADA KULLANILAN ALGORİTMALAR, YÖNTEM VE VERİ KÜMELERİ

Bu bölümde uygulama sırasında kullanılan optimizasyon algoritmaları, önerilen adaptif optimizasyon algoritmaları ve gerçek dünya sahte haber veri kümeleri tanıtılmaktadır.

5.1. Gri Kurt Optimizasyon Algoritması

5.1.1. Algoritmanın İlham Kaynağı

Gri kurtlar (*Canis lupus*), köpekgiller familyasına üyedir. Gri kurtlar süper yırtıcı olarak kabul edilirler, bu da onların besin zincirinin en üstünde oldukları anlamına gelir. Gri kurtlar genellikle sürü halinde yaşamayı tercih eder. Sürü büyüklüğü ortalama 5-12'dir. En ilgi çekici özellikleri, Şekil 5.1'de gösterildiği gibi çok katı sosyal baskın bir hiyerarşiye sahip olmalarıdır [77].



Şekil 5.1. Gri kurt hiyerarşisi (Baskınlık yukarıdan aşağıya doğru azalır)

Liderler, alfa adı verilen bir erkek ve bir dişidir. Alfa, çoğunlukla, avlanma, uyuma alanı, uyanma zamanı vb. hakkında karar vermekten sorumludur. Alfa'nın kararları gruba kabul ettirilir. Ayrıca, bir alfanın grubundaki diğer kurtları takip ettiği bir tür demokratik davranış da gözlemlenmiştir. Toplantılarda, tüm sürü kuyruklarını aşağıda tutarak alfanın kararlarını kabul ettiklerini gösterirler. Tüm sürü alfa kurdun emirlerine uymak zorunda olduğu için, alfa kurda aynı zamanda baskın kurt da denir [78]. Alfa kurtlarının yalnızca sürüden biri ile çiftleşmesine izin verilir. Alfa, sürünün en güçlü üyesi olmak zorunda değildir, grubu yönetme bakımından en iyisidir. Bu durum, bir sürünün organizasyonu ve disiplinin gücünden daha önemli olduğunu göstermektedir [77].

Gri kurt hiyerarşisinde ikinci seviye betadır. Betalar, karar verme sürecinde veya diğer sürü faaliyetlerinde alfaya yardımcı olan alt kurtlardır. Beta kurt erkek veya dişi olabilir ve alfa

kurtlardan birinin ölmesi ya da çok yaşlanması durumunda alfa olmaya en güçlü adaydır. Beta kurt alfaya saygı duymalıdır, fakat diğer alt seviye kurtlara da emretmelidir. Beta kurt alfaya yardımcı rolünü oynar ve sürünün disiplininden sorumludur. Sürü içinde alfanın komutlarını destekler ve alfaya geri bildirimde bulunur [77].

En düşük seviyeli gri kurt omegadır. Omegaya günah keçisi rolü düşer. Omegalar her zaman baskın olan diğer kurtlara boyun eğmek zorundadırlar. Yemek yemelerine en son izin verilen kurtlar omegalardır. Omega, sürüde önemli bir üye değilmiş gibi görünse de, sürüde omega kaybı olması durumunda sürünün iç savaşa maruz kaldığı ve sorunlar olduğu görülmüştür. Bunun nedeni, sürüdeki tüm kurtların şiddet ve geriliminin omegalar tarafından yatıştırılmasıdır. Bu durum, tüm sürüyü tatmin etmeyi ve sürünün baskınlık yapısının korunmasını sağlar. Bazı durumlarda, omega aynı zamanda sürüdeki bebek bakıcısı rolünü de yerine getirir [77].

Alfa, beta veya omega olmayan kurtlara, ast (veya bazı kaynaklarda delta) adı verilir. Delta kurtlar, alfa ve betalara boyun eğmek zorundadırlar, fakat omegalar üzerinde hâkimiyet kurarlar. İzçiler, nöbetçiler, yaşlılar, avcılar ve bekçiler bu kategoriye girer. İzçiler, bölge sınırlarını izlemek ve herhangi bir tehlike durumunda sürüyü uyarmaktan sorumludur. Nöbetçiler sürüyü korur ve güvenliğini sağlar. Yaşlılar, eskiden alfa veya beta olan deneyimli kurtlardır. Avcılar avdayken ve sürü için yiyecek sağlarken alfa ve betalara yardımcı olurlar. Bakıcılar ise, sürüdeki zayıf, hasta ve yaralı kurtların korunmasından sorumludur [77].

Kurtların sosyal hiyerarşisine ek olarak, grup avı da gri kurtların ilginç bir sosyal davranıştır. Gri kurt avının ana evreleri aşağıdaki gibidir [79]:

- Avı takip etmek, kovalamak ve yaklaşmak.
- Av hareket etmeyi bırakana kadar avı takip etmek, etrafını kuşatmak ve taciz etmek.
- Ava saldırmak.

Bu adımlar Şekil 5.2’de gösterilmiştir.

5.1.2. Matematiksel Model ve Algoritma

Bu alt bölümde sosyal hiyerarşi, takip etme, çevreleme ve ava saldırmanın matematiksel modelleri verilmiştir. Ardından ana hatlarıyla GKO algoritması anlatılmıştır [77].

Sosyal Hiyerarşi

GKO’yu tasarlarırken kurtların sosyal hiyerarşisini matematiksel olarak modellemek için en uygun çözüm alfa (α) olarak isimlendirilir. İkinci ve üçüncü en iyi çözümler ise sırasıyla beta (β) ve delta (δ) olarak adlandırılır. Aday çözümlerin geri kalanının ise omega (ω) olduğu varsayılmaktadır. GKO algoritmasında, avlanma α , β ve δ tarafından yönlendirilir. ω kurtlar ise bu üç kurdu takip eder [77].



Şekil 5.2. Gri kurtların avlanma davranışı: (A) Avı takip etme, kovalama ve yaklaşma. (B-D) Avın peşinde olma, etrafını kuşatma ve taciz etme. (E) Sabit konum alma ve ava saldırı [79]

Avın Çevresini Kuşatma

Yukarıda belirtildiği gibi, gri kurtlar avlanma sırasında avlarının çevresini sarmalar. Çevresini sarma davranışını modellemek için aşağıdaki denklemler kullanılır:

$$\vec{D} = |\vec{C} \cdot \vec{X}_p(t) - \vec{X}(t)| \quad (5.1)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (5.2)$$

Denklemlerde t mevcut yinelemeyi gösterir. \vec{A} ve \vec{C} katsayı vektörlerdir. \vec{X}_p , avın konum vektörünü ve \vec{X} ise bir gri kurdun konum vektörünü gösterir.

\vec{A} ve \vec{C} vektörleri Denklem 5.3 ve 5.4 ile hesaplanır.

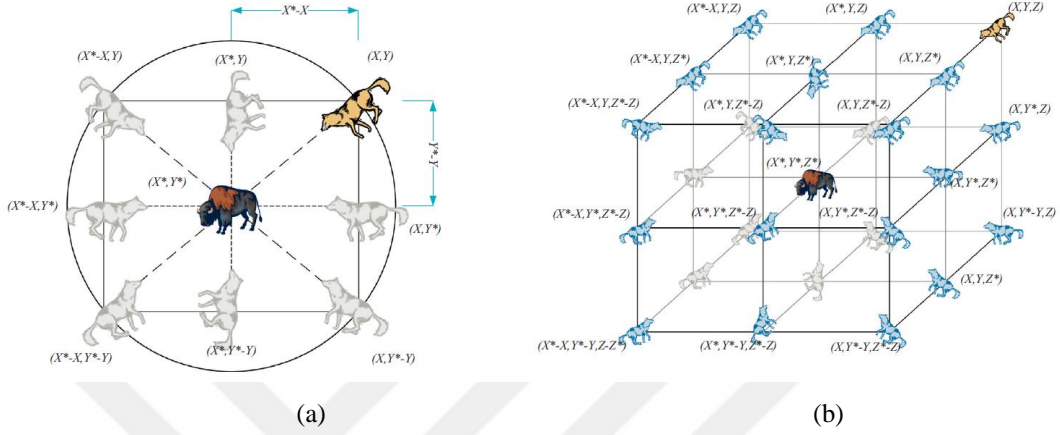
$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \quad (5.3)$$

$$\vec{C} = 2 \cdot \vec{r}_2 \quad (5.4)$$

Burada \vec{a} 'nın bileşenleri, yinelemeler boyunca doğrusal olarak 2'den 0'a düşürülmüştür. \vec{r}_1 ve \vec{r}_2 ise $[0,1]$ aralığında rastgele vektörlerdir.

Denklem 5.1 ve 5.2'nin etkilerini görmek için, iki-boyutlu (2B) bir konum vektörü ve olası komşuların bazıları Şekil 3(a)'da gösterilmiştir. Şekilden de görülebileceği gibi, (X, Y) konumundaki bir gri kurt, (X^*, Y^*) konumundaki avına göre kendi konumunu güncelleyebilir. \vec{A} ve \vec{C} vektörlerinin değerleri ayarlanarak, mevcut pozisyona göre en iyi ajanın etrafındaki farklı yerlere

ulaşılabilir. Örneğin; (X^*-X, Y^*) konumuna, $\vec{A} = (1,0)$ ve $\vec{C} = (1,1)$ ayarlarıyla ulaşılabilir. Gri bir kurdun üç-boyutlu (3B) uzayda muhtemel güncellenmiş konumları, Şekil 5.3(b)'de gösterilmektedir. \vec{r}_1 ve \vec{r}_2 rastgele vektörleri kurtların Şekil 5.3'te gösterilen noktalar arasında herhangi bir konuma erişmesine izin verir [77].



Şekil 5.3. (a) 2B ve 3B konum vektörleri (b) Bir sonraki olası konumları

Böylece, gri kurtların herhangi bir rastgele bir konumda, avının etrafındaki alanda konumu Denklem 5.1 ve 5.2 eşitlikleri kullanılarak güncellenebilir.

Aynı kavram n boyutlu bir arama alanına ile genişletilebilir ve gri kurtlar bugüne kadar elde edilen en iyi çözüm etrafında hiper-küpler (veya hiper-küreler) içinde hareket edecektir.

Avlanma

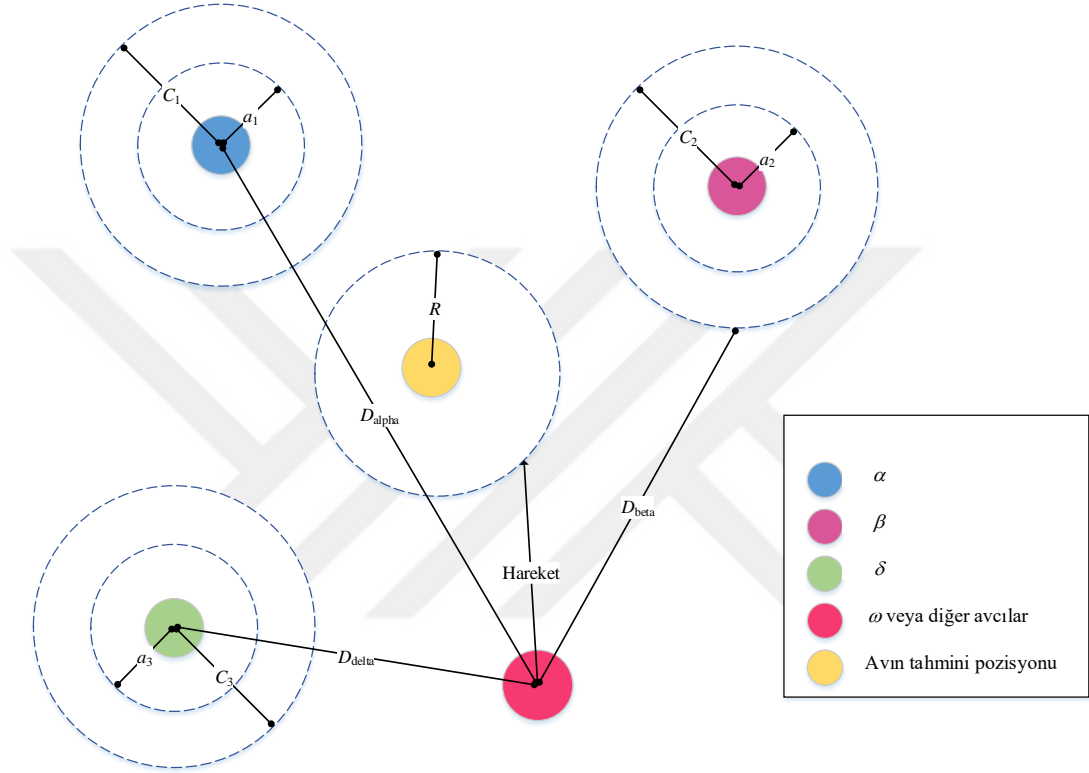
Gri kurtlar, avın konumu tanıma ve çevresini sarma yeteneğine sahiptir. Avlanma genellikle alfa tarafından yönlendirilir. Ara sıra, beta ve delta da avlanmaya katılabilir. Ancak, soyut bir arama alanında, en iyi avın konumu hakkında hiçbir fikrimiz yoktur. Gri kurtların avlanma davranışını benzetim yapabilmek için, alfa (en iyi aday çözüm), beta ve deltanın avın potansiyel konumu hakkında daha iyi bilgiye sahip olduğu varsayılır. Bu nedenle elde edilen en iyi ilk üç çözüm kaydedilerek ve diğer (omegalar dâhil) arama ajanları, konumlarını en iyi arama ajanlarının konumuna göre güncellemeye zorlanmıştır. Bu amaçla aşağıdaki formüller önerilmiştir [77]:

$$\vec{D}_a = |\vec{C}_1 \cdot \vec{X}_a - \vec{X}|, \vec{D}_\beta = |\vec{C}_2 \cdot \vec{X}_\beta - \vec{X}|, \vec{D}_\delta = |\vec{C}_3 \cdot \vec{X}_\delta - \vec{X}| \quad (5.5)$$

$$\vec{X}_1 = \vec{X}_a - \vec{A}_1 \cdot (\vec{D}_a), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta), \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta), \quad (5.6)$$

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (5.7)$$

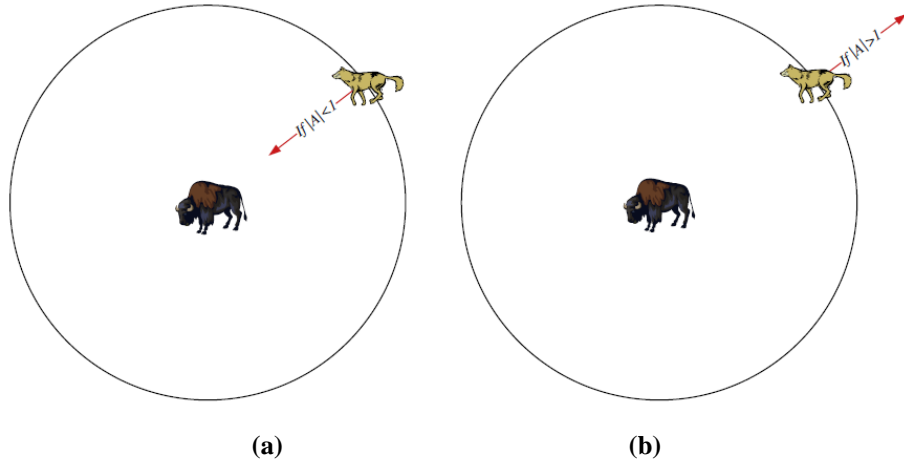
Şekil 5.4'te, bir arama ajanının konumunu 2B arama alanındaki alfa, beta ve deltaya göre nasıl güncellediği gösterilmektedir. Son pozisyonun, arama alanındaki alfa, beta ve delta pozisyonları tarafından tanımlanan bir daire içinde rastgele bir yerde olacağı gözlemlenebilir. Başka bir ifade ile alfa, beta ve delta, avın konumunu tahmin eder ve diğer kurtlar da konumlarını avın çevresinde rastgele olarak günceller [77].



Şekil 5.4. GKO'da konum güncelleme

Ava Saldırma (Sömürü)

Yukarıda da belirtildiği av durduğunda gri kurtlar avına saldırarak avlanmayı bitirir. Ava yaklaşmayı matematiksel olarak modellemek için \vec{a} 'nın değeri azaltılmıştır. \vec{A} değişim aralığında \vec{a} ile birlikte azalır. Diğer bir ifade ile \vec{A} , yineleme süresince 2'den 0'a düşen $[-2a, 2a]$ aralığında rastgele bir sayıdır. \vec{A} 'nın rastgele değerleri $[-1,1]$ olduğunda, bir arama ajanının bir sonraki konumu, mevcut konumu ile avın konumu arasında herhangi bir noktada olabilir. Şekil 5.5(a)'da, $|A| < 1$ durumunun kurtları ava saldırmaya zorladığı gösterilmektedir [77].



Şekil 5.5. (a) Ava saldırı (b) Av arayışı [77]

Şimdiye kadar önerilen operatörlerle, GKO algoritması, arama ajanlarının alfa, beta ve deltanın konumuna göre konumlarını güncellemelerine ve ava saldırmalarına olanak tanır. Ancak, GKO algoritması bu operatörlerle yerel çözümlerde hareketsizliğe eğilimlidir. Önerilen çevreleme mekanizmasının belirli ölçüde keşif özelliği gösterdiği doğrudur, ancak GKO'nun keşfi vurgulamak için daha fazla operatöre ihtiyacı vardır.

Av Arama (Keşif)

Gri kurtlar çoğunlukla alfa, beta ve deltanın konumuna göre arama yaparlar. Av aramak için birbirlerinden uzaklaşır, avlarına saldırmak için bir araya gelirler. Birbirinden uzaklaşmayı matematiksel olarak modellemek için, arama ajanını avdan uzaklaştırmaya zorlamak amacıyla \vec{A} ile birlikte 1'den büyük veya -1'den küçük rastgele değerler kullanılmıştır. Bu durum, keşfi ön plana çıkarmayı ve GKO algoritmasının global olarak avlanma yapmasına imkan sağlar. $|A| > 1$ durumunun, gri kurtları olası başka daha iyi avlar bulmak için avdan ayrılmaya zorladığı da Şekil 5.5(b)'de gösterilmektedir [77].

GKO'nun keşfi destekleyen diğer bir bileşeni de \vec{C} 'dir. Denklem 4.4'ten de görülebileceği gibi, \vec{C} vektörü [0,2]'den rastgele değerler içerir. Bu bileşen, rassal biçimde avın Denklem 5.1'deki mesafeyi belirleme üzerindeki etkisini öne çıkarmak ($C > 1$) veya bunun önemini azaltmak ($C < 1$) amacıyla av için rastgele değerler sağlar. Bu durum, optimizasyon boyunca GKO'nun, keşif ve yerel optimadan kaçınma lehinde daha rastgele bir davranış göstermesine yardımcı olur. Burada, C 'nin A 'dan farklı olarak, doğrusal olarak azaltılmadığını belirtmek gerekir. Keşfi vurgulamak amacıyla, kasıtlı olarak C 'nin sadece ilk tekrarlamalar sırasında değil, son tekrarlamalar sırasında da daima rastgele değerler vermesi istenmiştir. Bu bileşen, özellikle son yinelemelerde, yerel optima hareketsizlik durumunda çok faydalıdır [77].

C vektörü ayrıca doğadaki avlara yaklaşmanın önündeki engellerin etkisi olarak da düşünülebilir. Genel olarak ifade edilirse, doğadaki engeller kurtların avlanma yollarında ortaya çıkar ve gerçekte avlarına hızlı ve kolay şekilde yaklaşmalarını engeller. Tam olarak C vektörünün yaptığı da budur. Bir kurdun konumuna bağlı olarak, av rastgele bir ağırlık verebilir ve ava ulaşmayı kurtlar için zorlaştırır veya avı uzaklaştırır ya da tam tersini yapar [77].

GKO algoritması arama sürecine rastgele gri kurt popülasyonu (aday çözümler) oluşturmakla başlar. Yinelemeler boyunca, alfa, beta ve delta kurtlar avın olası konumunu tahmin eder. Her aday çözüm, av ile mesafesini günceller. Keşif ve sömürüyü vurgulamak için a parametresi 2'den 0'a düşürülür. Aday çözümler $|A| > 1$ olması durumunda avdan uzaklaşma ve $|A| < 1$ olması durumunda ise ava yaklaşma eğilimindedir. Son olarak, GKO bir sonlandırma kriterinin yerine getirilmesi ile sona erer [77].

GKO algoritmasının optimizasyon problemlerini teorik olarak nasıl çözdüğünü görmek için bazı noktalara dikkate alınabilir:

- Önerilen sosyal hiyerarşi, GKO'nun yineleme süreci boyunca şimdiye kadar elde edilen en iyi çözümleri kaydetmesine yardımcı olur.
- Önerilen çevreleme mekanizması, çözümlerin çevresinde hiper-küre olarak daha üst boyutlara genişletilebilecek daire şeklinde bir alanı tanımlar.
- A ve C rastgele parametreleri, farklı rastgele yarıçapları olan hiper-kürelere sahip aday çözümlere yardımcı olur.
- Önerilen avlanma yöntemi, aday çözümlerin avın muhtemel pozisyonunu tespit etmesini sağlar.
- Keşif ve sömürü uyarlanabilen a ve A değerleri tarafından güvence altına alınmıştır.
- a ve A parametrelerinin uyarlanabilir değerleri, GKO'nun keşif ve sömürü arasında sorunsuzca geçiş yapmasını sağlar.
- A 'nın azalmasıyla birlikte, yinelemelerin yarısı keşfe ($|A| \geq 1$) ve diğer yarısı ise sömürmeye ($|A| < 1$) aittir.
- GKO'nun, (a ve C olmak üzere) sadece iki ana parametresi vardır [77].

GKO algoritmasının sözde kodu Şekil 5.6'da verilmiştir.


```

Gri kurtların başlangıç popülasyonunun ayarlanması  $X_i$  ( $i = 1, 2, \dots, n$ )
 $a$ ,  $A$  ve  $C$  parametrelerine ilk değerlerin verilmesi
Her bir arama ajanının uygunluk değerinin hesaplanması
 $X_a =$  en iyi arama ajanı
 $X_\beta =$  en iyi ikinci arama ajanı
 $X_\delta =$  en iyi üçüncü arama ajanı
while ( $t <$  Maksimum yineleme sayısı)
for her bir arama ajanı
Mevcut arama ajanının konumunu (5.7)'ye göre güncelle
end for
 $a$ ,  $A$  ve  $C$  parametrelerini güncelle
Tüm arama ajanlarının uygunluğunu hesapla
 $X_a$ ,  $X_\beta$  ve  $X_\delta$ 'yi güncelle
 $t = t + 1$ 
end while
return  $X_a$ 

```

Şekil 5.6. GKO algoritmasının sözde kodu

5.2. Salp Sürüsü Optimizasyon Algoritması

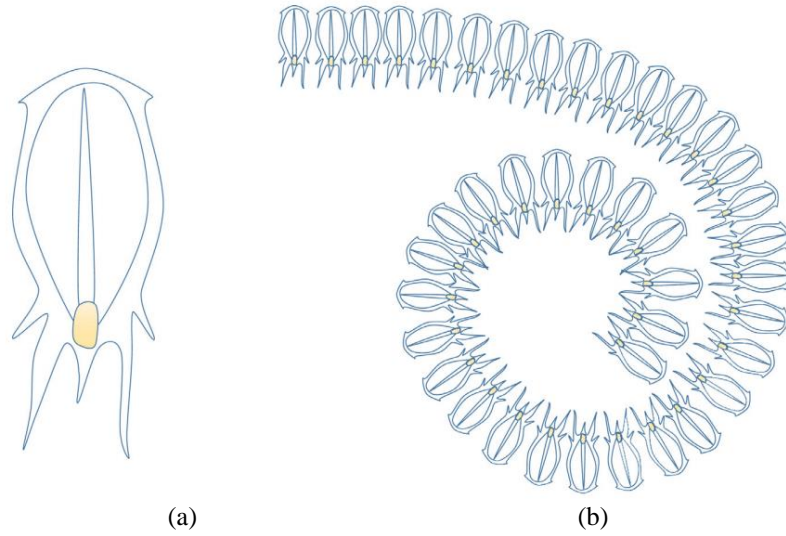
5.2.1. Algoritmanın İlham Kaynağı

Salplar “Salpidae” familyasına aittir ve şeffaf bir fiçı şeklindeki gövdeye sahiptir. Dokuları denizanelerine çok benzemektedir. Ayrıca, denizanelerinde olduğu gibi suyu ileriye doğru hareket ettirmek vücutlarından su pompalayarak itme gücü oluştururlar [80]. Bir salpın şekli, Şekil 5.7(a)'da gösterilmiştir.

Bu yaratıklarla ilgili biyolojik araştırmalar, yaşam alanlarına erişim son derece zor olduğu ve asgari laboratuvar ortamlarını korumak gerçekten zor olduğu için erken dönüm noktalarında. Salpların en ilginç davranışlarından biri, sürü davranışlarıdır. Derin okyanuslarda, salplar sıklıkla salp zinciri olarak adlandırılan bir sürüyü oluşturur. Bu zincir Şekil 5.7(b)'de gösterilmektedir. Bu davranışın temel nedeni henüz tam olarak açıklanmasa da, bazı bilim adamları bunun daha iyi bir hareketlilik elde etmek için yapıldığına inanmaktadır [81].

5.2.2. Salp Zincirlerinin Hareketi için Önerilen Matematiksel Model

Literatürde salpların sürü davranışlarını [82] ve popülasyonunu [83] matematiksel olarak modellemek için çok az bilgi vardır. Salp zincirlerinin matematiksel olarak modellemek için, popülasyon lider ve takipçiler olmak üzere ikiye ayrılır. Zincirin önündeki salp lider, geri kalan salplar ise takipçi olarak kabul edilirler. İsminden de anlaşılacağı gibi lider, sürüye rehberlik eder ve takipçiler birbirlerini takip ederler (dolaylı olarak lideri takip ederler).



Şekil 5.7. (a) Bireysel salp. (b) Salp sürüsü (Salp zinciri) [84]

Diğer sürü tabanlı yöntemlere benzer olarak, SSO algoritmasında da salpların pozisyonları n -boyutlu bir arama uzayında tanımlanır (n : Verilen problemdeki değişkenlerin sayısıdır). Bu nedenle bütün salpların pozisyonları iki-boyutlu x matrisinde saklanır. Ayrıca, arama uzayında sürünün hedefi olan ve F olarak adlandırılan bir yiyecek kaynağı olduğu varsayılır [84].

Liderin konumunu güncellemek için Denklem 5.8 kullanılır.

$$x_j^1 = \begin{cases} F_j + c_1 \left((ub_j - lb_j)c_2 + lb_j \right) & c_3 \geq 0 \\ F_j - c_1 \left((ub_j - lb_j)c_2 + lb_j \right) & c_3 < 0 \end{cases} \quad (5.8)$$

x_j^1 j . boyutta ilk salpın (liderin) konumunu gösterir. F_j , j . boyutta yiyecek kaynağının konumudur. lb_j ve ub_j sırasıyla j . boyutun alt ve üst sınırını gösterir. c_1 , c_2 ve c_3 ise rastgele sayılardır.

Denklemler, liderin konumunu sadece ilgili yiyecek kaynağına göre güncellediğini göstermektedir. c_1 katsayısı, keşif ve sömürüyü dengelediği için SSO algoritması için çok önemli bir parametredir ve Denklem 5.9 ile tanımlanır:

$$c_1 = 2e^{-\left(\frac{I}{L}\right)^2} \quad (5.9)$$

Denklemden I mevcut yineleme sayısını ve L maksimum yineleme sayısını gösterir. c_2 ve c_3 $[0,1]$ aralığında tekdüze oluşturulan rastgele sayılardır [84].

Takipçilerin konumlarını güncellemek için Denklem 5.10'da verildiği gibi Newton'un hareket yasası kullanılır:

$$x_j^i = \frac{1}{2}at^2 + v_0t \quad (5.10)$$

$i \geq 2$ olduğu durumda, x_j^i j. boyuttaki i. takipçi salpın konumunu gösterir. t zamanı, v_0 başlangıç hızı gösterir. $v = \frac{x-x_0}{t}$ olduğu durumda $a = \frac{v^{final}}{v_0}$ olur [84].

Optimizasyondaki süre yineleme olduğu için, yinelemeler arasındaki tutarsızlık 1'e eşittir ve $v_0 = 0$ olduğunda bu Denklem 5.11 gibi ifade edilebilir:

$$x_j^i = \frac{1}{2}(x_j^i + x_j^{i-1}) \quad (5.11)$$

$i \geq 2$ ve x_j^i j. boyuttaki i. takipçi salpın konumunu gösterir.

SSO algoritmasının sözde kodu Şekil 5.8'de verilmiştir.

```

Salpların başlangıç popülasyonunun  $x_i(i = 1, 2, \dots, n)$  ub ve lb'ye göre ayarlanması
while (Sonlandırma şartı sağlanıncaya kadar)
  Her bir arama ajanının (salp) uygunluk değerini hesapla
  F = en iyi arama ajanı
   $c_1$  değişkenini (5.9)'a göre güncelle
  for her bir salp ( $x_i$ )
    if ( $i == 1$ )
      Lider salpın konumunu (5.8)'e göre güncelle
    else
      Lider salpın konumunu (5.11)'e göre güncelle
    end
  end for
  Değişkenlerin alt ve üst sınırlarına göre salpları değiştir
end while
return F

```

Şekil 5.8. SSO algoritmasının sözde kodu

Sözde kod, SSO algoritmasının rastgele pozisyonlara sahip çoklu salpları başlatarak küresel en iyi noktaya yaklaşmaya başladığını göstermektedir. Daha sonra her bir salpın uygunluğunu hesaplanır, en iyi uygunluğa sahip olan salp bulunur ve en iyi salpın konumu, salp zinciri tarafından kovalanacak olan yiyecek kaynağı olarak F değişkenine atanır. Aynı zamanda, c_1 katsayısı Denklem 5.9 ile güncellenir. Her boyut için, lider salpın konumu Denklem 5.8 ile güncellenir ve takipçi salpların konumu Denklem 5.11 kullanılarak güncellenir. Salplardan biri arama uzayı alanının dışına çıkarsa, sınırlara geri getirilecektir. Başlatma süreci dışındaki tüm adımlar, sonlandırma şartı sağlanıncaya kadar tekrarlı bir biçimde çalıştırılır [84].

Optimizasyon süreci boyunca yiyecek kaynağının güncelleneceği unutulmamalıdır, çünkü salpların etrafındaki alanı keşif ve sömürü ile daha iyi bir çözüm bulma olasılığı çok yüksektir [84].

Önerilen salp zinciri modelinin ve SSO algoritmasının optimizasyon problemlerini çözmede nasıl etkili olduğunu görmek için bazı açıklamalar aşağıda listelenmiştir [84]:

- SSO algoritması, şu ana kadar elde edilen en iyi çözümü kaydeder ve bunu gıda kaynağı değişkenine atar, böylece tüm popülasyon bozulursa bile asla kaybolmaz.
- SSO algoritması, şimdiye kadar elde edilen en iyi çözüm olan, yalnızca besin kaynağına göre lider salpın konumunu güncellemektedir, bu yüzden lider daima etrafındaki alanı araştırır ve kullanır.
- SSO algoritması, takipçi salpların birbirlerine göre konumlarını günceller, böylece kademeli olarak lider salplara doğru hareket ederler.
- Takipçi salpların kademeli hareketleri, SSO algoritmasının yerel en iyi noktada kolayca durmaktan korur.
- c_1 parametresi yinelemeler boyunca uyarlamalı olarak düşürülür, bu nedenle SSO algoritması önce arama alanını keşfeder ve sonra sömürür.
- SSO algoritması sadece bir ana kontrol parametresine (c_1) sahiptir.
- SSO algoritması basit ve uygulaması kolaydır [84].

5.3. Adaptif Salp Sürüsü Optimizasyon Algoritmaları

Bu bölümde SSO algoritmasının çözüm doğruluğunu, güvenilirliğini ve yakınsama hızını artırmak için SSO yapısı iyileştirilmeye çalışılmıştır. Mevcut en iyi çözümü elde etmek için atalet ağırlığı adında yeni bir kontrol parametresi eklenmiştir. Bu çalışmada önerilen SSO algoritmasının yeni versiyonu Adaptif SSO (ASSO) algoritması olarak adlandırılmıştır. Atalet ağırlığı parametresi, en uygun çözümün kesin olarak anlaşılmasını sağlamak için keşif ve sömürü arasındaki denge kurmak için kullanılmıştır. Önerilen ASSO için, SSO algoritmasındaki Denklem 5.8, Denklem 5.12'deki gibi güncellenmiştir.

$$x_j^1 = \begin{cases} w * F_j + c_1 \left((ub_j - lb_j)c_2 + lb_j \right) & c_3 \geq 0 \\ w * F_j - c_1 \left((ub_j - lb_j)c_2 + lb_j \right) & c_3 < 0 \end{cases} \quad (5.12)$$

Denklemlerde verilen semboller ve anlamları Tablo 5.1'de verilmiştir:

Önerilecek ASSO algoritması için atalet ağırlığı dikkate alınarak iki farklı yaklaşım kullanılmıştır. Bunlardan ilki doğrusal olmayan atalet ağırlığı diğeri ise osilasyon (salınım) atalet ağırlığı yaklaşımıdır.

Tablo 5.1. ASSO algoritmasının sembol ve anlamları

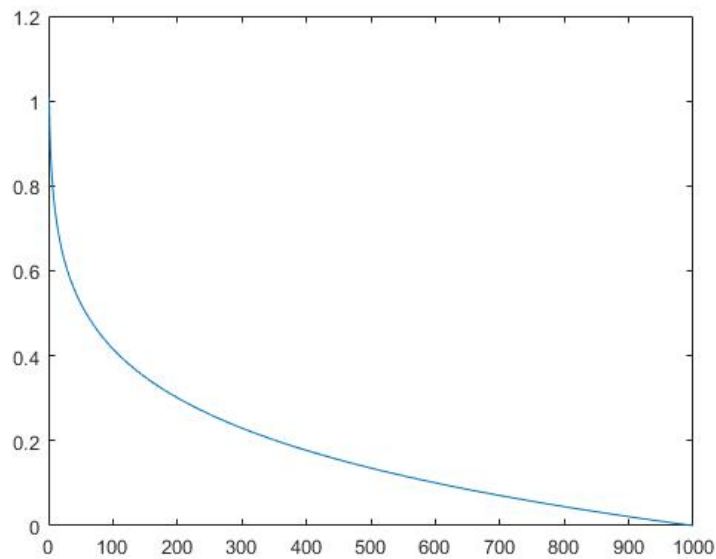
Sembol	Anlamı
w	Atalet ağırlığı
x_j^1	j . boyutta ilk salpın (liderin) konumu
F_j	j . boyutta yiyecek kaynağının konumu
lb_j	j . boyutun alt sınırı
ub_j	j . boyutun üst sınırı

5.3.1. Doğrusal Olmayan Şekilde Azalan Atalet Ağırlıklı Salp Sürüsü Optimizasyon Algoritması

Doğrusal olmayan şekilde azalan atalet stratejisi Yang vd. [85] tarafından PSO algoritmasının performansını artırmak için önerilmiştir. Bu yöntem doğrusal güncelleme stratejisini kullanır. Her bir yinelemede w değeri Denklem 5.13 kullanılarak hesaplanır:

$$w(t) = w_{max} - (w_{max} - w_{min})\left(\frac{t}{T}\right)^a \quad (5.13)$$

Denklemdaki a parametresi kullanıcı tarafından belirlenen bir değerdir. Yang vd. deneysel olarak a parametresinin değerini $a = \frac{1}{\pi^2}$ ile hesaplamışlardır. $w_{min} = 0$ ve $w_{max} = 2$ olarak belirlenmiştir. $T = 1000$ olarak verildiğinde atalet ağırlığı değerlerinin değişimi Şekil 5.9'da verilmiştir. Bu yöntemin ismi tez çalışmaları boyunca Adaptif SSO-1 (ASSO1) olarak kullanılmıştır.



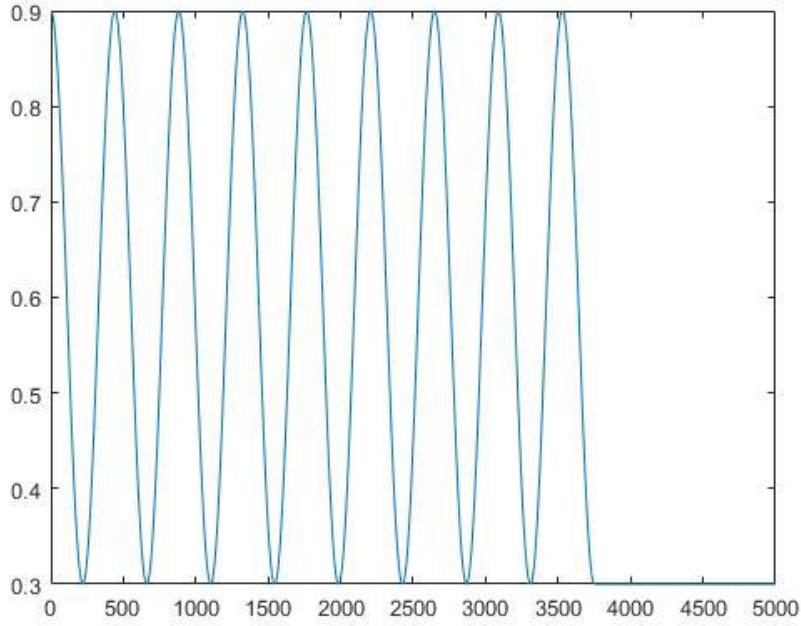
Şekil 5.9. Doğrusal olmayan atalet ağırlığı değerleri

5.3.2. Osilasyon (Salınım) Atalet Ağırlıklı Salp Sürüsü Optimizasyon Algoritması

Osilasyon (Salınım) atalet ağırlığı kavramı, bir keşif dalgası ardından bir sömürü dalgasının optimizasyon süreci boyunca tekrarlanan bir döngü oluşturduğu bir strateji uygular [86]. Salımlı azaltma stratejisi, arama işleme sırasında tekdüze bir azaltma yerine w parametresinin değerini güncellemek için sinüzoidal bir fonksiyon kullanır:

$$f(x) = \begin{cases} \frac{w_{min} + w_{max}}{2} + \frac{w_{max} - w_{min}}{2} \left(\cos \frac{2\pi t(4k+6)}{3T} \right), & \text{Eğer } t < \frac{3T}{4} \\ w_{min} & , \text{diğer durumda} \end{cases} \quad (5.14)$$

$w_{min} = 0.3$, $w_{max} = 0.9$ ve $k = 7$ olarak ayarlanmıştır. $T = 5000$ olarak verildiğinde atalet ağırlığı değerlerinin değişimi Şekil 5.10'da verilmiştir. Bu yöntemin ismi tez çalışmaları sırasında Adaptif SSO-2 (ASSO2) olarak kullanılmıştır.



Şekil 5.10. Osilasyon atalet ağırlığı değerleri

5.4. Denetimli Yapay Zekâ Algoritmaları

Denetimli yapay zekâ algoritmaları, eğitim setine dayalı bir öğrenme şeklidir [87]. Veri kümelerindeki örnek etiketlerinin zaten bilindiğini varsayarlar. Etiketli bir eğitim veri kümesine ihtiyaç duyarlar ve denetlenecek test verilerine örnek olan bir işlev döndürürler. İşlevin görevi girdiden minimum hata oranına sahip çıktıyı tahmin etmektir. Aşağıdaki bölümde, tez çalışmada kullanılan denetimli yapay zekâ algoritmaları hakkında kısa bilgiler verilmiştir.

• **Naive Bayes:** Naive Bayes belirli bir veri kümesindeki değerlerin sıklığını ve kombinasyonunu sayarak olasılık kümesini hesaplayan seçici bir sınıflandırıcıdır. Naive Bayes sınıflandırıcısı bayes teoremi ve toplam olasılık teoremine dayanmaktadır. Sınıfların etiketleri, belirli bir örneğin olasılığı ile tahmin edilir. Gerekli olan sınıf etiketini tahmin etmek için sadece eğitim verisine ihtiyaç vardır [88].

• **Sequential Minimal Optimization (SMO):** Bu algoritma ilk olarak 1998’de Platt tarafından tanıtılmıştır. SMO algoritması temel olarak SVM eğitimi güçlendirmek için kullanılır. SMO’nun çalışma mantığı, bir problemi daha sonra analitik olarak çözülecek en küçük olası alt problemler grubuna böler. Daha sonra, problemin en küçük ve en büyük çözüm noktalarında bulunan doğrusal eşitlik kısıtları nedeniyle, olası en küçük problem iki faktöre ayrılır. Daha sonra, iki faktöre uygulanan kısıtlamalar azaltılır. Bu indirgeme problemi, minimum bir boyutlu ikinci dereceden bir fonksiyon bulmak için analitik olarak çözülür. Her yinelemede, eşitliğin sürekli kısıtındaki terimler, geri kalan kısmındaki toplamın negatifi olarak hesaplanır [89].

• **JRip:** JRip (RIPPER), temel ve en popüler sınıflandırma algoritmalarından biridir. Sınıflar artan boyutta incelenir ve artan azaltılmış hata (RIPPER) kullanılarak sınıf için başlangıç kuralları kümesi oluşturulur. RIPPER, bir sınıfın eğitim verilerindeki belirli yargıların tüm örnekleri göz önüne alınarak ve o sınıfın tüm üyelerini kapsayan bir kural kümesi bulunarak elde edilir [90].

• **OneR:** “Bir Kural” kelimelerinin kısaltması olan OneR, tek seviyeli bir karar ağacı oluşturan basit bir sınıflandırma algoritmasıdır. OneR, bir takım örneklerden tipik olarak basit ama kesin sınıflandırma kurallarını çıkarabilir. OneR basit olmasına rağmen esneklik gösteren eksik değerleri ve sayısal özellikleri işleyebilir. OneR algoritması, eğitim verilerindeki her bir özellik için bir kural oluşturur, ardından kuralı ‘Bir Kural’ olarak minimum hata oranıyla seçer [91].

• **Stochastic Gradient Descent (SGD):** SGD, bir amaç fonksiyonunu optimize etmek için yinelemeli bir yöntem kullanan modern bir sınıflandırıcıdır. Bu sınıflandırıcı, gradyanları değerlendirmek için rastgele seçilen örnekleri kullanır, bu nedenle, stokastik olarak adlandırılır [92].

• **Bagging:** Mevcut bir eğitim kümesinden yeni bir eğitim seti çıkararak temel öğrenmeyi yeniden eğitmenin bir yöntemidir. Eğitim kümesi, n numuneden oluşan eğitim seti ile bir örnek kümesi değiştirilerek rastgele üretilir. Seçilen her bir örnek eğitim kümesine iade edilir. Bazı örnekler yeni eğitim kümesine dâhil edilmezken, diğerleri birden fazla kez yer alabilir. [93].

• **Decision Stump:** Karar ağacının yalnızca bir seviyesine sahip bir makine öğrenmesi modelidir. Yapraklarına bağlı bir kökü vardır. Adaboost gibi topluluk algoritmalarına uygulanır. Algoritma, tahmin için tek bir giriş özelliğinin değerlerini kullanır [94].

• **ZeroR:** Hedefe bağlı olan ve özelliklere en az önem veren en basit sınıflandırma yöntemidir. ZeroR sınıflandırıcısı, çoğunluk kategorisini tahmin eder. Algoritma, hedef için bir

frekans tablosu oluşturur ve en sık değeri seçer. ZeroR, diğer sınıflandırma metotları için bir temel performansın belirlenmesi için kullanışlıdır [95].

- **Cross Validation Parameter Selection (CVPS):** Bu sınıflandırıcı, herhangi bir sınıflandırıcı için çapraz doğrulama parametreleri seçmek için kullanılan bir yöntemdir [96].

- **Randomizable Filtered Classifier (RFC):** Yalnızca eğitim verilerine göre oluşturulmuş verileri isteğe bağlı bir filtreden geçirdikten sonra veriler üzerinde rasgele bir sınıflandırıcı kullanır. Test örnekleri, yapıları değiştirilmeden filtre tarafından ele alınır [97].

- **Logistic Model Tree (LMT):** Sınıflandırıcı, tek bir ağaç elde etmek için karar ağacı öğrenmesini ve lojistik regresyonu birleştirir. LogitBoost algoritması, ağaçtaki her düğümde bir Lojistik Regresyon modeli elde etmek için kullanılır. LogitBoost, tüm veriler kullanılarak elde edilen bir lojistik model oluşturmak için kullanılır. Verileri kökte bölmek için bir eşik değeri kullanılır. Bölünme işlemi durma koşulu sağlanana kadar devam eder [98].

- **Locally Weighted Learning (LWL):** Bu yöntemde, örneklerin ağırlığını atamak için örnek tabanlı bir algoritma kullanılır. Daha sonra bu algoritma sınıflandırma veya regresyon için kullanılır. Algoritma, tüm fonksiyon alanı için komşu verilere dayanan yerel bir model oluşturur ve test verilerini hafızada saklar [99].

- **Classification via Clustering (CvC):** Bu yöntem sınıflandırma için bir küme kullanır. Kümeleme benzer örnekleri gruplama işlemidir. Kümeleme, verileri sınıflandırmadan önce ön işleme adımı olarak kabul edilebilir. Sınıflandırma kümeleme işlemine dayanmaktadır. Eğitim ve test verileri birlikte kümelenir ve test verilerinin etiketini bulmak için sınıf küme dağılımları kullanılır [100].

- **Weighted Instances Handler Wrapper (WIHW):** Bu yöntemde, eğitim ağırlıkları örnek ağırlıklarıyla başa çıkabiliyorsa temel sınıflandırıcıya iletilir. Bununla birlikte, yeniden örnekleme kullanımını ağırlıklarla da zorlamak mümkündür [101].

- **Ridor:** Ridor, Ripple-Down Rule [102] sınıflandırma algoritmasının bir uygulamasıdır. Ripple-Down Rule doğrudan bir sınıflandırma algoritmasıdır. Sıradan ağaç sınıflandırıcılarından farklı bir ikili karar ağacı oluşturur, standart ağaçların aksine bir iç düğümde bir karara varılabilir. Tüm kural düğümleri iki yönlü bir ilişki ile bağlantılıdır. Bir düğüm için öncüllerin tümü doğruysa, o zaman test konusu bu düğüm tarafından iddia edilecektir [103].

- **Multi-Layer Perceptron (MLP):** MLP algoritması, 1950'de Rosenblatt tarafından önerilmiştir. Bir MLP, giriş katmanı, gizli katman ve çıkış katmanı olmak üzere en az üç düğüm katmanına sahiptir. Bu algoritma, eğitim için geri yayılım adı verilen denetimli bir öğrenme tekniği kullanır. MLP, lineer bir algılayıcıdan çoklu tabakaları ile ayrılır. MLP'de öğrenme, çıktıdaki hata miktarına bağlıdır. Her veri parçası işlendikten sonra bağlantıların ağırlıkları değişir [104].

- **Ordinal Learning Model (OLM):** OLM, monotonluk sınırlamaları ile sıralı sınıflandırma

amacı ile önerilmiştir. Öğrenme aşamasında, her örnek bir kural tabanındaki her bir kurala göre kontrol edilir. Eğer bir örnek kural tabanındaki bir kural ile tutarlı değilse, elenirken bunlardan biri rastgele seçilir. Ancak örnek seçilirse, diğer tüm monotonluk kurallarına uygunluğu kontrol edilir [105].

- **Simple Cart:** Metot Leo Breiman [106] tarafından veri keşfi ve tahmininde kullanılan geleneksel metotlara alternatif olarak önerilmiştir. Simple Cart, sınıflandırma için ikili karar ağacını yaratır. Budama aşamasında, en iyi ağacı seçmek için çapraz doğrulama veya en büyük veri kullanılır. Bu yöntemde, veriler farklılıkları olan iki alt gruba ayrılır. Bu işlem, alt grup büyüklüğü minimumda olduğunda sona erer [107].

- **Attribute Selected Classifier (ASC):** Sınıflandırıcıya geçilmeden önce eğitim ve test verilerinin boyutu özellik seçimi ile azaltılır. Bu yöntem, sınıflandırıcının yanında bir arama algoritması ve değerlendirici olarak bulunur. Temel sınıflandırıcı C4.5 algoritması kullanılarak oluşturulan karar ağacıdır [108].

- **J48:** J48 algoritması genellikle sınıflandırma uygulamaları için tercih edilen algoritmadır. Bu bağlamda, J48, ID3 ve C4.5 algoritmalarına dayanan istatistiksel bir karar ağacı algoritmasıdır. J48, her bir ağaçtan bir düğüm kullanma mantığı ile düğümler ve alt sıralardan çocuk satırlar üzerinde çalışır. Bu bağlamda, sınıflandırma algoritmaları arasında genellikle en hızlı ve en yüksek hassasiyetli algoritmalarından biridir [109].

- **REPTree:** REPTree ilk kez Quinlan tarafından önerilmiştir [110]. REPTree, regresyon ağacı mantığını kullanır ve farklı yinelemelerde birden fazla ağaç oluşturur. Daha sonra, üretilen tüm ağaçlardan en iyisini seçer ve temsilci olarak kabul eder. Ağacın budamasında kullanılan ölçü, ağaç tarafından yapılan tahminlerde ortalama karesel hatadır. REPTree algoritması, entropi ile bilgi kazancının hesaplanması ve varyanstan kaynaklanan hatayı azaltmak ilkesine dayanır [108].

- **IBk:** Instance-based k -en yakın komşuluk algoritması (IBk), k -NN algoritmasının bir alternatifidir. Belirli örnekler kullanarak sınıflandırma tahimleri yapan bir çerçeve oluşturur. IBk, örneklerden türetilmiş bir soyutlama kümesi tutmaz. Bu yöntem en yakın komşu algoritmasını kullanır ve genişletir. Bununla birlikte, bu en yakın komşunun daha büyük depolama alanı gerektirdiği düşünülmektedir. IBk kullanarak, bu depolama ihtiyacı önemli ölçüde azaltılabilir ve öğrenme veya doğruluk oranları daha fazla azalmaz [111].

- **Kernel Logistic Regression (KLR):** Sınıflandırma için en iyi bilinen istatistiksel model olan Logistik Regresyonun (LR) çekirdek modelidir. Bu yöntemin temel amacı, çekirdek fonksiyonlarını kullanarak orijinal girdi alanını yüksek boyutlu bir özellik alanına dönüştürmek için sınıflandırma problemlerini çözen diskriminant bir işlev bulmaktır [112].

- **Fuzzy Unordered Rule Induction Algorithm (FURIA):** Bu algoritma en gelişmiş kural öğrenicisi olarak tanınan RIPPER algoritmasının, basit ve anlaşılır kural kümeleri gibi avantajlarını

korurken genişletmiştir. Ek olarak, birçok değişiklik ve uzantı içerir. Özellikle FURIA, geleneksel kurallar yerine bulanık kuralları ve kural listeleri yerine sıralanmamış kural kümelerini öğrenir. [113].

• **Random Forest:** Popüler makine öğrenme modellerinden biri olan denetimli bir sınıflandırma yöntemidir. Onu popüler yapan en önemli özelliği, hiper parametre kestirimi yapılmasına gerek duymadan iyi sonuçlar üretmesi ayrıca regresyon ve sınıflandırma problemlerine uygulanabilmesidir. Algoritmanın temel çalışma prensibinde modelin temel bloğu olan karar ağaçları kullanılmaktadır. En basit şekliyle algoritma rastgele olarak bir orman yaratmaktadır, bu algoritmadaki ağaç sayısı ve elde edebileceği sonuç arasında doğrudan bir ilişki bulunmaktadır. Ağaç sayısı artırıldıkça daha kesin bir sonuç elde edilebilmektedir. Random Forest algoritması ile Decision Tree algoritması arasındaki en önemli fark, Random Forest yöntemindeki kök düğümün (Root Node) bulunması ve düğümlerin bölünmesi işlemlerinin rastgele çalışıyor olmasıdır. Random Forest algoritmasında kullanılan model farklı veri setleri üzerinde eğitim gerçekleştirdiği için karar ağaçlarının en büyük problemlerinden olan varyans, diğer bir deyişle overfitting azalmaktadır [114].

• **Random Tree:** Birkaç zayıf ağaç sınıflandırıcısı arasından ortalama bir sonuç elde etmek veya belirli bir ağaç sınıflandırıcısının performansı artırılmak istenebilir. Ancak aynı verilerle aynı sonuçların elde edildiği birden fazla ağaç yapısı inşa edilememektedir. Bu bakımdan iki farklı türün rastgeleliği ortaya konmaktadır; her ağaç orijinalden tekrarlanan örneklerle farklı satırlar üzerine kurulmakta ve her bir ağaçtan üretilen sütunlardan (veya bazı durumlarda dallardan) rastgele seçilen küçük bir alt sütun kümesi kullanılarak oluşturulmaktadır. Random Tree genellikle makine öğrenmesiyle ilgisi olmayan rastgele oluşturulmuş ağaçları ifade etmektedir. Bununla birlikte, popüler makine öğrenme çerçevesi olan terimler, rastgele bir sütun alt kümesi üzerine kurulu bir karar ağacına atıfta bulunmak için kullanılmaktadır [108].

• **Simple Logistic:** Logistic Regression, bağımlı değişkenlerin binary yani ikilik yapıda olduğunda başvurulması gereken en uygun regresyon analizidir. Tüm regresyon analizleri gibi, lojistik regresyon da öngörücü bir analizdir. Verileri tanımlamak ve bir bağımlı ikili değişken ile bir veya daha fazla nominal, sıra, aralık veya oran düzeyinde bağımsız değişkenler arasındaki ilişkiyi açıklamak için lojistik regresyon kullanılmaktadır. Bazen lojistik regresyonların gerilemelerinin yorumlanması zordur; Intellectus istatistikler aracı, bu analizleri kolayca yapmanıza izin vermektedir [115].

• **Iterative Classifier Optimization:** Bu algoritmada, LogitBoost gibi çapraz doğrulama ya da yüzde bölünme değerlendirmesi kullanılarak en iyi yineleme sayısı seçilmektedir. Çapraz doğrulama veya yüzde bölünme değerlendirmesi kullanılarak yinelemeli sınıflandırıcının yineleme sayısı optimize edilmektedir. İlk aşamada, daha önceden işlenmiş bir eğitim veri seti modele tanıtılmaktadır. Veriler birtakım işleme tabi tutulduktan ve model oluşturulduktan sonra model test

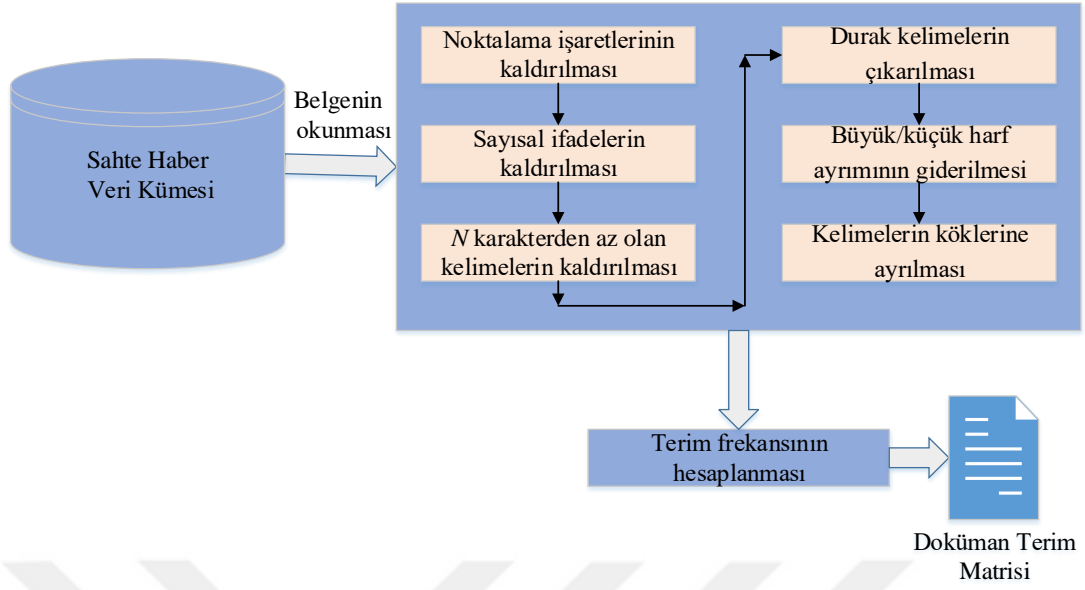
edilmekte ve sonuçlar istenen / beklenen sonuç ya da çıktı ile eşleştirilmektedir. Daha sonra algoritmanın sonuçlarını daha iyi öğrenmek ve ince ayarlamalar yapılması için geri bildirim sisteme geri gönderilmektedir. Bu durum, veri yineleme ve model eğitim iterasyonu olmak üzere iki yineleme işleminin bu kısımda gerçekleştiğini göstermektedir [116].

• **PART:** Bu algoritma, $\{0, 1, \dots, n - 1\}$ kümesindeki bölümler için kullanılan algoritmalarından biridir. Soyut olarak nitelendirilen ormanlar, bölümlerin her bir bloğu için bir ağaç koleksiyonu olarak temsil edilmektedir. Ağaç hakkında sadece ebeveyn bilgisine ihtiyaç duyulmaktadır. Bir bloğu temsil eden bir vektöre ait her bir bloğun kökü, o bloğun en az bir elementine ya da her bir kök olmayan kökünün köküne ait ise derlenmektedir. Bu biçim benzersizdir, yani, iki vektör aynı normal forma sahip ise aynı bölümü temsil etmektedirler. Buradaki örneklerin tamamı normal formdadır. Algoritma, herhangi bir elementin kökünü bulmak için basit bir özyinelemeli prosedür üretmektedir [117].

• **Stacking:** Çoklu sınıflandırma veya regresyon modeli oluşturmanın yollarından birisidir. Modelleri bir araya getirmenin çeşitli yolları bulunmaktadır, yaygın olarak bilinen modeller Boosting veya Bagging'dir. Torbalamada, varyans değerini azaltmak için yüksek değişkenliğe sahip birden fazla benzer modelin ortalaması alınmaktadır. Artırmada, sapma değeri küçük tutulurken yanlılığı azaltmak için çoklu artımlı modeller oluşturulmaktadır. İstifleme yığılmış genelleme olarak adlandırılan farklı bir paradigmadır. İstiflemenin amacı, aynı problem için farklı modellerdeki bir alanı keşfetmek üzerine kuruludur. Buradaki fikir, problemin bir bölümünü öğrenebilecek, ancak problemin tüm alanını öğrenemeyen farklı model tipleri ile bir öğrenme probleminin üstesinden gelmektir. Böylelikle, çok sayıda farklı öğrenici oluşturabilir ve bunlar, her bir öğrenilen model için bir tahmin niteliğinde ara tahmin oluşturmak için kullanılmaktadır. Ardından, ara tahminlerden aynı hedefi öğrenen yeni bir model eklenmektedir. Böylece, genel performans artırılabilir ve çoğu zaman da herhangi bir ara modelden daha iyi bir model elde edilebilmektedir [118].

5.5. Veri Ön İşleme Adımları

Öğrenme sisteminde, verilerin temsili sonuçların doğruluğunu büyük ölçüde etkiler. Özel olarak, metin analizi problemlerinde, verinin uygulanacak yöntemeye uygun olan bir gösterime dönüştürülmesi gerekir. Kullanıcılar tarafından sosyal medyada paylaşılan metin tabanlı veriler genellikle yapılandırılmamış biçimlerde. Bu nedenle sosyal medyadan çıkarılan yapılandırılmamış veriler veri ön işleme yöntemleri ile yapısal bir formata dönüştürülmelidir. Bu tez çalışmasında kullanılan sahte haber veri kümeleri üzerinde yapılan ön işlemlerin adımları Şekil 5.11'de verilmiştir.



Şekil 5.11. Veri üzerinde yapılan ön işlem adımları

Veri ön işleme adımı, sahte haber veri kümesinden verilerin okunması ile başlar. Daha sonra, tüm noktalama işaretleri, boşluk ve satır sonu karakterleri metin verisinden kaldırılır. Rakam içeren bütün terimler kaldırılır. Arından karakter sayısı N 'den küçük olan kelimeler N -karakter filtresi ile silinir [119]. Metin içinde çok sık geçen fakat önemli olmayan kelimeler durak kelimelerdir [120]. Bağlaçlar, edatlar ve zarflar gibi kelimeler, durak kelimeler olarak kabul edilir. Durak kelimeler çıkarılarak terim uzayının boyutu azaltılır. Ön işlem adımları büyük/küçük harf ayrılması giderilmesi ve tek bir biçime dönüştürülmesi ile devam eder. Bu çalışmada metin içerisinde geçen terimlerin tamamı küçük harflere dönüştürülmüştür. Bazı kelimeler aynı kökten oluşmalarına rağmen aldıkları eklere göre farklı anlamlara kazanmaktadır. Bu tip kelimelerin metin içerisinde geçme sıklığını belirlemek için kelimelerin köklerine ayrılması gerekir [121].

Metin içerikli veriler ile yapılacak çalışmalar için en büyük sorun yüksek boyutlu verilerdir. Bu nedenle, önerilen modelin başarımını artırmak için gereksiz özelliklerin metinden çıkarılması gerekir. Özellik çıkarımı, metin özellik boyutunu azaltmak ve kullanışlı veri kümesi elde etmek için kullanılır. Veri kümesindeki her bir doküman için terimler ağırlıklandırılır ve her bir doküman terimlerin ağırlık vektörüne dönüştürülür. Bu temel gösterim Vektör Uzay Modeli (VUM) olarak adlandırılır [122]. VUM'un amacı metin dokümanlarını bir özellik vektörüne dönüştürmektir [123]. VUM'da, her bir kelime, kelimenin doküman içindeki ağırlığını gösteren bir değer ile ifade edilir. Bu çalışmada terimleri ağırlıklandırmak için TF yöntemi kullanılmıştır. TF bir kelimenin bir dokümanda görülme sıklığını ifade eder ve Denklem 5.15 ile hesaplanır:

$$\text{Terim Frekansı: } TF = \frac{n_{ij}}{|d_i|} \quad (5.15)$$

Denklemdede d_i i . dokümandaki tüm terimlerin toplam sayısıdır. n_{ij} ise i . dokümanda j . kelimenin sayısını temsil eder.

Her bir dokümandaki kelimelerin TF değerleri hesaplandıktan sonra, terim ağırlıklarına göre Şekil 5.12’de gösterilen $m \times n$ boyutunda bir Doküman Terim Matrisi (DTM) oluşturulur. Matriste her bir satır dokümanları, sütunlar ise terimleri ifade eder. Her bir hücre ise, dokümandaki terimlerin ağırlıklarını gösterir.

$$\begin{matrix} D_1 \\ D_1 \\ \cdot \\ \cdot \\ \cdot \\ D_m \end{matrix} \begin{bmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mn} \end{bmatrix}$$

Şekil 5.12. Doküman Terim Matrisi

5.6. Sahte Haber Veri Kümeleri

Sahte haber tespiti, bir haber makalesinin gerçek mi yoksa sahte mi olduğuna karar vermek için yapılan işlemlerin bütünüdür. Tez çalışmasında kullanılan GKO, SSO ve iki yeni ASSO algoritmalarının sahte haber tespiti problemi üzerindeki başarılarını test etmek için genel erişime açık dört farklı veri kümesi kullanılmıştır. Veri kümelerinde sahte ve gerçek etiketine sahip iki farklı kategoride veriler bulunmaktadır. Veri ön işleme teknikleri kullanılarak sayısallaştırılan metin verileri eğitim ve test verilerine ayrılarak optimizasyon algoritmaları ile işlenebilecek uygun hale getirilmiştir.

5.6.1. BuzzFeed Politik Haber Veri Kümesi

Bu veri kümesi BuzzFeed’in Facebook’taki [124] sahte seçim haberleriyle ilgili 2016 tarihli makalesinde bulunan haberleri toplanarak oluşturulmuştur [125]. 2016 ABD Başkanlık seçimlerinden önce, BuzzFeed bu verileri elde etmek için üç aylık ayrı bölümler halinde dokuz ay boyunca gerçek ve sahte hikâyeleri incelemiştir. Sahte makaleler için, önemli seçim terimleriyle ilgisi en yüksek olan makaleleri hedeflemiş ve bunları bilinen sahte haber kaynaklarına göre filtrelemiştir. Gerçek hikâyeler için, aynı dönemde tanınmış haber kuruluşlarından en fazla ilgiyi çeken hikâyeleri bulmuşlardır. Veri kümesinde, 2016 ABD seçimiyle ilgili Facebook’tan toplanan 1,627 haber makalesi yer almaktadır. Şekil 5.13’te BuzzFeed politik haberler veri kümesinden alınan bir kesit gösterilmiştir.

1	Hillary in hot water over her email server, again.	Fake
2	The list of Republicans supporting Hillary Clinton is still growing.	Real
3	Democratic nominee Hillary Clinton is in hot water again after nearly 5 milli...	Fake
4	For months, we have been keeping track of all of the GOP politicians, adm...	Real
5	The election commission has an emergency meeting scheduled for tomorro...	Fake
6	The most recent high-profile announced was from former secretary of sta...	Real
7	Ms. Clinton has already come under fire during this election cycle over usi...	Fake
8	General Powell said at a meeting of the Long Island Association that he w...	Real
9	Ms. Clinton's public relations official released a brief statement saying, "H...	Fake
10	Powell's support isn't much of a surprise, given his support for President O...	Real
11	These ballots could change the fortune of Bernie Sanders in his bid for the...	Fake
12	In his leaked emails from last month, Powell also appeared to be talking ov...	Real
13	South African Billionaire, Femi Adenugame, has released a statement offe...	Fake
14	But at least it's official now. And Powell is now the third Bush Cabinet offic...	Real
15	Concerns about Donald Trump becoming president has prompted a South ...	Fake
16	Kathleen Kennedy Townsend said in a Facebook post that Bush told her t...	Real
17	Americans have trouble trusting the Clintons for many reasons, but near t...	Fake
18	The president's office isn't confirming it, but she stands by it.	Real
19	Now, it appears that yet another name has been added to this list.	Fake
20	That's what he said, she told Politico	Real
21	Former U.N. President John Ashe was reportedly found dead last Wednes...	Fake
22	a moderate Republican who is retiring this year, told Syracuse.com that h...	Real
23	Ashe's death became even more suspicious when police learned that he ...	Fake
24	He cited Trump's criticism of Khizr Khan.	Real
25	It was later revealed that Seng also illegally funneled several hundred tho...	Fake
26	I think Trump is a national embarrassment, Hanna said. Is he really the gu...	Real
27	?It would have been very embarrassing,? the source added. ?His death w...	Fake
28	Loose lips sink ships. Got that, Trump? Loose lips sink ships. Warner adde...	Real

Şekil 5.13. BuzzFeed politik haberler veri kümesinden bir kesit

5.6.2. Rastgele Politik Sahte Haber Veri Kümesi

BuzzFeed politik haberler veri kümesi yalnızca politik haberleri içermektedir. Horne ve Adali, BuzzFeed politik haberler veri kümesinin sınırlarını kontrol etmek ve analiz bakımında güçlendirmek için yeni bir politik haber veri kümesi oluşturmuşlardır [125]. Sahte haber kaynakları, Zimdars'ın sahte ve yanıltıcı haber siteleri listesinden [126], gerçek haber kaynakları Business Insider'ın "MostTrusted" listesinden [127] alınmıştır. Gerçek ve sahte haberlerin alındığı kaynakların bir kısmı Tablo 5.2'de listelenmiştir. Şekil 5.14'te rastgele sahte haber veri kümesinden alınan bir kesit gösterilmiştir.

1	A quick trip down memory lane causes a stumble over this gem from Obama. He was gearing up for his first run at the office of President and was spewing lies all over the American public.	Fake
2	After publicly conceding electoral defeat last week, President Yahya Jammeh of Gambia has reversed course and is calling for a new election.	Real
3	"I'll make our government open and transparent so that anyone can ensure that our business is the people's business."	Fake
4	Jammeh has ruled the tiny West African country since seizing power in a coup in 1994, and his public concession to President-elect Adama Barrow on Dec. 2 led to hopes of the first pea...	Real
5	He is finally living up to that with less than a month left to his Presidency. He is not bothering to hide the dirty business he conducts, because he just doesn't care anymore. Recent mont...	Fake
6	On Friday, Jammeh said the Independent Electoral Commission made errors in vote tallies.	Real
7	Many are speculating that he will step up and stop talking about the Hillary email scandal and simply pardon her from any prosecution. It is in his best interest to do so. His attempts to m...	Fake
8	In the same way that I accepted faithfully the results, believing that the IEC was independent and honest and reliable, I hereby reject the results in totality, Jammeh said in a televised ...	Real
9	A reporter with first-hand knowledge sums up Obama's "transparent" government, "Obama and his team criminalized reporting, becoming the most secretive White House since Richard ...	Fake
10	He also said there were other irregularities and problems in the electoral process.	Real
11	A message on the White House website from Obama makes it clear he still maintains the illusion that his office upheld the pledge of transparency;	Fake
12	Our investigations reveal that in some cases, voters were told that the opposition has already won and there was no need for them to vote, Jammeh said.	Real
13	"My Administration is committed to creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency...	Fake
14	Last week, supporters of the opposition took to the streets to celebrate Barrow's win. Jammeh's allegations are now plunging Gambia into confusion and uncertainty, NPR's Ofebea Quis...	Real
15	Every action taken by Barack Hussein Obama was calculated. He knew, as did his advisors and supporters, that he was actively engaged in deception against the United States. David A...	Fake
16	President-elect Barrow responded, telling reporters, The outgoing president has no constitutional authority to reject the result of the election and order for fresh elections to be held, R...	Real
17	He was positioned to not only witness the secrecy that Obama demanded but he was also part of it. After leaving the Obama administration, Axelrod began to hint that things were quit...	Fake
18	The U.S. joined international bodies in condemning the announcement.	Real
19	Some of his revelations regarding Obama continue to come as a shock to people who the President painted with the colors of truth.	Fake
20	This action is a reprehensible and unacceptable breach of faith with the people of The Gambia and an egregious attempt to undermine a credible election process and remain in power ille...	Real
21	Obama lied about his support FOR same-sex marriage before announcing his support for it in 2012. "Yet if Obama's views were 'evolving' publicly, they were fully evolved behind closed ...	Fake
22	The Economic Community of West African States, the U.N. and African Union echoed the sentiment in a joint statement:	Real
23	Obama chewed out Maureen Dowd on the campaign plane during the 2008 campaign. "No one got under Barack's skin more than Maureen ... He was patronizing and disrespectful." This i...	Fake
24	They call on the government of The Gambia to abide by its constitutional responsibilities and international obligations. It is fundamental that the verdict of the ballots should be respecte...	Real
25	"Obama can have a short fuse, though he hides it from the public; The president called Axelrod a "mother---" and stalked out of a meeting after his strategist criticized the president's ...	Fake
26	Wire services report the streets of the capital Banjul were calm Saturday, with a heavy presence of police and soldiers. Gambians closed down shops and stayed home out of fear of viol...	Real
27	While those incidents speak to Obama's two-faced behavior, some of the secrecy of his administration can be summed up as simply lies;	Fake
28	Human rights groups have criticized Jammeh for abuses during his 22-year rule. In its 2016 report, Human Rights Watch said his government frequently committed serious human rights ...	Real

Şekil 5.14. Rastgele sahte haber veri kümesinden bir kesit

Tablo 5.2. Rastgele sahte haber veri kümesinin haber kaynakları

Gerçek kaynaklar	Sahte Kaynaklar
Wall Street Journal	Ending the Fed
The Economist	True Pundit
BBC	abc.news.com.co
NPR	DC Gazette
ABC	libertywritersnews
CBS	Before its News
USA Today	Infowars
The Guardian	Real News Right Now
NBC	
Washington Post	

5.6.3. LIAR Sahte Haber Veri Kümesi

Bu veri seti, toplanan toplam 12,836 kısa ifadeden oluşmaktadır. Siyasi tartışmalar, Facebook gönderileri, tweetler ve röportajlar dâhil olmak üzere çeşitli içeriklerden elde edilmiştir. Bu ifadeler PolitiFact'in Application Programming Interface (API)'sinden alınmıştır, her bir ifade PolitiFact editörü tarafından doğruluk açısından değerlendirilmiştir [61]. Şekil 5.15'te LIAR sahte haber veri kümesinden alınan bir kesit gösterilmiştir.

1	Building a wall on the U.S.-Mexico border will take literally years.	Real
2	Wisconsin is on pace to double the number of layoffs this year.	Fake
3	Says John McCain has done nothing to help the vets.	Fake
4	Suzanne Bonamici supports a plan that will cut choice for Medicare Advantage seniors.	Fake
5	When asked by a reporter whether hes at the center of a criminal scheme to violate campaign laws, Gov. Scott Walker nodded yes.	Fake
6	Over the past five years the federal government has paid out \$601 million in retirement and disability benefits to deceased former federal employees.	Real
7	Says that Tennessee law requires that schools receive half of proceeds -- \$31 million per year -- from a half-cent increase in the Shelby County sale...	Real
8	Says Vice President Joe Biden admits that the American people are being scammed with the economic stimulus package.	Fake
9	Donald Trump is against marriage equality. He wants to go back.	Real
10	We know that more than half of Hillary Clintons meetings while she was secretary of state were given to major contributors to the Clinton Foundation.	Fake
11	We know there are more Democrats in Georgia than Republicans. We know that for a fact.	Fake
12	PolitiFact Texas says Congressman Edwards attacks on Bill Flores are Fake.	Fake
13	Denali is the Kenyan word for black power.	Fake
14	Says 57 percent of federal spending goes to the military and just 1 percent goes to food and agriculture, including food stamps.	Fake
15	On residency requirements for public workers	Fake
16	Says the unemployment rate for college graduates is 4.4 percent and over 10 percent for noncollege-educated.	Real
17	Unfortunately we have documented instances where people defecated in the (Statehouse) building.	Fake
18	A recent Gallup poll found that 72 percent of Americans and 56 percent of Democrats say the biggest threat to our nations security is big governme...	Fake
19	Each year, 18,000 people die in America because they don't have health care.	Real
20	Ronald Reagan faced an even worse recession than the current one.	Fake
21	There have not been any public safety issues in cities that allow transgender people to use the bathroom of the gender they identify as.	Fake
22	Says Mitt Romney was one of the first national Republican leaders to endorse Marco Rubio.	Fake
23	The number of illegal immigrants could be 3 million. It could be 30 million.	Fake
24	Marijuana is less toxic than alcohol.	Fake
25	Says Charlie Crist is embroiled in a fraud case for steering taxpayer money to a de facto Ponzi scheme.	Fake
26	Now, there was a time when someone like Scalia and Ginsburg got 95-plus votes.	Real
27	I was gone when there was a red line against Syria.	Fake
28	Tim Kaine hiked tuition as governor, but now claims to champion affordability.	Fake

Şekil 5.15. LIAR sahte haber veri kümesinden bir kesit

5.6.4. ISOT Sahte Haber Veri Kümesi

Bu veri kümesi sahte ve gerçek olmak üzere iki tip haber içermektedir. Veri kümesi 21,417 tanesi gerçek ve 23,481 tanesi sahte haber olmak üzere toplamda 44,848 haber makalesi içermektedir. Sahte ve gerçek haberler gerçek dünya kaynaklarından elde edilmiştir. Gerçek haberler Reuters.com'dan toplanırken, sahte haberler için PolitiFact ve Wikipedia gibi farklı web

sayfalarından elde edilmiştir [55]. ISOT sahte haber veri kümesinden alınan bir kesit Şekil 5.16'da gösterilmiştir.

1	This is Disturbing,Donald Trump just couldn t wish all Americans a Happy New Year and leave it at that. Instead, he had to give a shout out t...	Fake
2	House Intelligence Committee Chairman Devin Nunes is going to have a bad day. He s been under the assumption, like many of us, that the ...	Fake
3	On Friday, it was revealed that former Milwaukee Sheriff David Clarke, who was being considered for Homeland Security Secretary in Donald ...	Fake
4	On Christmas day, Donald Trump announced that he would be back to work the following day, but he is golfing for the fourth day in a row. ...	Fake
5	Pope Francis used his annual Christmas Day message to rebuke Donald Trump without even mentioning his name. The Pope delivered his mes...	Fake
6	The number of cases of cops brutalizing and killing people of color seems to see no end. Now, we have another case that needs to be shared...	Fake
7	Donald Trump spent a good portion of his day at his golf club, marking the 84th day he s done so since taking the oath of office. It must have...	Fake
8	In the wake of yet another court decision that derailed Donald Trump s plan to bar Muslims from entering the United States, the New York Ti...	Fake
9	Many people have raised the alarm regarding the fact that Donald Trump is dangerously close to becoming an autocrat. The thing is, democ...	Fake
10	Just when you might have thought we d get a break from watching people kiss Donald Trump s ass and stroke his ego ad nauseam, a pro-Tru...	Fake
11	A centerpiece of Donald Trump s campaign, and now his presidency, has been his white supremacist ways. That is why so many of the public ...	Fake
12	Republicans are working overtime trying to sell their scam of a tax bill to the public as something that directly targets middle-class and workin...	Fake
13	Republicans have had seven years to come up with a viable replacement for Obamacare but they failed miserably. After taking a victory lap f...	Fake
14	The media has been talking all day about Trump and the Republican Party s scam of a tax bill	Fake
15	Abigail Disney is an heiress with brass ovaries who will profit from the GOP tax scam bill but isn t into f-king poor people over. Ms. Disney pe...	Fake
16	Donald Trump just signed the GOP tax scam into law. Of course, that meant that he invited all of his craven, cruel GOP sycophants down fro...	Fake
17	A new animatronic figure in the Hall of Presidents at Walt Disney World was added, where every former leader of the republic is depicted in a...	Fake
18	Trump supporters and the so-called president s favorite network are lashing out at special counsel Robert Mueller and the FBI. The White Ho...	Fake
19	Right now, the whole world is looking at the shocking fact that Democrat Doug Jones beat Republican Roy Moore in the special election to rep...	Fake
20	Senate Majority Whip John Cornyn (R-TX) thought it would be a good idea to attack Special Counsel Robert Mueller over the Russia probe. A...	Fake
21	Trump Invites NRA To Xmas Party On Sandy Hook Anniversary	Fake
22	In this #METOO moment, many powerful men are being toppled. It spans many industries, from entertainment, to journalism, to politics and ...	Fake
23	As a Democrat won a Senate seat in deep-red Alabama, social media offered up everyone s opinion because that s what social media does. D...	Fake
24	Alabama is a notoriously deep red state. It s a place where Democrats always think that we have zero chances of winning especially in state...	Fake
25	A backlash ensued after Donald Trump launched a sexist rant against Kirsten Gillibrand Thursday morning, saying that the Democratic Senato...	Fake
26	Donald Trump is afraid of strong, powerful women. He is a horrific misogynist, and has shown himself to be so over and over again. That is n...	Fake
27	Ronald Reagan is largely seen as the Messiah of the Republican Party. Despite how long it has been since the man was president, he has alw...	Fake
28	Judge Jeanine Pirro has continued her screamy ragey meltdown over special counsel Robert Mueller s investigation into any possible collusio...	Fake

Şekil 5.16. ISOT sahte haber veri kümesinden bir kesit

5.7. Performans Değerlendirme Ölçütleri

Tez çalışmasında kullanılan optimizasyon yöntemleri ve denetimli yapay zekâ algoritmalarının performanslarını ölçmek ve etkinliklerini değerlendirmek için dört farklı performans değerlendirme ölçütünden yararlanılmıştır. Öncelikle sahte haber tespiti problemi için Tablo 5.3'te verilen karmaşık matrisi oluşturulmuştur.

Tablo 5.3. Sahte haber için karmaşıklık matrisi

Karmaşıklık matrisi		Gerçek sınıf	
		Pozitif Sınıf	Negatif Sınıf
Tahmin edilen sınıf	Pozitif Sınıf	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif Sınıf	Yanlış Negatif (YN)	Doğru Negatif (DN)

Matriste verilen terimler (DP, YP, YN, DN) sahte haber problemi için aşağıdaki gibi ifade edilmiştir.

- Doğru Pozitif (DP): Tahmin edilen sahte haber aslında sahte bir haber ise, tahmin DP'dir.
- Yanlış Pozitif (YP): Tahmin edilen sahte haber aslında gerçek bir haber ise, tahmin YP'dir.
- Doğru Negatif (DN): Tahmin edilen gerçek haber aslında gerçek bir haber ise, tahmin DN'dir.

- Yanlış Negatif (YN): Tahmin edilen gerçek haber aslında sahte bir haber ise, tahmin YN'dir.

Karmaşıklık matrisindeki bu değerlere göre, kullanılan yöntemlerin performanslarını karşılaştırmak için Tablo 5.4'teki değerlendirme ölçütleri kullanılmıştır.

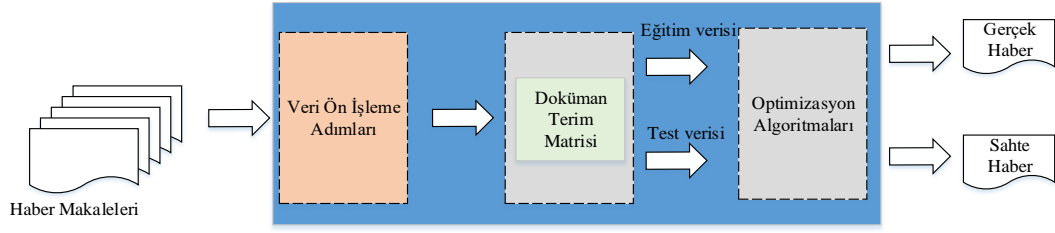
Tablo 5.4. Kullanılan performans değerlendirme ölçütleri

Performans Değerlendirme Ölçütü	Formül
Doğruluk	$\frac{DP + DN}{DP + YP + DN + YN}$
Kesinlik	$\frac{DP}{DP + YP}$
Duyarlılık	$\frac{DP}{DP + YN}$
F-Ölçütü	$\frac{2 * Kesinlik * Duyarlılık}{Kesinlik + Duyarlılık}$

Sahte haber tespiti problemi için doğruluk değeri, tüm veri kümesi içerisinde doğru tahmin edilen haberlerin orandır. Kesinlik, sahte olarak tahmin edilen haberler içerisindeki, doğru tahmin edilen sahte haberlerin oranını gösterir. Duyarlılık ise, tahmin edilen sahte haberlerin, tüm sahte haberlere oranını ifade eder. F-ölçütü ise, kesinlik ve duyarlılık değerlerinin harmonik ortalamalarıdır.

5.8. Optimizasyon Algoritmalarının Sahte Haber Tespiti için Modellenmesi

Ön işlemden geçirilen veri kümeleri sonrasında elde edilen DTM'nin boyutu optimizasyon algoritmalarında popülasyon bireylerinin oluşturulmasında kullanılan temel kriterdir. GKO, SSO ve ASSO algoritmaları için başlangıç popülasyonları 0-1 aralığında rastgele değerlerden oluşmuştur. Oluşturulan başlangıç popülasyonunun boyutu her bir algoritma için 30 olarak belirlenmiştir. Problem boyutu (M) ise, her bir veri kümesi için ön işlemler yapıldıktan sonra elde edilen doküman matrisindeki terim sayısı ile ilişkilendirilmiştir. İterasyon sayısı standart olarak tüm algoritmalar için 1000 olarak belirlenmiştir. Tez çalışmasında optimizasyon yöntemlerini kullanarak sahte haber tespiti yapmak için izlenen adımlar genel itibarı ile Şekil 5.17'deki gibidir.



Şekil 5.17. Sahte haber tespiti için önerilen yöntemin yapısı

Tez çalışmasında metinler arasındaki benzerliği ölçmek için Jaccard benzerliği kullanılmıştır. $X_i = X_{i1}, X_{i2}, \dots, X_{iM}$ popülasyondaki i . birey ile $D_j = D_{j1}, D_{j2}, \dots, D_{jM}$ doküman matrisindeki j . haber satırına karşılık gelen vektör değerleri olarak alınırsa, dokümanlar arasındaki Jaccard benzerliği (5.16) ile hesaplanır [128].

$$jaccard_değeri_{ij} = \frac{|X_i \cap D_j|}{|X_i \cup D_j|} \quad (5.16)$$

Bireylerin uygunluğunu hesaplamak için kullanılan uygunluk fonksiyonu (5.17) ile ifade edilir. Sonuç olarak en iyi uygunluk değerine sahip birey en iyi çözüm olarak kabul edilir [43].

$$Uygunluk\ fonksiyonu = a_1 * (Doğruluk) + a_2 * (Kesinlik) + a_3 * (Duyarlılık) + a_4 * (F - ölçütü) \quad (5.17)$$

Denklemden 4 farklı performans değerlendirme ölçütü eş zamanlı olarak ele alınarak bir uygunluk fonksiyonu oluşturulmuştur.

5.9. Friedman Testi

Bu tez çalışmasında GKO, SSO ve önerilen iki yeni adaptif SSO algoritmalarının performanslarını analiz etmek için parametrik olmayan Friedman testi kullanılmıştır. Friedman testi birbiri ile ilişkili iki ya da daha fazla değişkene ait dağılımların karşılaştırılarak dağılımlar arasında anlamlı fark olup olmadığını test etmek amacıyla kullanılır. Diğer bir deyişle, örneklemin tekrarlanan ölçümleri arasında fark olup olmadığı test edilir.

Friedman testi için aşağıdaki varsayımlar geçerlidir:

- Rastgele örnekleme
- Tekrarlanan ölçümler
- Her gruptaki örneklem kendi içinde bağımsız olmalıdır.

Friedman testi için X_R^2 test istatistiği Denklem 5.18 ile hesaplanır.

$$X_R^2 = \frac{12}{nk(k+1)} \left[\sum_{j=1}^k R_j^2 \right] - [3n(k+1)] \quad (5.18)$$

Denklemdaki semboller ve anlamları Tablo 5.5'te verilmiştir.

Tablo 5.5. (5.18) denklemindeki semboller ve anlamları

Sembol	Anlamı
n	Satır sayısı
k	Grup sayısı
R_j	Her bir gruba ilişkin sıra sayıları toplamı
X_R^2	İstatistiği seçilen yanılma düzeyinde $k-1$ serbestlik dereceli ki-kare dağılımı

Friedman testi için uygulanan adımlar aşağıdaki gibidir:

1. Sıfır hipotez (H_0) ve karşıt hipotez (H_1) belirlenir. Tez çalışmamızdaki sahte haber tespiti problemine uygulanan GKO, SSO ve önerilen iki yeni adaptif SSO algoritmaları için hipotezler şöyle belirlenmiştir:
 H_0 : Kullanılan dört optimizasyon algoritması ile alınan sonuçlar arasındaki fark istatistiksel olarak anlamlı değildir.
 H_1 : Kullanılan dört optimizasyon algoritması ile alınan sonuçlar arasındaki fark istatistiksel olarak anlamlıdır.
2. Bir alpha (α) belirlenir. $\alpha = 0.05$ olarak alınmıştır.
3. Serbestlik derecesi (df) hesaplanır. Serbestlik derecesi $df = k - 1$ ile hesaplanır. $k = 4$ olduğundan, bu çalışmada $df = 3$ olarak hesaplanır.
4. Karar kuralı belirlenir. α ve df değeri kullanılarak ki-kare dağılım tablosuna bakılırsa X_R^2 değeri 7.81473 olduğu görülür.
5. Test istatistiği hesaplanır.
6. Sonuçlar değerlendirilir. Elde edilen X_R^2 değeri 7.81473'ten büyük ise H_0 reddedilir ve elde edilen sonuçlar arasındaki fark anlamlıdır.

5.10. Kutu Grafiği

Kutu grafiği istatistiksel bir grafik gösterimi olup niceliksel verileri görsel şekilde özetlemek için kullanılır. İlgilenilen değişken bakımından veri için hazırlanan beş sayılı özetleme tablosu gösterimini grafiksel olarak özetlemeye dayalıdır. Tez çalışmasında, kutu grafiği ile ifade edilecek ilgili değişken 30 kez çalıştırılan optimizasyon algoritmalarından elde edilen uygunluk fonksiyonu değerleridir. Bu değerler için kutu grafiği oluşturmak için öncelikle veri kümesi için aşağıdaki değerler hesaplanır.

X_{\min} : En küçük uygunluk fonksiyonu değeri

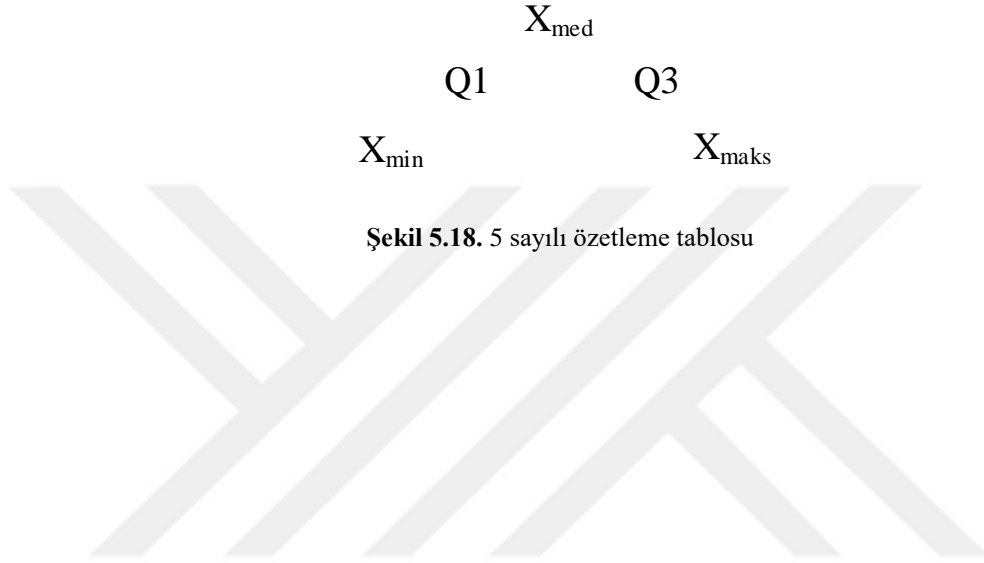
Q1: Birinci dörtebirlik

X_{med} : Medyan (ortanca)

$Q3$: Üçüncü dörttebirlik

X_{maks} : En büyük uygunluk fonksiyonu değeri

Bu değerleri hesaplamak için içerisinde optimizasyon algoritmalarından elde edilen ve 30 farklı değer bulunan tablo küçükten büyüğe sıralanır ve Şekil 5.18'de gösterildiği gibi 5 sayılı özetleme tablosu çıkarılır. Bu değerlere göre kutu grafikleri çizdirilir.



Şekil 5.18. 5 sayılı özetleme tablosu

6. DENEYSEL BULGULAR VE TARTIŞMA

Bu tez çalışmasında bir çevrimiçi sosyal ağ analizi problemi olan “sahte haber tespiti” probleminin sınıflandırılmasında ilk defa bir çözüm yöntemi olarak GKO, SSO ve ilk defa bu tez kapsamında önerilen yeni iki adaptif SSO metasezgisel optimizasyon algoritmaları kullanılmıştır. Bu algoritmalar gerçek dünya verilerinden toplanan genel erişime açık dört farklı veri kümesi üzerinde uygulanmıştır. İlk olarak, bu algoritmalarından elde edilen sonuçlar ile bu veri kümelerinin WEKA yazılımı içerisinde yer alan 30 ayrı sınıflandırma algoritmasından elde edilen sonuçlar sınıflandırma ölçütleri kullanılarak karşılaştırılmıştır. Optimizasyon algoritmalarının 30 defa çalıştırılması sonucunda elde edilen en iyi uygunluk değerleri, bu değerler için yapılan Friedman istatistik testi sonuçları ve kutu grafikleri verilmiştir. Son olarak ise, sahte haber tespiti problemi için kullanılan dört optimizasyon algoritması ve 30 ayrı sınıflandırma algoritmasından elde edilen sonuçların dört farklı performans değerlendirme kriterine göre çizilmiş grafiksel gösterimleri verilmiştir. Performans karşılaştırılması yapılırken denetimli yapay zekâ algoritmaları için standart parametre değerleri kullanılmıştır. SSO, GKO, ASSO1 ve ASSO2 optimizasyon algoritmaları için problem ve algoritma başlangıç parametrelerinin değerleri Tablo 6.1’de verilmiştir.

Tablo 6.1. Optimizasyon algoritmalarının parametreleri ve değerleri

GKO	Katsayı vektörü (\vec{A}): $[-2a, 2a]$ (\vec{a}): 2’den 0’a doğru azalan değer Katsayı vektörü (\vec{C}): Rastgele bir sayı	Alt sınır (lb): 0 Üst sınır (ub): 1 En yüksek iterasyon sayısı: 1000 Popülasyon sayısı: 30 Deney sayısı: 30
SSO	Denge parametresi (c_1): $(2e^{-\frac{it}{L}})^2$	
ASSO1	w_{min} : 0 w_{max} : 2	
ASSO2	w_{min} : 0.3 w_{max} : 0.7 k : 7	

6.1. BuzzFeed Politik Haberler Veri Kümesinden Elde Edilen Sonuçlar

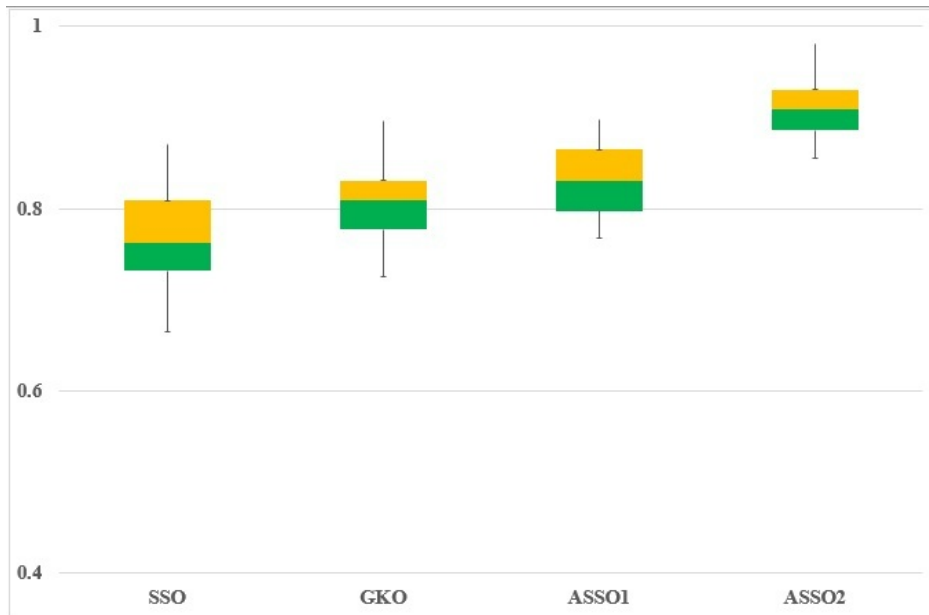
Öncelikle BuzzFeed politik haberler veri kümesi üzerinde, bir önceki bölümde söz edilen ön işlem adımları yapılmış ve doküman terim matrisi oluşturulmuştur. Ön işlemler sonrasında bu veri kümesi için 38 adet özellik elde edilmiştir. Bu nedenle optimizasyon algoritmaları BuzzFeed politik haberler veri kümesi için çalıştırılırken problem boyutu 38 olarak belirlenmiştir. Tüm optimizasyon yöntemleri veri kümesi için 30 kez çalıştırdıktan elde edilen sonuçların dört farklı performans değerlendirme ölçütü bakımından değerleri Tablo 6.2’de verilmiştir. Tablo 6.2’de görüldüğü gibi

ASSO2 algoritması bütün performans değerlendirme ölçütleri bakımından en iyi değeri vermiştir. ASSO2 ise tüm ölçütlere göre ikinci en iyi değeri vermiştir.

Tablo 6.2. Optimizasyon algoritmalarının BuzzFeed politik haberler veri kümesine uygulanması sonucunda elde edilen sonuçlar

	Doğruluk	Kesinlik	Duyarlılık	F-ölçütü
SSO	0.803	1.000	0.723	0.839
GKO	0.875	1.000	0.722	0.838
ASSO1	0.884	1.000	0.783	0.878
ASSO2	0.991	1.000	0.964	0.982

BuzzFeed politik haberler veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarının 30 kez çalıştırılmasıyla elde edilen uygunluk değerlerine göre çizilen kutu grafiği Şekil 6.1’de verilmiştir. Grafikteki değerler, elde edilen uygunluk değerlerinin dağılımını gösterir. Kutunun alt ve üst uçları, değerlerin alt ve üst çeyrekleridir. Kutu içindeki çizgi medyan değerini gösterir. Grafik incelendiğinde, SSO algoritmasından elde edilen alt çeyrek, üst çeyrek ve medyan değerleri GKO, ASSO1 ve ASSO2 algoritmalarından elde edilen değerlerden küçük olduğu görülmektedir. Ayrıca, ASSO2 algoritmasından elde edilen uygunluk fonksiyonu değerlerinin dağılımı diğer 3 algoritmadan daha yüksektir. Aynı zamanda bu değerler Friedman testi ile değerlendirilerek arasında anlamlı bir fark olup olmadığına bakılmıştır. Uygunluk fonksiyon değerlerine göre Friedman testi ile elde edilen algoritma ortalama sıralamaları Tablo 6.3’te verilmiştir.



Şekil 6.1. SSO, GKO, ASSO1 ve ASSO2 algoritmaları ile BuzzFeed veri kümesi için elde edilen sonuçların kutu grafiği

Tablo 6.3. BuzzFeed politik veri kümesi için ortalama fonksiyon değerlerine göre Friedman testi ile elde edilen algoritma ortalama sıralamaları

Algoritma isimleri	Ortalama Sıralama
SSO	1.43
GKO	2.20
ASSO1	2.60
ASSO2	3.87

Tablo 6.4. BuzzFeed politik veri kümesi için Friedman testi sonuçları

n	30
Ki-kare değeri	57.160
df	3
p -değeri	0.000

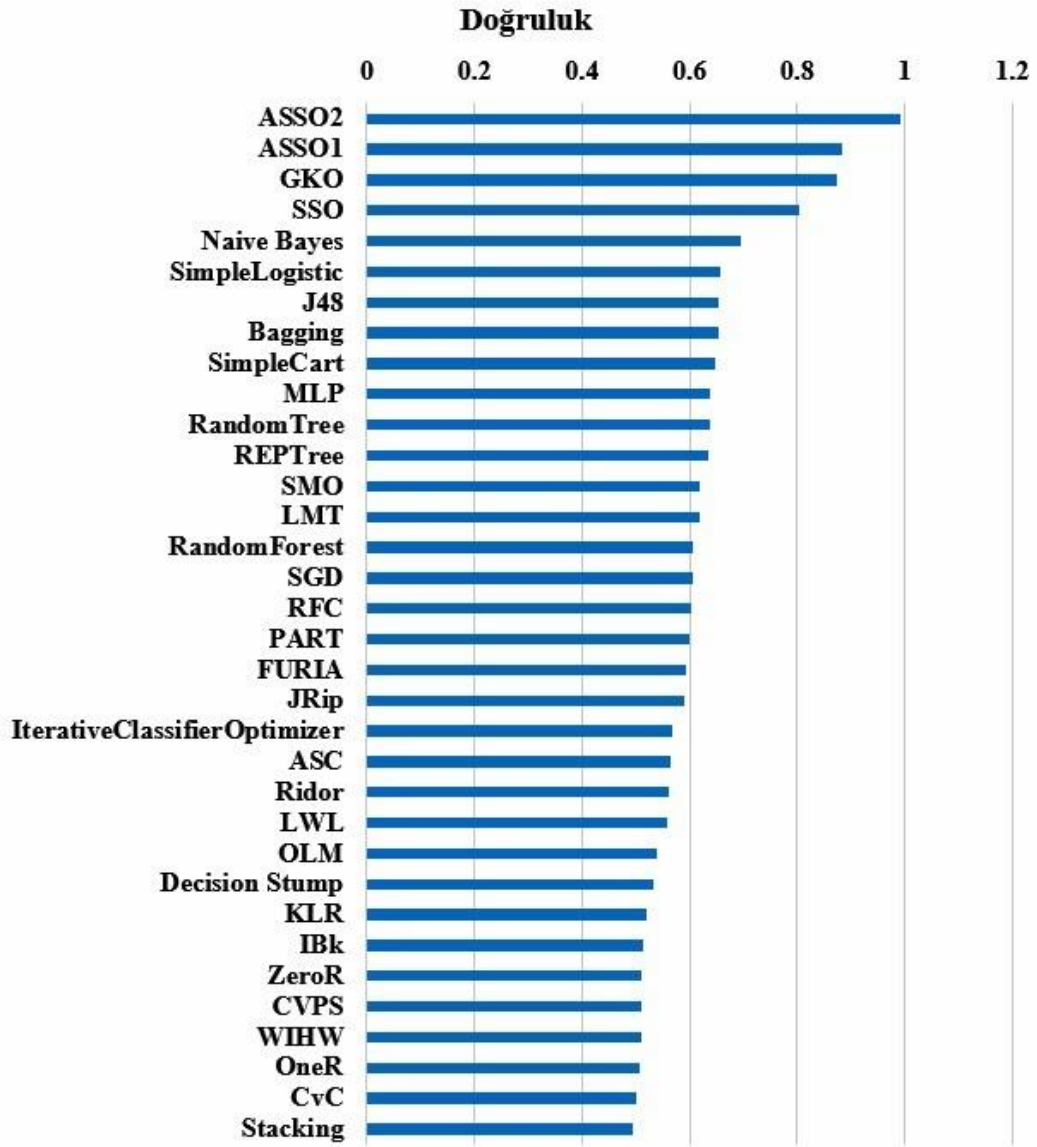
Tablo 6.4'te görüldüğü gibi BuzzFeed veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarının elde ettiği sonuçlara göre elde edilen Friedman testi değerlerinde X_R^2 değeri 57.160'dır. Ki-kare tablosunda $\alpha = 0.05$ ve $df = 3$ değeri için gösterilen değer (X_{tablo}^2) 7.81473 olarak görülmektedir. $57.160 > 7.81473$ olduğu için H_0 hipotezi reddedilir. Yani, Friedman testine göre BuzzFeed politik haberler veri kümesinde SSO, GKO, ASSO1 ve ASSO2 algoritmalarından alınan sonuçlar arasındaki fark istatistiksel olarak anlamlıdır.

BuzzFeed politik haberler veri kümesi için 30 farklı yapay zekâ algoritması dört farklı performans değerlendirme kriterine göre elde edilen sonuçları Tablo 6.5'te verilmiştir. Tabloda gösterilen değerlere Naive Bayes yapay zekâ algoritmaları içerisinde en yüksek doğruluk değerini veren algoritmadır. Kesinlik kriteri için sonuçlar incelendiğinde Decision Stump algoritması ile en iyi değerin elde edildiği görülmektedir. CvC algoritması 30 farklı yapay zekâ algoritması içinde duyarlılık bakımından en yüksek değeri vermiştir. Doğruluk değerindeki gibi Naive Bayes, F-ölçütü içinde en iyi değeri veren algoritmadır.

Tablo 6.5'te verilen 30 farklı yapay zekâ algoritması ile elde edilen sonuçlar, SSO, GKO ve ilk kez bu çalışmada önerilen ASSO1 VE ASSO2 adaptif yöntemlerinin sonuçları ile kıyaslanmıştır. İçerisine optimizasyon yöntemlerinin de dahil edildiği 34 farklı yöntemin doğruluk, kesinlik, uyarlılık ve F-ölçütü için elde edilen sonuçları sırasıyla Şekil 6.2, Şekil 6.3, Şekil 6.4 ve Şekil 6.5'te grafiksel olarak gösterilmiştir. Şekillerdeki yöntemlerden elde edilen değerler büyükten küçüğe olacak şekilde sıralanmıştır.

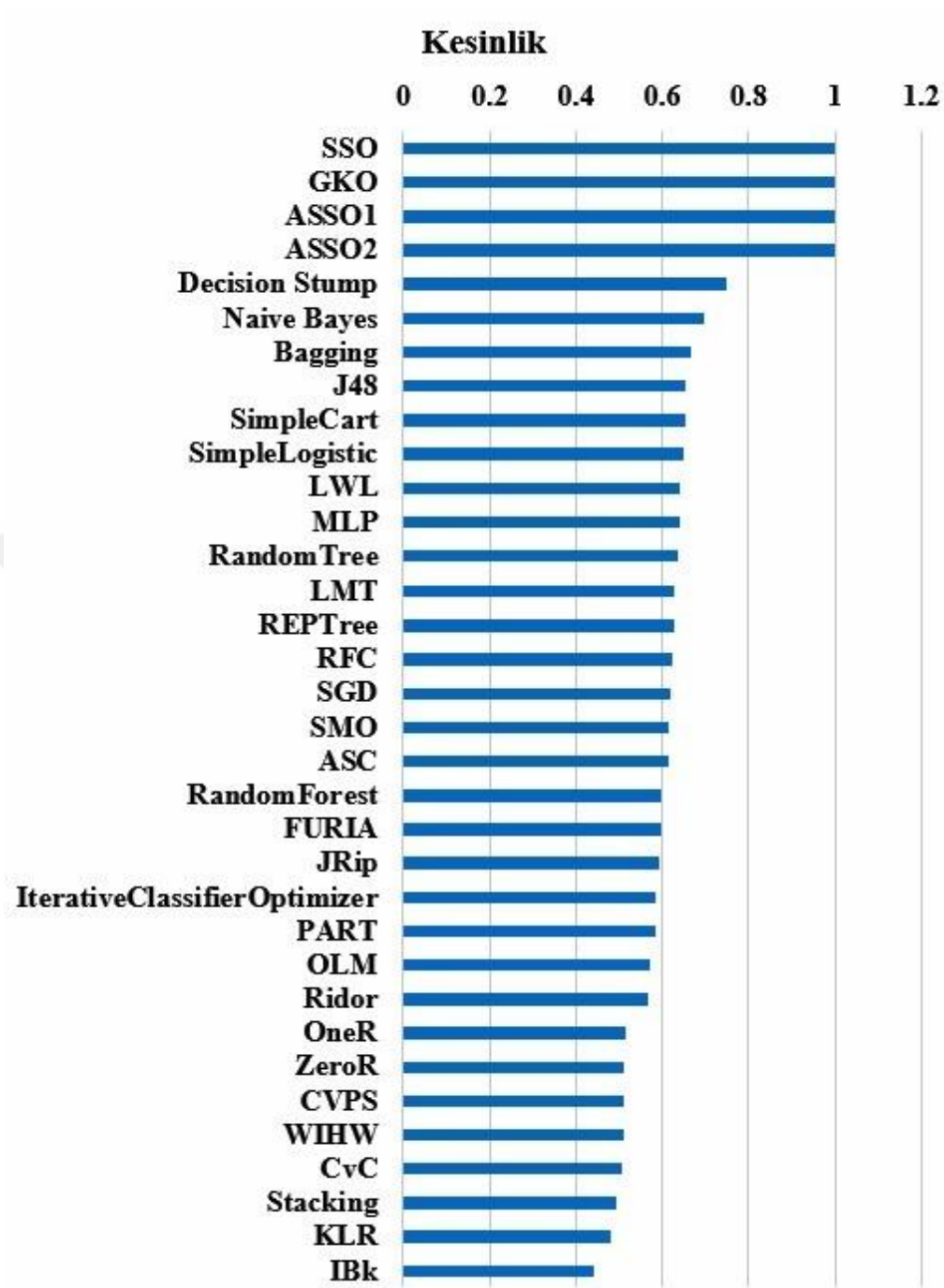
Tablo 6.5. BuzzFeed politik haberler veri kümesi için denetimli yapay zekâ algoritması ile elde edilen sonuçlar

	Doğruluk	Kesinlik	Duyarlılık	F-ölçütü
Naive Bayes	0.696	0.697	0.716	0.706
SMO	0.619	0.616	0.629	0.622
JRip	0.589	0.592	0.626	0.609
OneR	0.507	0.514	0.639	0.569
SGD	0.605	0.619	0.590	0.604
Bagging	0.653	0.666	0.642	0.653
Decision Stump	0.532	0.747	0.534	0.534
ZeroR	0.510	0.509	1.000	0.675
CVPS	0.510	0.509	1.000	0.675
RFC	0.604	0.621	0.574	0.604
LMT	0.619	0.627	0.621	0.627
LWL	0.558	0.642	0.558	0.490
CvC	0.501	0.507	0.777	0.613
WIHW	0.510	0.509	1.000	0.675
Ridor	0.562	0.567	0.592	0.579
MLP	0.638	0.640	0.639	0.639
OLM	0.538	0.573	0.573	0.489
SimpleCart	0.646	0.654	0.649	0.652
ASC	0.563	0.616	0.563	0.516
J48	0.655	0.655	0.681	0.668
REPTree	0.634	0.626	0.707	0.664
IBk	0.513	0.441	0.480	0.460
KLR	0.521	0.481	0.583	0.527
FURIA	0.594	0.596	0.595	0.592
RandomForest	0.607	0.597	0.634	0.615
RandomTree	0.638	0.634	0.634	0.634
SimpleLogistic	0.656	0.647	0.670	0.658
IterativeClassifierOptimizer	0.568	0.583	0.568	0.543
PART	0.599	0.583	0.661	0.619
Stacking	0.493	0.493	0.521	0.661



Şekil 6.2. BuzzFeed politik haberler veri kümesi için elde edilen doğruluk değerleri

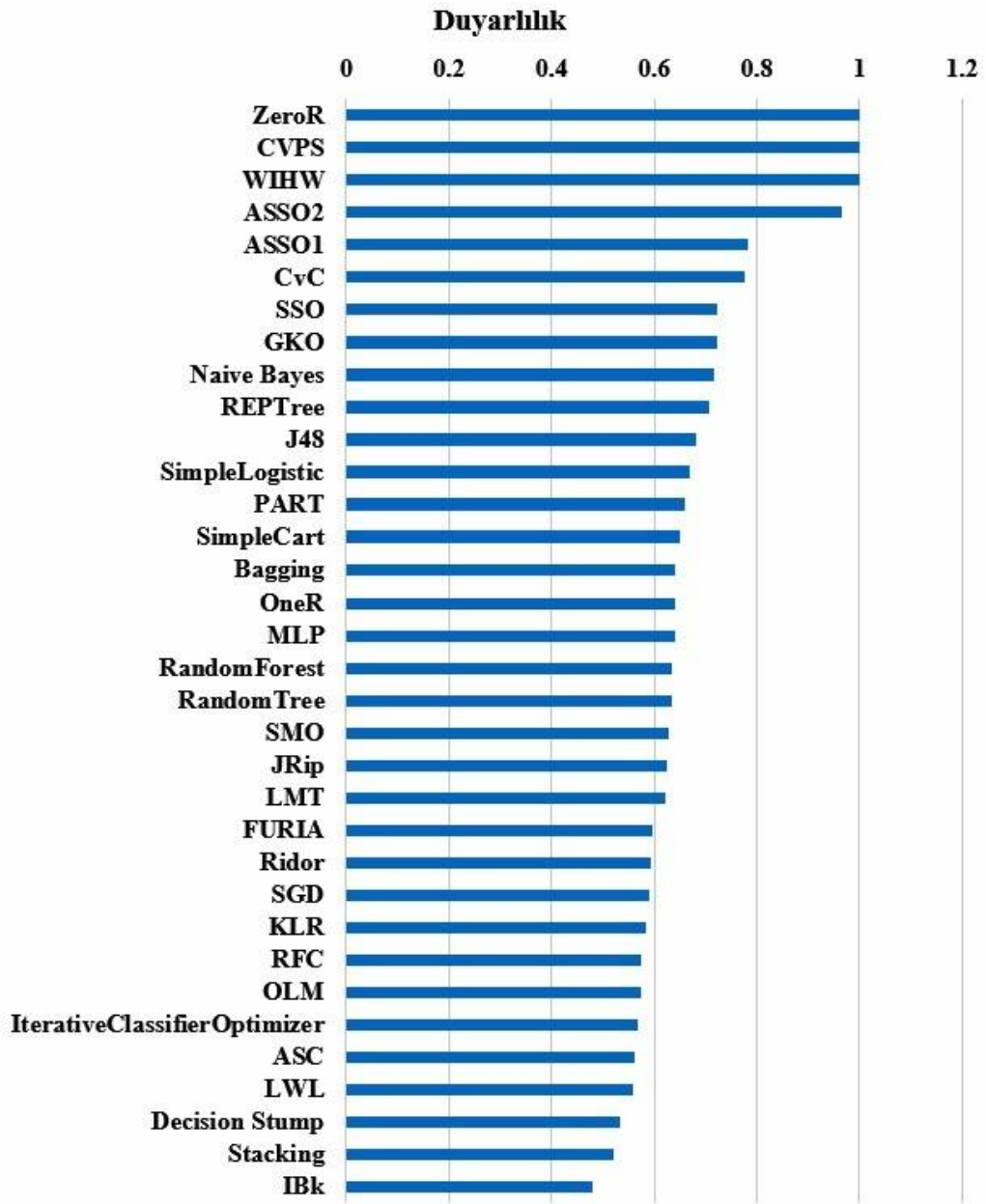
Şekil 6.2’de görüldüğü ASSO2 doğruluk kriterine göre en yüksek değeri vermiştir. ASSO1, GKO ve SSO algoritmaları da onu izlemektedir.



Şekil 6.3. BuzzFeed politik haberler veri kümesi için elde edilen kesinlik değerleri

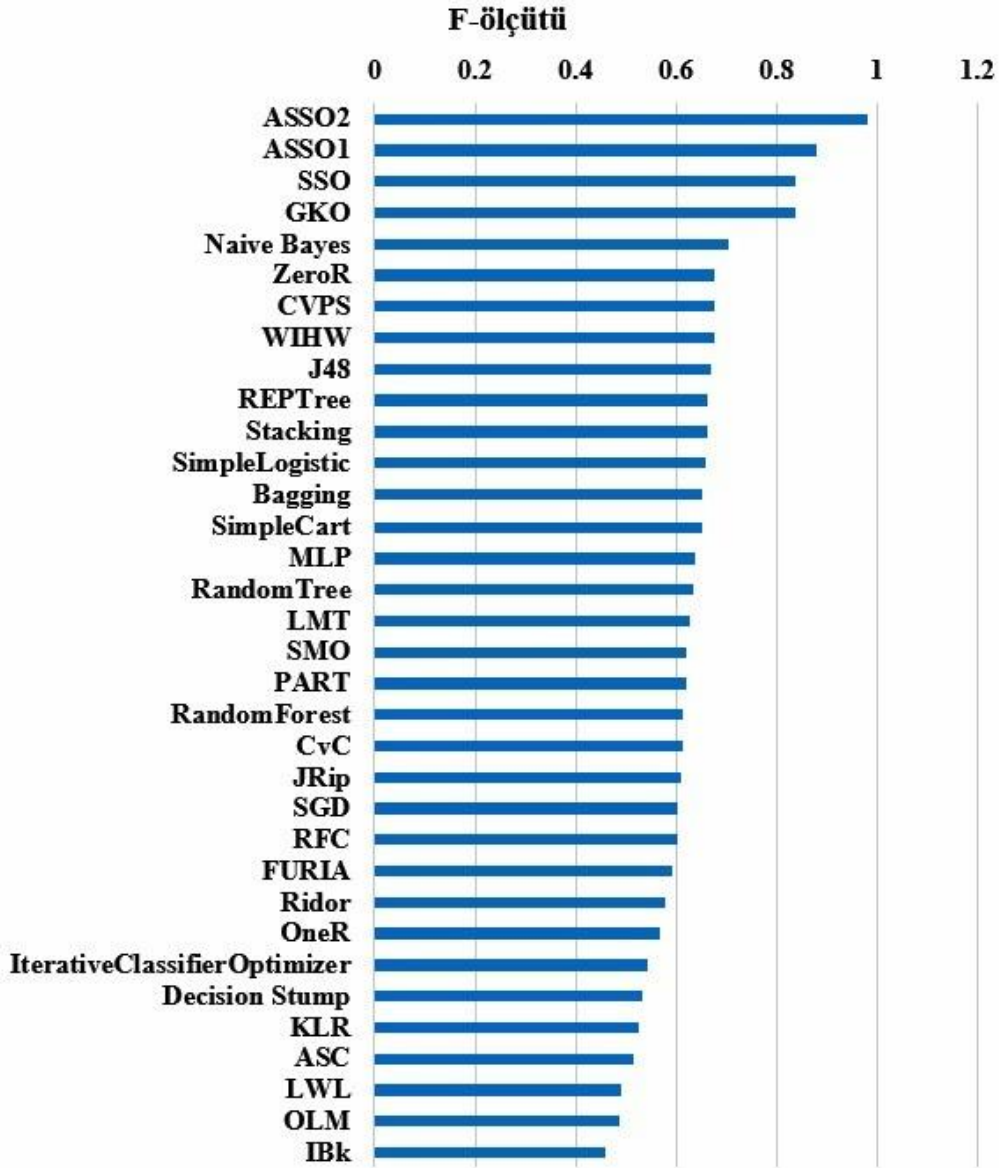
Şekil 6.3'te görüldüğü gibi, yöntemler kesinlik kriterine göre sıralandığında SSO ile en yüksek kesinlik değeri elde edilirken; GKO, ASSO1 ve ASSO2 optimizasyon algoritmaları da ikinci, üçüncü ve dördüncü yüksek değerleri vermiştir.

Duyarlılık ölçütü bakımından ZeroR, CVPS, WIHW değerleri 1.000 değeri ile en yüksek değeri vermiştir. ASSO2 algoritması ise 0.964 değeri ile dördüncü en iyi değeri vermiştir.



Şekil 6.4. BuzzFeed politik haberler veri kümesi için elde edilen duyarlılık değerleri

Şekil 6.5'te gösterildiği gibi en yüksek F-ölçütü değeri ASSO2 algoritması ile elde edilmiştir. ASSO2'den sonra en yüksek değerler ASSO1, SSO VE GKO algoritmaları ile elde edilmiştir.



Şekil 6.5. BuzzFeed politik haberler veri kümesi için elde edilen F-ölçütü değerleri

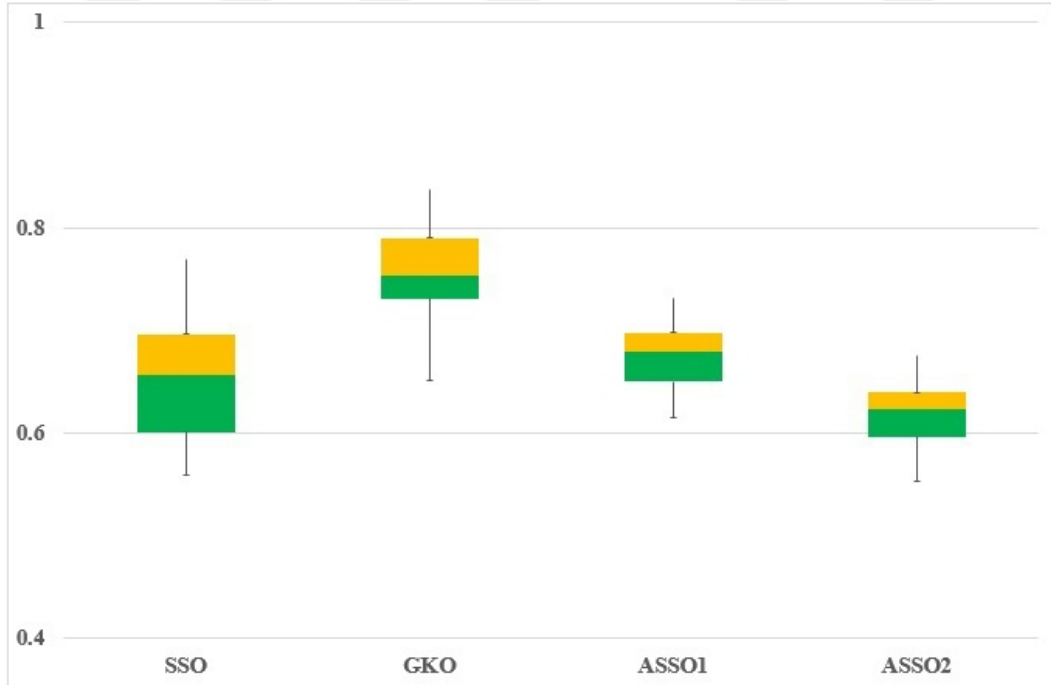
6.2. Rastgele Politik Haberler Veri Kümesinden Elde Edilen Sonuçlar

Öncelikle veri kümesi üzerinde, bir önceki bölümde söz edilen ön işlem adımları yapılmış ve doküman terim matrisi oluşturulmuştur. Ön işlemler sonrasında bu veri kümesi için 39 adet özellik elde edilmiştir. Bu nedenle optimizasyon algoritmaları rastgele politik haberler veri kümesi için çalıştırılırken problem boyutu 39 olarak belirlenmiştir. Tüm optimizasyon yöntemleri veri kümesi için 30 kez çalıştırıldıktan elde edilen sonuçların dört farklı performans değerlendirme ölçütü bakımından değerleri Tablo 6.6'da verilmiştir. Tablo 6.6'da görüldüğü gibi GKO algoritması doğruluk, duyarlılık ve F-ölçütü bakımından en iyi sonucu vermiştir. Kesinlik kriteri için GKO, ASSO1 ve ASSO2 algoritmaları 1.000 değeri ile SSO'dan daha iyi bir sonuç vermiştir.

Tablo 6.6. Optimizasyon algoritmalarının rastgele politik haberler veri kümesine uygulanması sonucunda elde edilen sonuçlar

	Doğruluk	Kesinlik	Duyarlılık	F-ölçütü
SSO	0.713	0.920	0.652	0.763
GKO	0.926	1.000	0.835	0.910
ASSO1	0.884	1.000	0.757	0.862
ASSO2	0.863	1.000	0.612	0.759

Rastgele politik haberler veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarının 30 kez çalıştırılmasıyla elde edilen uygunluk değerlerine göre çizilen kutu grafiği Şekil 6.6'da verilmiştir. Grafikteki değerler, elde edilen uygunluk değerlerinin dağılımını gösterir. Grafik incelendiğinde, GKO algoritmasından elde edilen alt çeyrek, üst çeyrek ve medyan değerleri SSO, ASSO1 ve ASSO2 algoritmalarından elde edilen değerlerden oldukça büyük olduğu görülmektedir. Rastgele sahte haber veri kümesi için ASSO2 algoritmasından elde edilen uygunluk fonksiyonu değerlerinin dağılımı diğer 3 algorithmadan çok daha düşüktür. Aynı zamanda bu değerler Friedman testi ile değerlendirilerek arasında anlamlı bir fark olup olmadığına bakılmıştır. Uygunluk fonksiyonu değerlerine göre Friedman testi ile elde edilen algoritma ortalama sıralamaları Tablo 6.7'de verilmiştir.



Şekil 6.6. SSO, GKO, ASSO1 ve ASSO2 algoritmaları ile rastgele politik veri kümesi için elde edilen sonuçların kutu grafiği

Tablo 6.7. Rastgele veri kümesi için Friedman testi ile elde edilen sıralamalar

Algoritma isimleri	Ortalama Sıralama
SSO	2.20
GKO	3.73
ASSO1	2.50
ASSO2	1.57

Tablo 6.8. Rastgele veri kümesi için Friedman testi ile elde edilen test istatistikleri

n	30
Ki-kare	44.680
df	3
p -değeri	0.000

Tablo 6.8’de görüldüğü rastgele veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarının elde ettiği sonuçlara göre elde edilen Friedman testi değerlerinde X_R^2 değeri 44.680’dir. Ki-kare tablosunda $\alpha = 0.05$ ve $df = 3$ değeri için gösterilen değer (X_{tablo}^2) 7.81473 olarak görülmektedir. $44.680 > 7.81473$ olduğu için H_0 hipotezi reddedilir. Yani, Friedman testine göre, rastgele politik haberler veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarından alınan sonuçlar arasındaki fark istatistiksel olarak anlamlıdır.

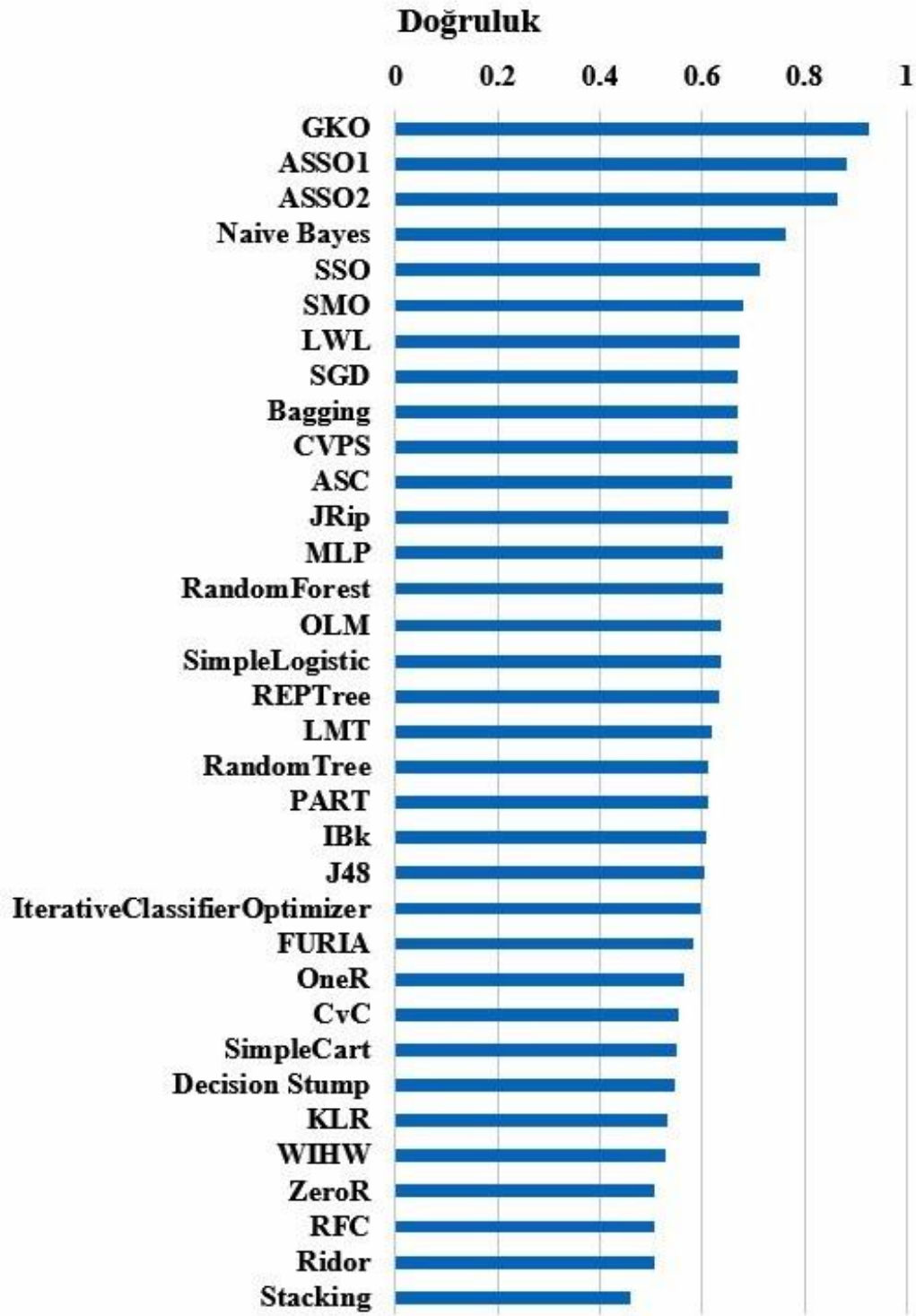
Rastgele politik haberler veri kümesi için 30 farklı yapay zekâ algoritması dört farklı performans değerlendirme kriterine göre elde edilen sonuçları Tablo 6.9’da verilmiştir. Tabloda gösterilen değerlere Naive Bayes yapay zekâ algoritmaları içerisinde en yüksek doğruluk değerini veren algoritmadır. Kesinlik kriteri için sonuçlar incelendiğinde Decision Stump algoritması ile en iyi değer elde edildiği görülmektedir. ZeroR, RFC ve Ridor algoritmaları 1.000 değeri ile 30 farklı yapay zekâ algoritması içinde duyarlılık bakımından en yüksek değeri vermişlerdir. Doğruluk değerindeki gibi Naive Bayes F-ölçütü içinde en iyi değeri veren algoritmadır.

Tablo 6.9’da verilen 30 farklı yapay zekâ algoritması ile elde edilen sonuçlar, SSO, GKO ve ilk kez bu çalışmada önerilen ASSO1 VE ASSO2 adaptif yöntemlerinin sonuçları ile kıyaslanmıştır. İçerisine optimizasyon yöntemlerinin de dahil edildiği 34 farklı yöntemin doğruluk, kesinlik, uyarlılık ve F-ölçütü için elde edilen sonuçları sırasıyla Şekil 6.7, Şekil 6.8, Şekil 6.9 ve Şekil 6.10’da grafiksel olarak gösterilmiştir. Şekillerdeki yöntemlerden elde edilen değerler büyükten küçüğe olacak şekilde sıralanmıştır.

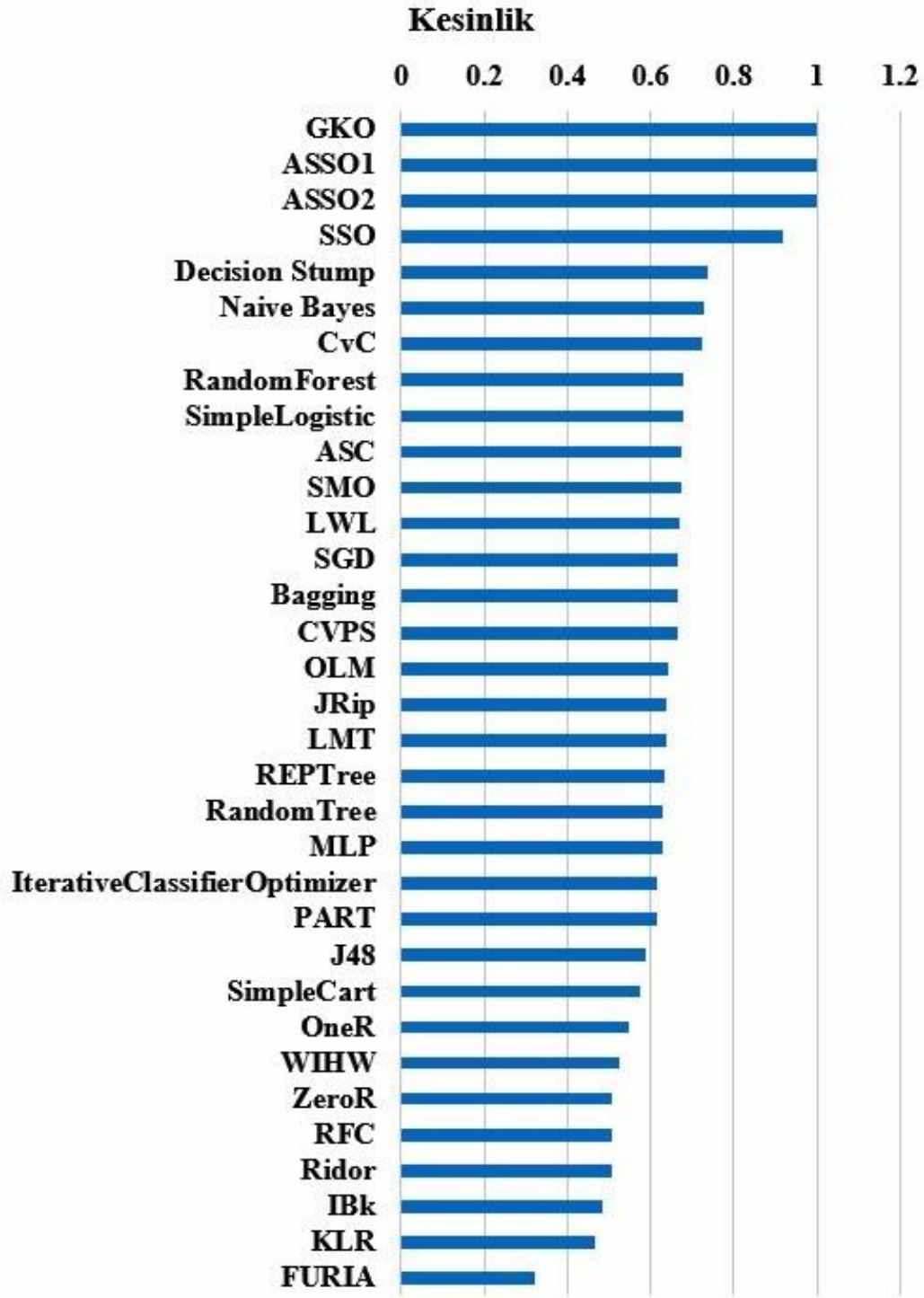
Tablo 6.9. Random veri kümesi için denetimli yapay zekâ algoritması ile elde edilen sonuçlar

	Doğruluk	Kesinlik	Duyarlılık	F-ölçütü
Naive Bayes	0.762	0.726	0.851	0.784
SMO	0.680	0.674	0.713	0.693
JRip	0.651	0.636	0.730	0.680
OneR	0.564	0.549	0.787	0.646
SGD	0.668	0.664	0.699	0.681
Bagging	0.668	0.664	0.699	0.681
Decision Stump	0.548	0.736	0.548	0.431
ZeroR	0.506	0.506	1.000	0.672
CVPS	0.668	0.664	0.699	0.681
RFC	0.506	0.506	1.000	0.672
LMT	0.619	0.636	0.583	0.608
LWL	0.672	0.669	0.699	0.684
CvC	0.555	0.723	0.556	0.447
WIHW	0.529	0.524	0.784	0.628
Ridor	0.506	0.506	1.000	0.672
MLP	0.642	0.628	0.721	0.671
OLM	0.637	0.643	0.639	0.641
SimpleCart	0.549	0.574	0.549	0.514
ASC	0.660	0.675	0.635	0.654
J48	0.605	0.587	0.746	0.657
REPTree	0.634	0.634	0.660	0.647
IBk	0.609	0.484	0.682	0.566
KLR	0.531	0.464	0.546	0.502
FURIA	0.584	0.322	0.585	0.571
RandomForest	0.640	0.677	0.637	0.657
RandomTree	0.613	0.629	0.614	0.611
SimpleLogistic	0.637	0.677	0.627	0.651
IterativeClassifierOptimizer	0.597	0.616	0.598	0.594
PART	0.610	0.615	0.608	0.609
Stacking	0.460	?	0.460	?

Şekil 6.7’de görüldüğü GKO algoritması doğruluk kriterine göre en yüksek değeri vermiştir. ASSO1 ve ASSO2 algoritmaları da onu izlemektedir.



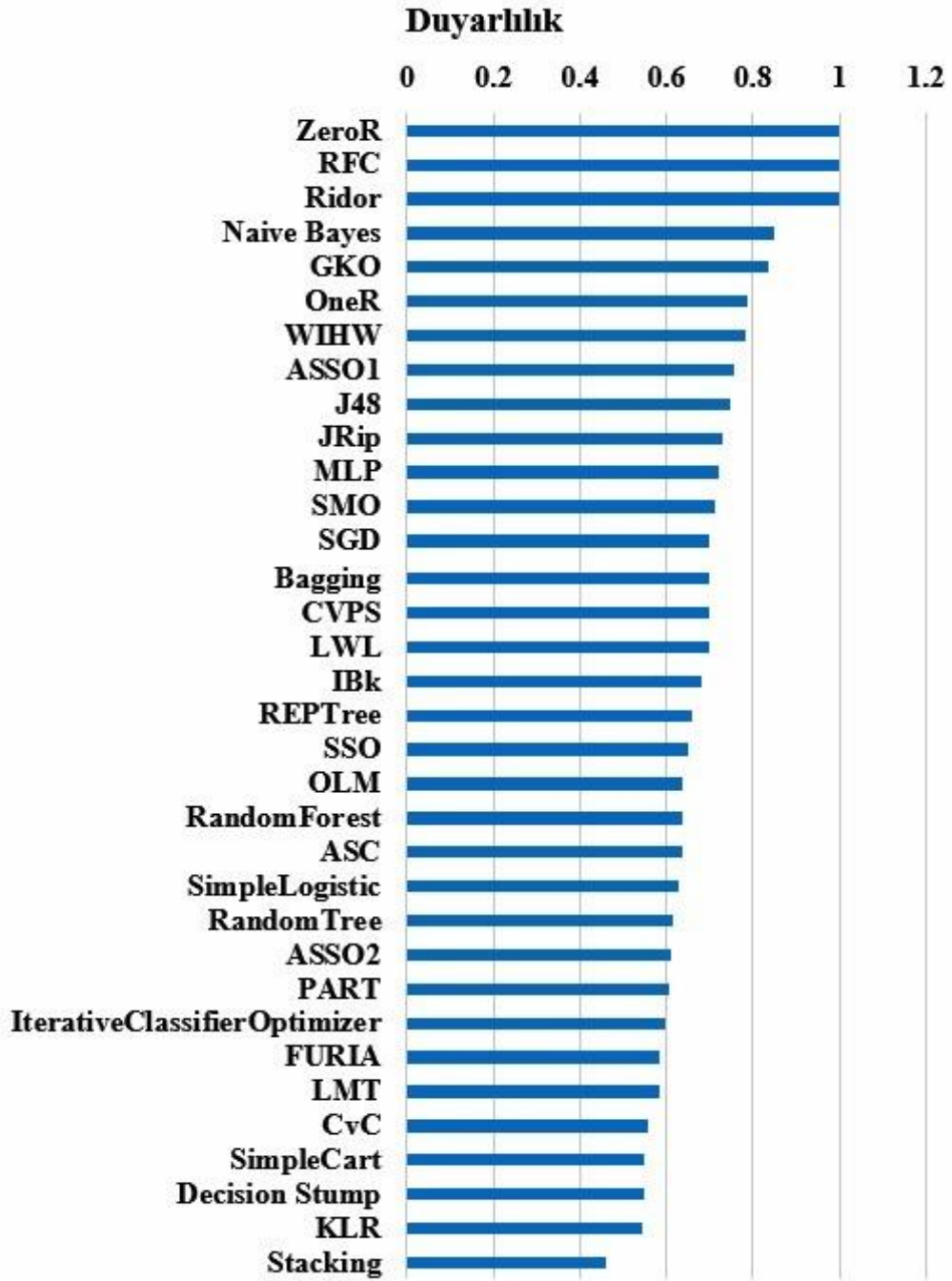
Şekil 6.7. Rastgele politik haberler veri kümesi için elde edilen doğruluk değerleri



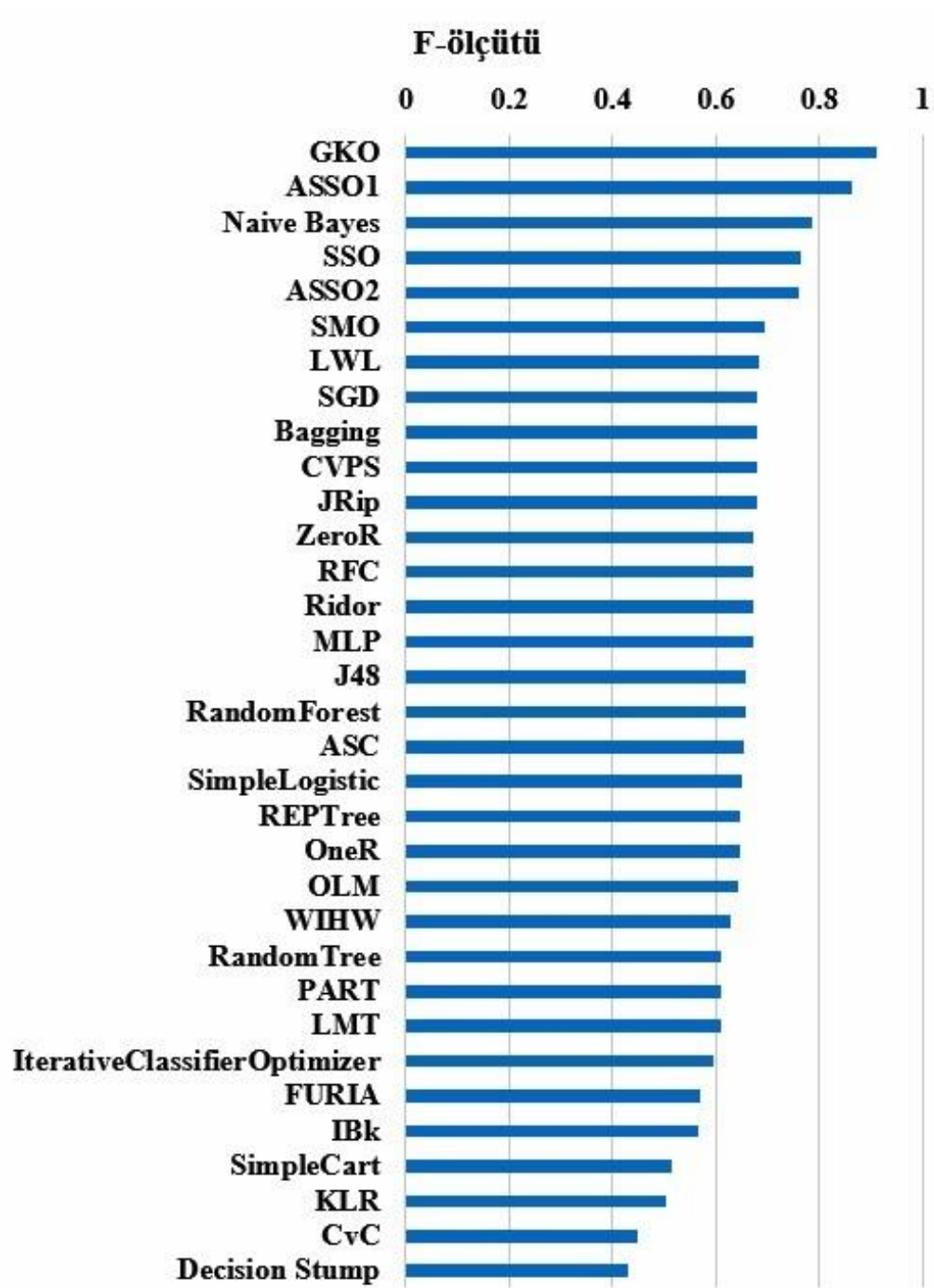
Şekil 6.8. Rastgele politik haberler veri kümesi için elde edilen kesinlik değerleri

Şekil 6.8’de görüldüğü gibi, kullanılan algoritmalar kesinlik kriterine göre sıralandığında GKO ile en yüksek kesinlik değeri elde edilirken; ASSO1, ASSO2 ve SSO algoritmaları da ikinci, üçüncü ve dördüncü yüksek değerleri vermiştir. Duyarlılık ölçütü bakımından ZeroR, RFC ve

Ridor algoritmaları 1.000 değeri ile en yüksek değeri vermiştir. GKO algoritması ise 0.835 değeri ile dördüncü en iyi değeri vermiştir.



Şekil 6.9. Rastgele politik haberler veri kümesi için elde edilen duyarlılık değerleri



Şekil 6.10. Rastgele politik haberler veri kümesi için elde edilen F-ölçütü değerleri

Şekil 6.10'da gösterildiği gibi en yüksek F-ölçütü değeri GKO algoritması ile elde edilmiştir. GKO'dan sonra en yüksek değer ASSO1 algoritması ile elde edilmiştir.

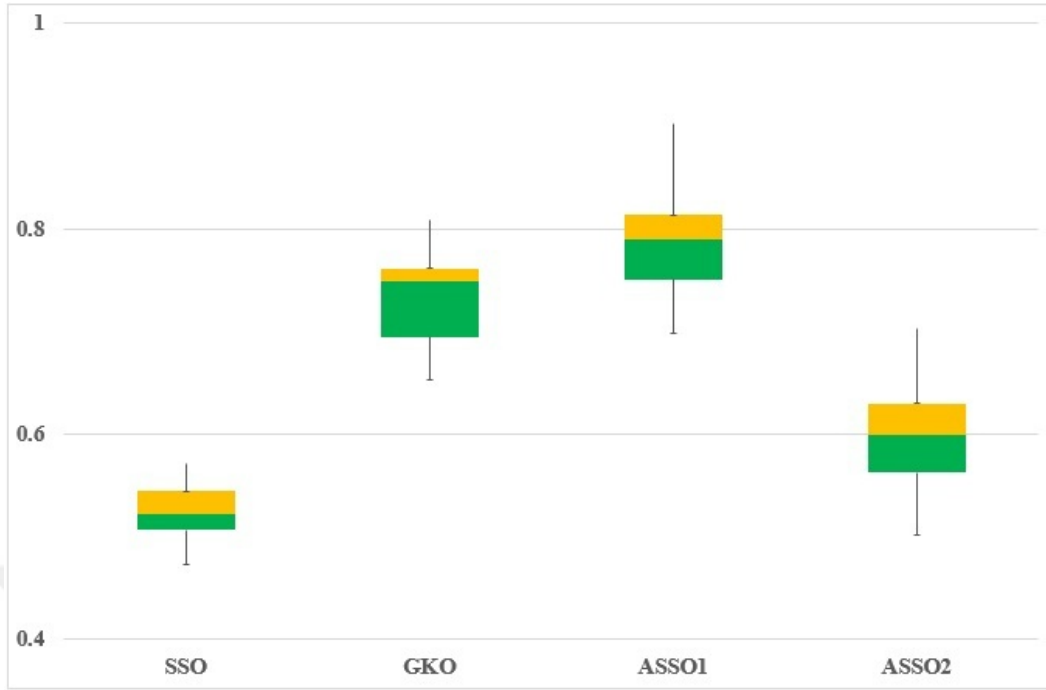
6.3. LIAR Veri Kümesinden Elde Edilen Sonuçlar

Öncelikle veri kümesi üzerinde, bir önceki bölümde söz edilen ön işlem adımları yapılmış ve doküman terim matrisi oluşturulmuştur. Ön işlemler sonrasında bu veri kümesi için 42 adet özellik elde edilmiştir. Bu nedenle optimizasyon algoritmaları rastgele politik haberler veri kümesi için çalıştırılırken problem boyutu 42 olarak belirlenmiştir. Tüm optimizasyon yöntemleri veri kümesi için 30 kez çalıştırdıktan elde edilen sonuçların dört farklı performans değerlendirme ölçütü bakımından değerleri Tablo 6.10'da verilmiştir. Tablo 6.10'da görüldüğü gibi GKO algoritması doğruluk, duyarlılık ve F-ölçütü bakımından en iyi sonucu vermiştir. Kesinlik kriteri için SSO ve ASSO2 algoritmaları 1.000 değeri ile en iyi sonucu vermiştir.

Tablo 6.10. Optimizasyon algoritmalarının LIAR veri kümesine uygulanması sonucunda elde edilen sonuçlar

	Doğruluk	Kesinlik	Duyarlılık	F-ölçütü
SSO	0.780	1.000	0.705	0.827
GKO	0.965	0.956	1.000	0.977
ASSO1	0.879	0.992	0.858	0.920
ASSO2	0.956	1.000	0.940	0.969

LIAR veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarının 30 kez çalıştırılmasıyla elde edilen uygunluk değerlerine göre çizilen kutu grafiği Şekil 6.11'de verilmiştir. Grafik incelendiğinde, SSO algoritmasından elde edilen alt çeyrek, üst çeyrek ve medyan değerleri SSO, ASSO1 ve ASSO2 algoritmalarından elde edilen değerlerden oldukça küçük olduğu görülmektedir. 30 kez çalıştırılma sonucunda ASSO1 algoritması ile elde edilen uygunluk fonksiyonu değerlerini diğer 3 algoritmaya göre yüksektir. Aynı zamanda bu değerler Friedman testi ile değerlendirilerek arasında anlamlı bir fark olup olmadığına bakılmıştır. Uygunluk fonksiyonu değerlerine göre Friedman testi ile elde edilen algoritma ortalama sıralamaları Tablo 6.11'de verilmiştir.



Şekil 6.11. SSO, GKO, ASSO1 ve ASSO2 algoritmaları ile LIAR veri kümesi için elde edilen sonuçların kutu grafiği

Tablo 6.11. LIAR veri kümesi için Friedman testi ile elde edilen sıralamalar

Algoritma isimleri	Ortalama Sıralama
SSO	1.13
GKO	3.17
ASSO1	3.80
ASSO2	1.90

Tablo 6.12. LIAR veri kümesi için Friedman testi ile elde edilen test istatistikleri

n	30
Ki-kare	78.520
df	3
p -değeri	0.000

Tablo 6.12’de görüldüğü rastgele veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarının elde ettiği sonuçlara göre elde edilen Friedman testi değerlerinde X_R^2 değeri 78.520’dir. Ki-kare tablosunda $\alpha = 0.05$ ve $df = 3$ değeri için gösterilen değer (X_{tablo}^2) 7.81473 olarak görülmektedir. $78.520 > 7.81473$ olduğu için H_0 hipotezi reddedilir. Yani, Friedman testine göre, LIAR veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarından alınan sonuçlar arasındaki fark istatistiksel olarak anlamlıdır.

LIAR veri kümesi için 30 farklı yapay zekâ algoritması dört farklı performans değerlendirme kriterine göre elde edilen sonuçları Tablo 6.13’de verilmiştir. Tablo 6.13’de gösterilen değerlere

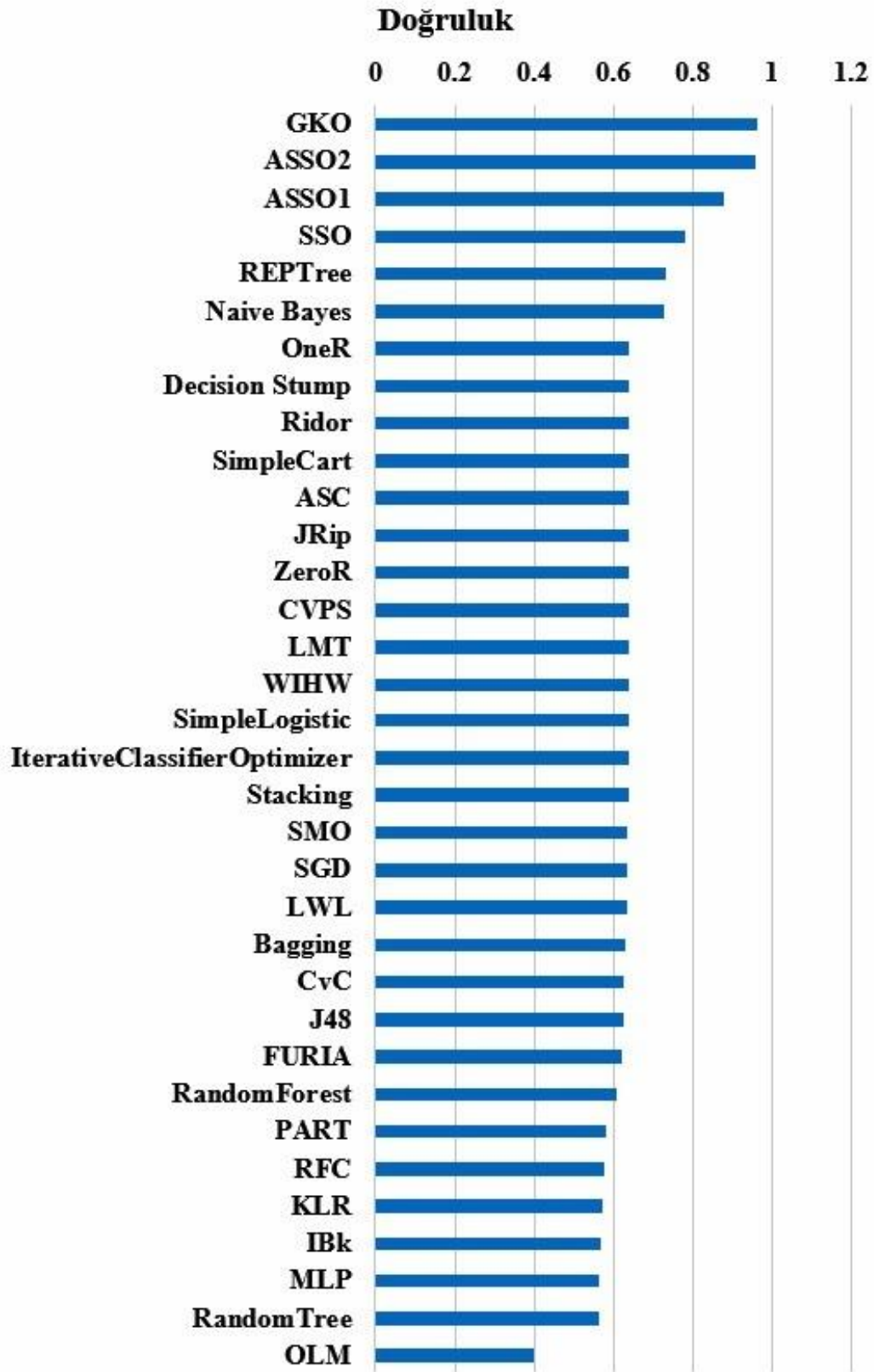
REPTree yapay zekâ algoritmaları içerisinde en yüksek doğruluk değerini veren algoritmadır. Kesinlik kriteri için sonuçlar incelendiğinde Naive Bayes algoritması ile en iyi değerin elde edildiği görülmektedir. ZeroR, CVPS, WIHW, REPTree ve Stacking algoritmaları 1.000 değeri ile 30 farklı yapay zekâ algoritması içinde duyarlılık bakımından en yüksek değeri vermişlerdir. Doğruluk değerindeki gibi REPTree F-ölçütü içinde en iyi değeri veren algoritmadır.

Tablo 6.13'te verilen 30 farklı yapay zekâ algoritması ile elde edilen sonuçlar, SSO, GKO ve ilk kez bu çalışmada önerilen ASSO1 VE ASSO2 adaptif yöntemlerinin sonuçları ile kıyaslanmıştır. İçerisine optimizasyon yöntemlerinin de dahil edildiği 34 farklı yöntemin doğruluk, kesinlik, uyarlılık ve F-ölçütü için elde edilen sonuçları sırasıyla Şekil 6.12, Şekil 6.13, Şekil 6.14 ve Şekil 6.15'te grafiksel olarak gösterilmiştir. Şekillerdeki yöntemlerden elde edilen değerler büyükten küçüğe olacak şekilde sıralanmıştır.

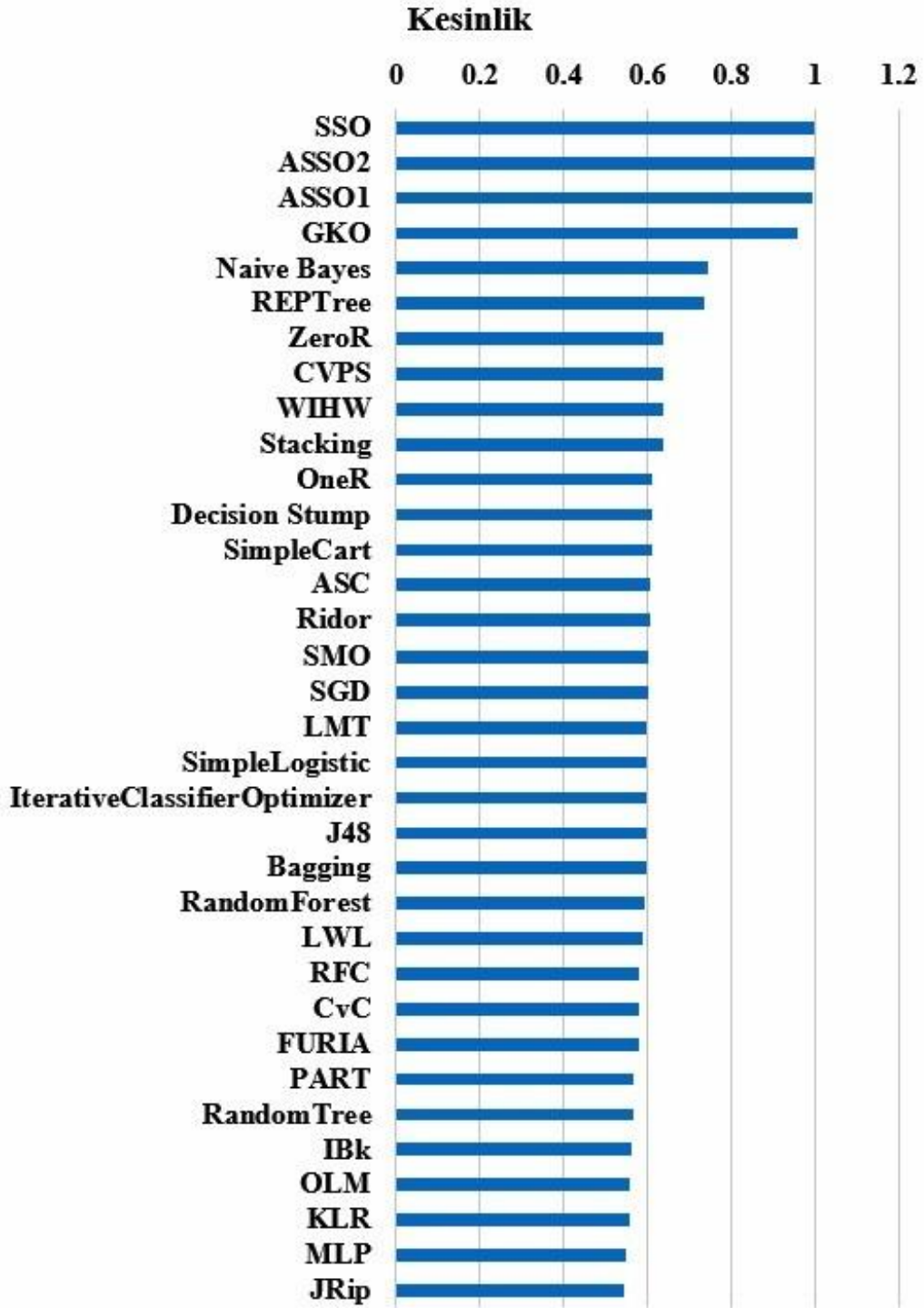
Tablo 6.13. LIAR veri kümesi için denetimli yapay zekâ algoritması ile elde edilen sonuçlar

	Doğruluk	Kesinlik	Duyarlılık	F-ölçütü
Naive Bayes	0.726	0.746	0.910	0.820
SMO	0.634	0.600	0.635	0.589
JRip	0.638	0.544	0.638	0.592
OneR	0.641	0.609	0.642	0.532
SGD	0.634	0.600	0.635	0.589
Bagging	0.631	0.595	0.631	0.587
Decision Stump	0.641	0.609	0.642	0.532
ZeroR	0.638	0.638	1.000	0.779
CVPS	0.638	0.638	1.000	0.779
RFC	0.576	0.579	0.577	0.578
LMT	0.638	0.598	0.638	0.568
LWL	0.634	0.590	0.635	0.562
CvC	0.627	0.579	0.628	0.561
WIHW	0.638	0.638	1.000	0.779
Ridor	0.641	0.605	0.642	0.555
MLP	0.563	0.549	0.563	0.554
OLM	0.399	0.558	0.399	0.315
SimpleCart	0.641	0.609	0.642	0.532
ASC	0.641	0.606	0.642	0.573
J48	0.624	0.597	0.625	0.599
REPTree	0.733	0.734	1.000	0.846
IBk	0.569	0.561	0.570	0.565
KLR	0.573	0.557	0.573	0.563
FURIA	0.621	0.579	0.621	0.574
RandomForest	0.610	0.591	0.611	0.596
RandomTree	0.563	0.567	0.563	0.565
SimpleLogistic	0.638	0.598	0.638	0.568
IterativeClassifierOptimizer	0.638	0.598	0.638	0.568
PART	0.583	0.568	0.584	0.574
Stacking	0.638	0.638	1.000	0.779

Şekil 6.12'de görüldüğü GKO algoritması doğruluk kriterine göre en yüksek değeri vermiştir. ASSO2, ASSO1 ve SSO algoritmaları da onu izlemektedir.

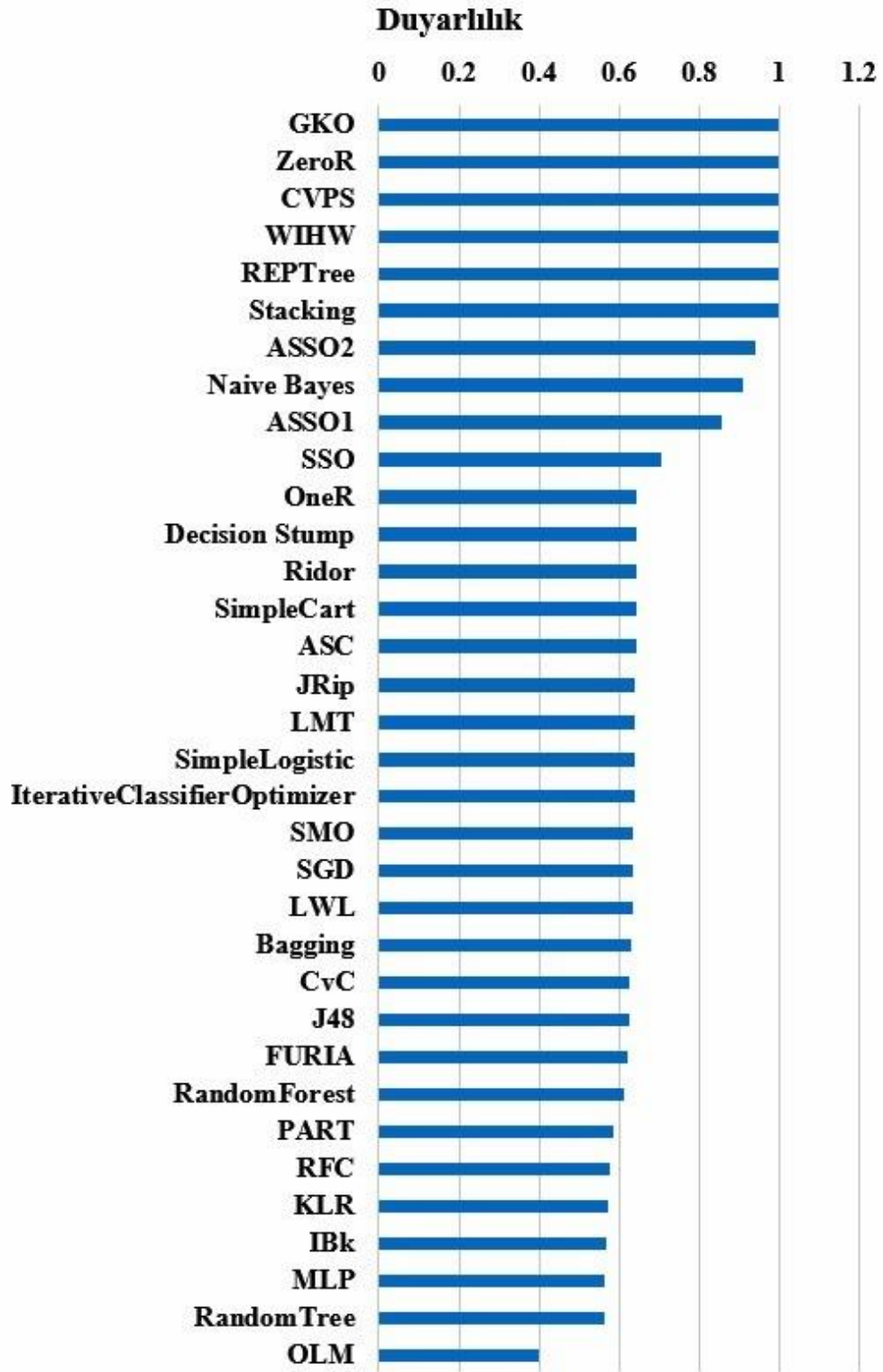


Şekil 6.12. LIAR veri kümesi için elde edilen doğruluk değerleri



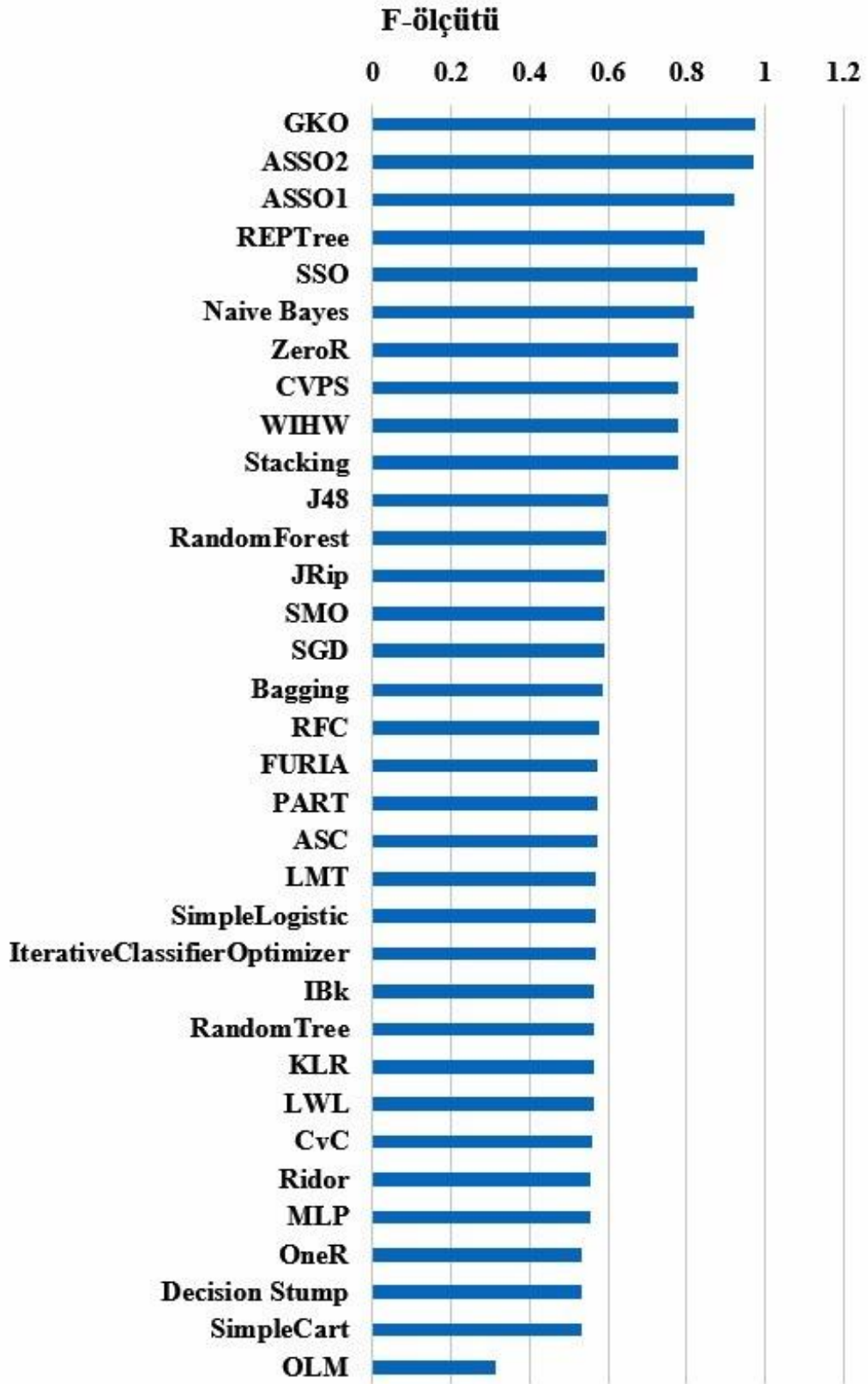
Şekil 6.13. LIAR veri kümesi için elde edilen kesinlik değerleri

Şekil 6.13'te görüldüğü gibi, kullanılan algoritmalar kesinlik kriterine göre sıralandığında SSO ve ASSO2 algoritmaları 1.000 değeri ile en yüksek kesinlik değerlerini verirken, ASSO1 ve GKO algoritmaları da bu iki yöntemden sonra en iyi değerleri vermiştir.



Şekil 6.14. LIAR veri kümesi için elde edilen duyarlılık değerleri

Şekil 6.14'te görüldüğü gibi duyarlılık ölçütü bakımından GKO algoritması, 1.000 değeri ile en yüksek değeri vermiştir.



Şekil 6.15. LIAR veri kümesi için elde edilen F-ölçütü değerleri

Şekil 6.15'te gösterildiği gibi en yüksek F-ölçütü değeri GKO algoritması ile elde edilmiştir. GKO'dan sonra en yüksek değer ASSO2 algoritması ile elde edilmiştir.

6.4. ISOT Veri Kümesinden Elde Edilen Sonuçlar

Öncelikle veri kümesi üzerinde, bir önceki bölümde söz edilen ön işlem adımları yapılmış ve doküman terim matrisi oluşturulmuştur. Ön işlemler sonrasında bu veri kümesi için 44 adet özellik elde edilmiştir. Bu nedenle optimizasyon algoritmaları rastgele politik haberler veri kümesi için çalıştırılırken problem boyutu 44 olarak belirlenmiştir. Tüm optimizasyon yöntemleri veri kümesi için 30 kez çalıştırıldıktan elde edilen sonuçların dört farklı performans değerlendirme ölçütü bakımından değerleri Tablo 6.14'te verilmiştir. Tablo 6.14'te görüldüğü gibi GKO algoritması doğruluk, duyarlılık ve F-ölçütü bakımından en iyi sonucu vermiştir. Kesinlik kriteri için SSO, ASSO1 ve ASSO2 algoritmaları 1.000 değeri ile en iyi sonucu vermiştir.

Tablo 6.14. Optimizasyon algoritmalarının ISOT veri kümesine uygulanması sonucunda elde edilen sonuçlar

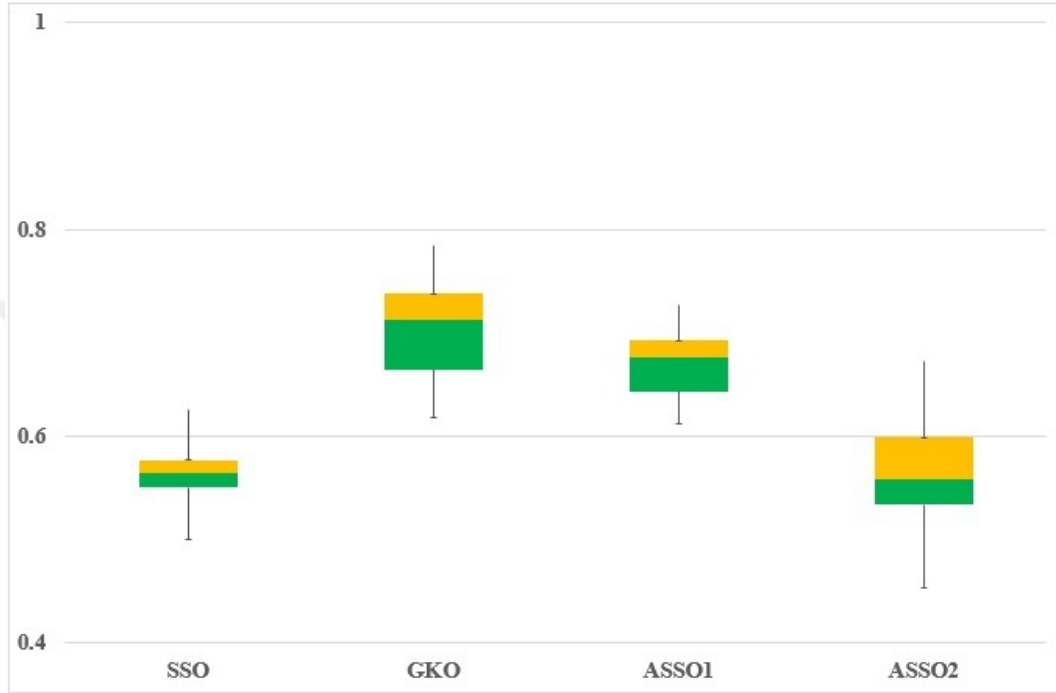
	Doğruluk	Kesinlik	Duyarlılık	F-ölçütü
SSO	0.859	1.000	0.791	0.883
GKO	0.995	0.994	1.000	0.997
ASSO1	0.972	1.000	0.972	0.986
ASSO2	0.942	1.000	0.893	0.943

ISOT veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarının 30 kez çalıştırılmasıyla elde edilen uygunluk değerlerine göre çizilen kutu grafiği Şekil 6.16'da verilmiştir. Aynı zamanda bu değerler Friedman testi ile değerlendirilerek arasında anlamlı bir fark olup olmadığına bakılmıştır. Grafik incelendiğinde, SSO ve ASSO2 algoritmasından elde edilen alt çeyrek ve medyan değerlerinin çok yakın olduğu görülmektedir. Ayrıca, GKO algoritmalarından elde edilen alt çeyrek, üst çeyrek ve medyan değerlerinin diğer 3 yöntemle kıyaslandığında yüksek olduğu gözlemlenmiştir. Uygunluk fonksiyonu değerlerine göre Friedman testi ile elde edilen algoritma ortalama sıralamaları Tablo 6.15'te verilmiştir.

Tablo 6.16'da görüldüğü rastgele veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarının elde ettiği sonuçlara göre elde edilen Friedman testi değerlerinde X_R^2 değeri 68.320'dir. Ki-kare tablosunda $\alpha = 0.05$ ve $df = 3$ değeri için gösterilen değer (X_{tablo}^2) 7.81473 olarak görülmektedir. $68.320 > 7.81473$ olduğu için H_0 hipotezi reddedilir. Yani, Friedman testine göre, ISOT veri kümesi için SSO, GKO, ASSO1 ve ASSO2 algoritmalarından alınan sonuçlar arasındaki fark istatistiksel olarak anlamlıdır.

ISOT veri kümesi için 30 farklı yapay zekâ algoritması dört farklı performans değerlendirme kriterine göre elde edilen sonuçları Tablo 6.17'de verilmiştir. Tablo 6.17'de gösterilen değerlere

REPTree yapay zekâ algoritmaları içerisinde en yüksek doğruluk, kesinlik ve F-ölçütü değerlerini veren algoritmadır. ZeroR, RFC ve Ridor algoritmaları 1.000 değeri ile 30 farklı yapay zekâ algoritması içinde duyarlılık bakımından en yüksek değeri vermişlerdir. Doğruluk değerindeki gibi REPTree F-ölçütü cinsinden en iyi değeri veren algoritmadır.



Şekil 6.16. SSO, GKO, ASSO1 ve ASSO2 algoritmaları ile ISOT veri kümesi için elde edilen sonuçların kutu grafiği

Tablo 6.15. ISOT veri kümesi için Friedman testi ile elde edilen sıralamalar

Algoritma isimleri	Ortalama Sıralama
SSO	1.57
GKO	3.63
ASSO1	3.30
ASSO2	1.50

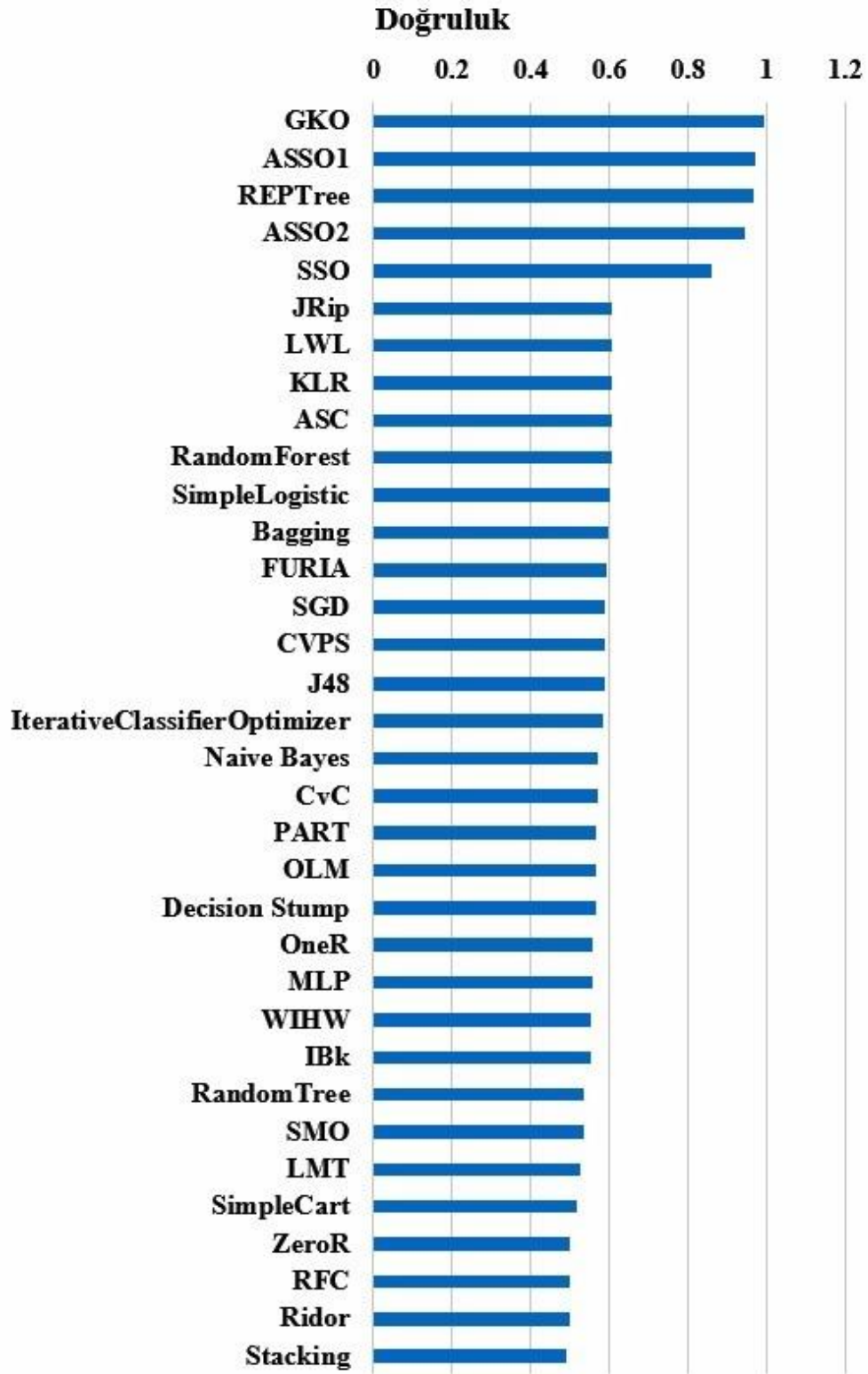
Tablo 6.16. ISOT veri kümesi için Friedman testi ile elde edilen test istatistikleri

n	30
Ki-kare	68.320
df	3
p -değeri	0.000

Tablo 6.17. ISOT veri kümesi için denetimli yapay zekâ algoritması ile elde edilen sonuçlar

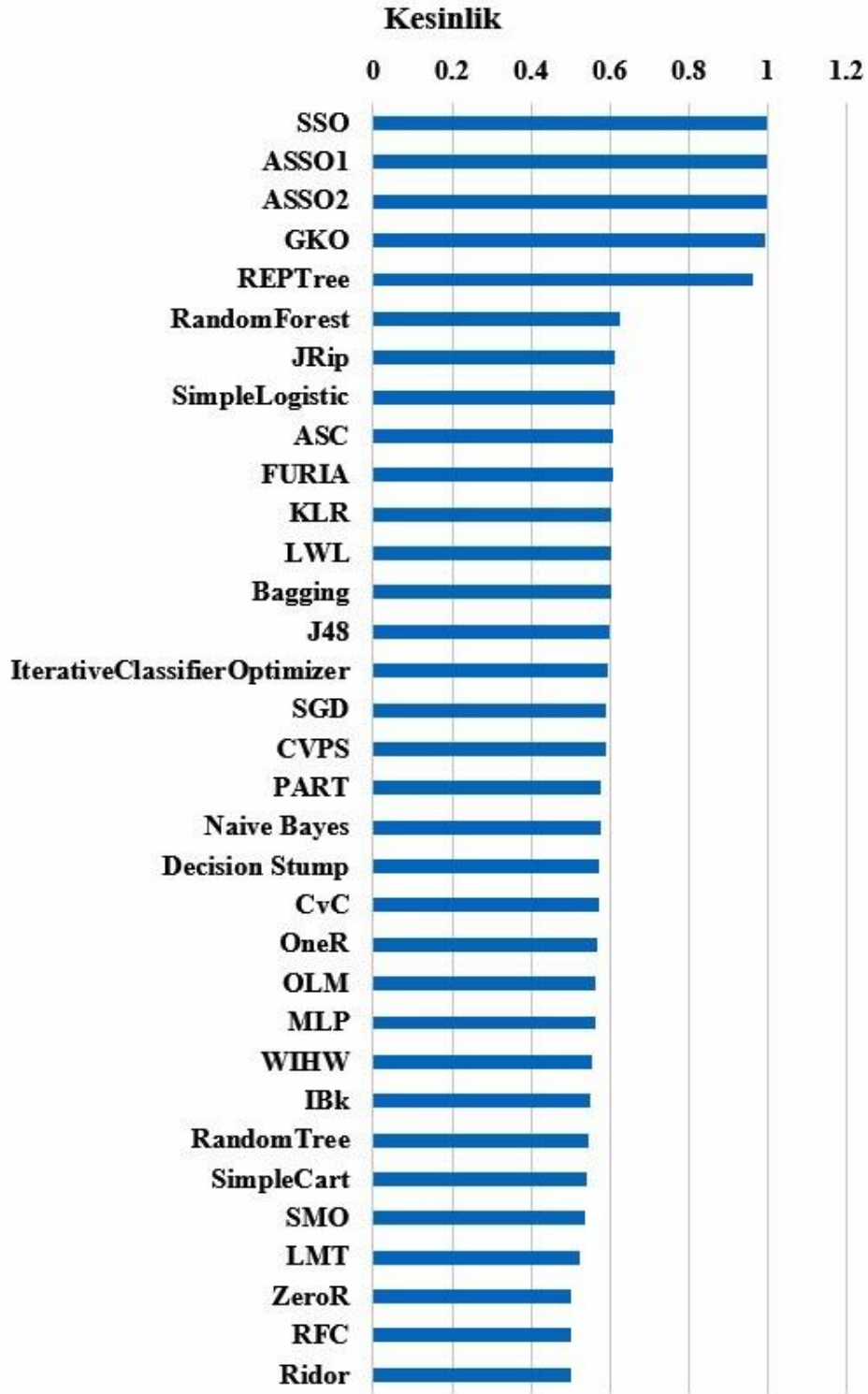
	Doğruluk	Kesinlik	Duyarlılık	F-ölçütü
Naive Bayes	0.571	0.576	0.599	0.587
SMO	0.534	0.536	0.489	0.512
JRip	0.607	0.611	0.588	0.599
OneR	0.559	0.567	0.560	0.547
SGD	0.589	0.590	0.583	0.586
Bagging	0.598	0.603	0.576	0.589
Decision Stump	0.564	0.574	0.564	0.549
ZeroR	0.501	0.501	1.000	0.667
CVPS	0.589	0.590	0.583	0.586
RFC	0.501	0.501	1.000	0.667
LMT	0.526	0.525	0.534	0.530
LWL	0.607	0.604	0.616	0.610
CvC	0.570	0.573	0.570	0.566
WIHW	0.553	0.556	0.526	0.541
Ridor	0.501	0.501	1.000	0.667
MLP	0.557	0.563	0.558	0.549
OLM	0.565	0.565	0.571	0.568
SimpleCart	0.516	0.540	0.516	0.430
ASC	0.604	0.607	0.586	0.597
J48	0.588	0.598	0.534	0.564
REPTree	0.968	0.963	0.973	0.968
IBk	0.551	0.551	0.551	0.550
KLR	0.606	0.605	0.614	0.609
FURIA	0.591	0.607	0.555	0.580
RandomForest	0.604	0.625	0.557	0.589
RandomTree	0.536	0.544	0.534	0.536
SimpleLogistic	0.601	0.611	0.592	0.601
IterativeClassifierOptimizer	0.585	0.596	0.592	0.584
PART	0.567	0.579	0.546	0.562
Stacking	0.491	?	0.492	?

Tablo 6.17’de verilen 30 farklı yapay zekâ algoritması ile elde edilen sonuçlar, SSO, GKO ve ilk kez bu çalışmada önerilen ASSO1 VE ASSO2 adaptif yöntemlerinin sonuçları ile kıyaslanmıştır. İçerisine optimizasyon yöntemlerinin de dahil edildiği 34 farklı yöntemin doğruluk, kesinlik, uyarlılık ve F-ölçütü için elde edilen sonuçları sırasıyla Şekil 6.17, Şekil 6.18, Şekil 6.19 ve Şekil 6.20’de grafiksel olarak gösterilmiştir. Şekillerdeki yöntemlerden elde edilen değerler büyükten küçüğe olacak şekilde sıralanmıştır.



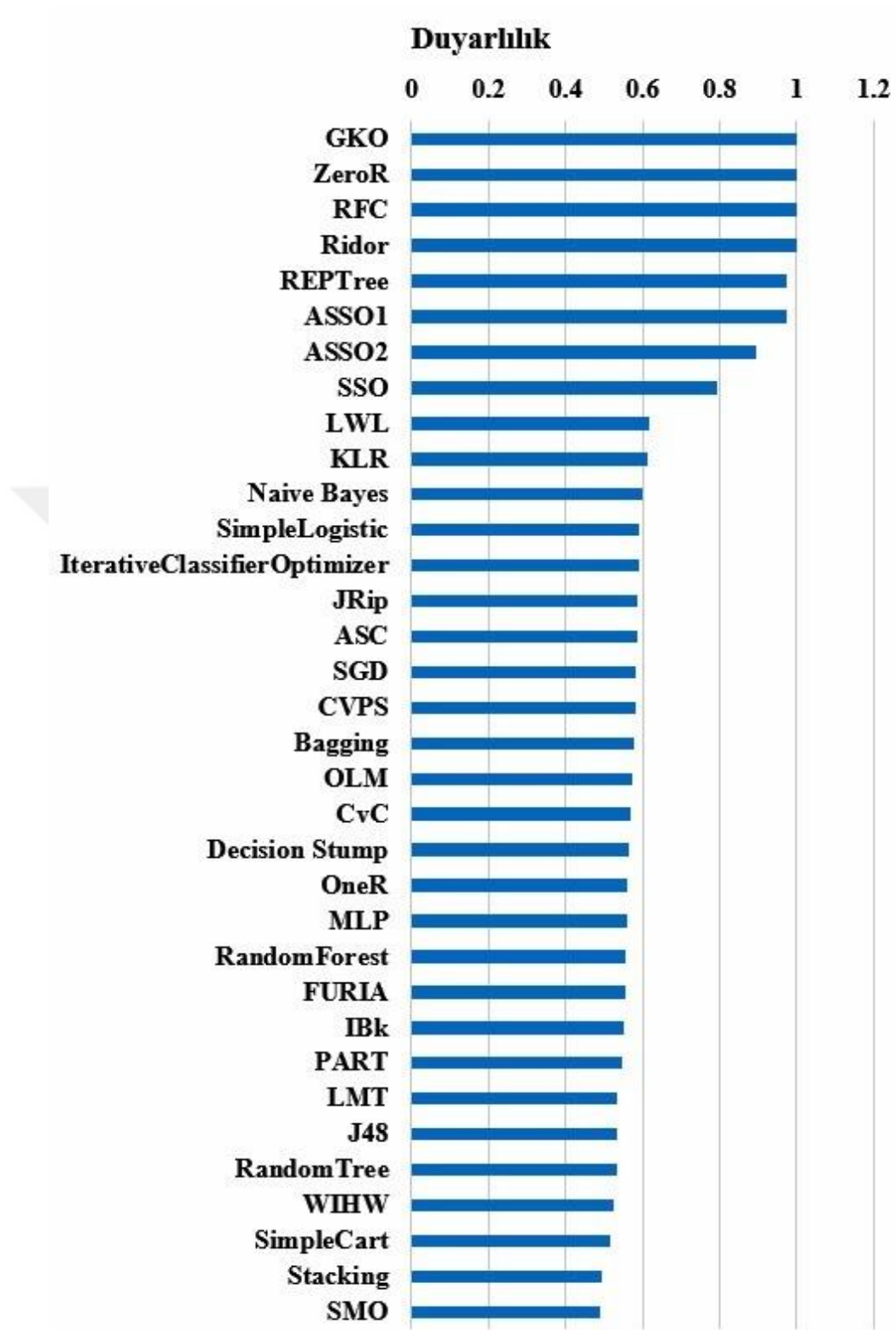
Şekil 6.17. ISOT veri kümesi için elde edilen doğruluk değerleri

Şekil 6.17’de görüldüğü GKO algoritması doğruluk kriterine göre en yüksek değeri vermiştir. ASSO1, ASSO2 ve SSO algoritmaları da onu izlemektedir.



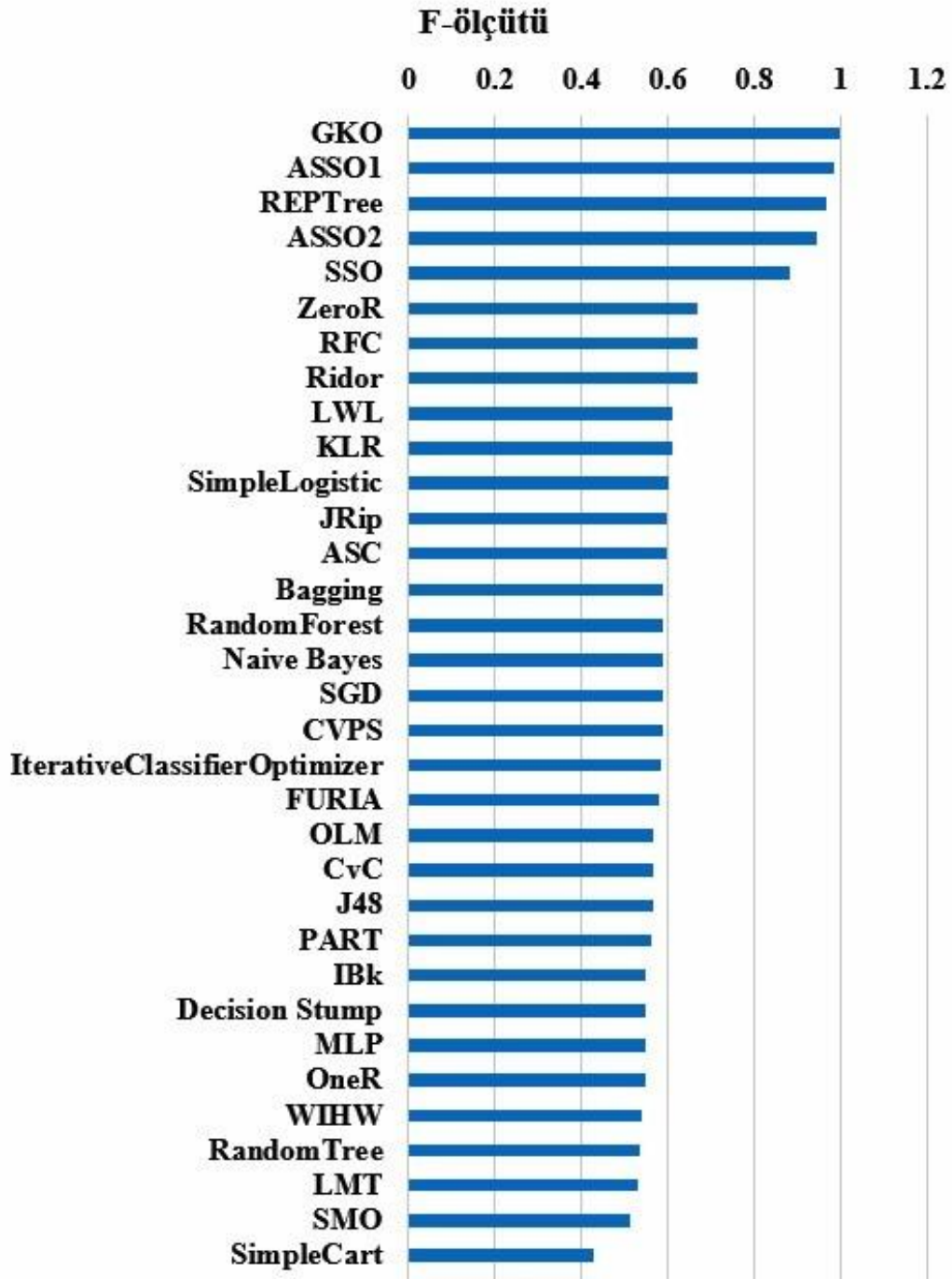
Şekil 6.18. ISOT veri kümesi için elde edilen kesinlik değerleri

Şekil 6.18’de görüldüğü gibi, kullanılan algoritmalar kesinlik kriterine göre sıralandığında SSO, ASSO1 ve ASSO2 algoritmaları 1.000 değeri ile en yüksek kesinlik değerlerini verirken, GKO algoritması da bu üç yöntemden sonra en iyi değerleri vermiştir.



Şekil 6.19. ISOT veri kümesi için elde edilen duyarlılık değerleri

Şekil 6.19’da görüldüğü gibi duyarlılık ölçütü bakımından GKO algoritması, 1.000 değeri ile en yüksek değeri vermiştir.



Şekil 6.20. ISOT veri kümesi için elde edilen F-ölçütü değerleri

Şekil 6.20’de gösterildiği gibi en yüksek F-ölçütü değeri GKO algoritması ile elde edilmiştir. GKO’dan sonra en yüksek değer ASSO1 algoritması ile elde edilmiştir.

7. SONUÇLAR

Son yıllarda, internetin hızlı gelişimi ve yayılımı ile birlikte sosyal, kültürel ve ekonomik hayatta gözle görülür birçok değişiklik olmuştur. Meydana gelen bu değişiklikler bireylerin toplumsal yaşamındaki iletişim kurma alışkanlıklarını da değiştirmiştir. Değişen iletişim alışkanlıklarıyla beraber sosyal medya kavramı ortaya çıkmıştır. İletişim kapasitesini arttırmanın yanında maliyetlerini azaltan ve buna ek olarak nitelik olarak sürekli gelişen sosyal medya her geçen gün hızla büyümekte ve internetin en popüler uygulamaları arasında yerini almaktadır. Sosyal medya insanları birbirine bağlayan, iletişim kurup sosyalleşmelerini sağlayan bir ortam olmasının yanı sıra bilgiye ulaşmak için en çok kullanılan ortam haline gelmiştir. Artık bireyler, siyasi, sosyal, kültürel, ekonomik ve sanatsal birçok bilgiye sosyal medya platformlarından erişebilmektedir. Sosyal medyanın insan hayatındaki artan kullanımı ile birlikte insanların haberleri tüketmek için kullandıkları platformlar da değişmiştir. Artık birçok insan gazete ve televizyon gibi geleneksel haber kaynakları yerine hızlı, düşük maliyetli ve kolay erişilebilir oluşu nedeniyle sosyal medya sitelerini tercih etmektedir. Ayrıca tartışma ve yorum yapma kolaylığı sağlaması ve sosyal medyadaki diğer kullanıcılarla kolaylıkla içerik paylaşabilme imkânı sunması sosyal medyanın haberlere erişim için tercih edilme nedenlerinden biridir.

Sosyal medyanın kullanıcılara sağladığı birçok kolaylığın yanında, sosyal medyadaki haberlerin kalitesi geleneksel haber kaynakları ile kıyaslandığında oldukça düşüktür. Çevrimiçi olarak haberler sunmak ucuz ve sosyal medya aracılığı ile yaymak daha hızlı ve kolay olduğu için, sosyal medyada finansal, politik veya şahsi çıkarlar sağlamak amacıyla kasıtlı olarak yanlış bilgiler içeren büyük miktarda sahte haber üretilmektedir. Sosyal medyadaki sahte haberler, bireylere ve toplumlara büyük ölçüde zarar vermektedir. Sahte haberler, insanların kafalarını karıştırarak onları yanlış bilgilere ve düşüncelere yöneltmektedir. Ayrıca insanların tüm haberlere kuşku ile yaklaşmasına neden olarak yanlış ile doğruyu ayırt etme yeteneklerini engellemektedir. Sosyal medyadaki sahte haberlerin kullanıcılar üzerindeki olumsuz etkilerinden korunmak için sahte haberlerin tespit edilmesi önemli bir problem haline gelmiştir.

Bu tez kapsamında popüler ve yeni bir ÇSM problemi olan Sahte Haber Tespiti problemi üzerinde durulmuştur. Sahte Haber Tespiti problemi ilk kez bu çalışmada bir optimizasyon problemi olarak ele alınmıştır. SSO ve GKO optimizasyon algoritmaları Sahte Haber Tespiti problemi için uyarlanarak, performansları test edilmiştir. Ayrıca daha iyi sonuçlar elde edebilmek için doğrusal olmayan şekilde azalan atalet ağırlıklı SSO ve osilasyon (salınım) atalet ağırlıklı SSO adlı iki yeni adaptif SSO algoritması ilk kez bu tez çalışması kapsamında önerilerek sahte haber tespiti problemi için modellenmiştir. Modellenen dört farklı optimizasyon algoritmalarının sonuçlarını test etmek amacıyla BuzzFeed politik haber, Rastgele politik haber, LIAR ve ISOT olmak üzere gerçek dünya haberlerini içeren dört farklı veri kümesi kullanılmıştır. İlk olarak veri

kümelerindeki metinler ön işlemlerden geçirilmiş ve doküman terim matrisi oluşturulmuştur. Daha sonra SSO, GKO, ASSO1 ve ASSO2 optimizasyon algoritmalarının bu probleme uyarlanabilmesi için esnek bir uygunluk fonksiyonu belirlenerek veri kümeleri üzerinde sahte haber tespiti yapılmıştır. Doğruluk, kesinlik, duyarlılık ve F-ölçütü olmak üzere dört tane performans değerlendirme kriteri için en iyi değeri elde etmeyi hedefleyen bir uygunluk fonksiyonu belirlenmiş ve bu uygunluk fonksiyonu ilk kez bu çalışmada sahte haber tespiti problemi için kullanılmıştır. Uygunluk fonksiyonu farklı amaçlar eklenerek kolaylıkla değiştirilebilecek şekilde olduğundan esnek bir fonksiyondur. Kullanılan optimizasyon algoritmaları her bir veri kümesi için 30 kez çalıştırılmış ve elde edilen uygunluk fonksiyonu sonucu değerlerini istatistiksel olarak yorumlamak için Friedman testi kullanılarak yöntemlerden elde edilen sonuçlar arasında anlamlı bir fark olup olmadığı incelenmiştir. Optimizasyon algoritmalarına ek olarak 30 tane denetimli yapay zekâ algoritması da bu problem için çalıştırılmış ve elde edilen tüm değerler doğruluk, kesinlik, duyarlılık ve F-ölçütü ölçütlerine göre karşılaştırılmıştır.

Sahte haber tespiti problemi için optimizasyon algoritmaları kullanılarak ilk kez uygulanan bu yöntemden elde edilen sonuçlar oldukça umut vericidir. Bu yönü ile gelecekte, sahte haber tespiti problemi için farklı metasezgisel optimizasyon algoritmaları uyarlanabilir. Algoritmaların hibrit versiyonları geliştirilerek daha iyi sonuçlar elde edilebilir. Aynı şekilde farklı sosyal medya problemleri için metasezgisel optimizasyon algoritmaları uyarlanabilir. Önerilen yöntemler çok büyük veri kümeleri için dağıtık ve paralel olarak gerçekleştirilebilir.

KAYNAKLAR

- [1] Maynard, D. C. (2019). *In The Internet and Health in Brazil*. Springer, Cham.
- [2] Kaplan, A. M., Haenlein, M. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media, *Business Horizons*, 53(1), 59–68. doi: 10.1016/j.bushor.2009.09.003
- [3] Tang, Q., Gu, B., Whinston, A. B. (2012). Content contribution for revenue sharing and reputation in social media: A dynamic structural model, *Journal of Management Information Systems*, 29(2), 41–76. doi: 10.2753/MIS0742-1222290203
- [4] Social media, https://en.wikipedia.org/wiki/Social_media, Eriřim: 21 Kasım 2019.
- [5] Lazer D.M., Baum, M.A., Benkler, Y., Berinsky, A.J, Greenhill, K.M., Menczer. F., Metzger, M., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E., Watts, D., Zittrain J. L. (2018). The science of fake news. *Science*. 2018; 359(6380), 1094–1096. doi: 10.1126/science.aao2998
- [6] Qayyum, S., Dar, H. (2018). A survey of data mining techniques for crime detection, *University of Sindh Journal of Information and Communication Technology*, 2(1), 1-6.
- [7] Lan, K., Wang, D. T., Fong, S., Liu, L. S., Wong, K. K., Dey, N. (2018). A survey of data mining and deep learning in bioinformatics, *Journal of medical systems*, 42(8), 139. doi: 10.1007/s10916-018-1003-9
- [8] Maingi, M.N. (2013). Survey on data preprocessing concept applicable in data mining, *International Journal of Science and Research*, 4(2), 1901-1902.
- [9] Krishnaiah, V., Narsimha, D. G., Chandra, D. N. S. (2013). Diagnosis of lung cancer prediction system using data mining classification techniques. *International Journal of Computer Science and Information Technologies*, 4(1), 39-45.
- [10] Geetha, M. C. S. (2015). A survey on data mining techniques in Agriculture. *International journal of innovative research in computer and communication engineering*, 3(2), 887-892.
- [11] Sato, Y., Izui, K., Yamada, T., Nishiwaki, S. (2019). Data mining based on clustering and association rule analysis for knowledge discovery in multiobjective topology optimization. *Expert Systems with Applications*, 119, 247-261. doi: 10.1016/j.eswa.2018.10.047
- [12] Altunbey, F., Alatař, B. (2015). Sosyal ađ analizi için sosyal tabanlı yapay zekâ optimizasyon algoritmalarının incelenmesi. *International Journal of Pure and Applied Sciences*, 1(1), 33-52.
- [13] Alatař, B. (2007). *Kaotik haritalı parçacık sürü optimizasyonu algoritmaları geliřtirme*, Doktora Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü.
- [14] Yang, X. S. 2010. *Nature-inspired metaheuristic algorithms*. Luniver press.

- [15] Karaboğa, D. 2011. *Yapay zekâ optimizasyon algoritmaları*. Nobel Yayın Dağıtım, Ankara
- [16] Sacco, W. F., De Oliveira, C. R. (2005). *A new stochastic optimization algorithm based on a particle collision metaheuristic*, 6th WCSMO, Rio de Janeiro, Brasil.
- [17] Kripka, M., Kripka, R. M. L. 2008. *Big crunch optimization method*, International conference on engineering optimization, Brazil.
- [18] Birbil SI, Fang SC (2003) An electromagnetism-like mechanism for global optimization, *Journal of Global Optimization*, 25(3), 263– 282. doi: 10.1023/A:1022452626305
- [19] Atashpaz-Gargari, E., Lucas, C. (2007). *Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition*, 2007 IEEE congress on evolutionary computation, Singapore.
- [20] Borji A, Hamidi, M. (2009) A new approach to global optimization motivated by parliamentary political competitions, *Journal of Innovative Computing, Information and Control*, 5(6), 1643-1653.
- [21] Ashrafi, S.M., Dariane, A.B. (2011). *A novel and effective algorithm for numerical optimization: melody search (MS)*. 11th IEEE international conference on hybrid intelligent systems, Malaysia.
- [22] Mora-Gutiérrez, R.A., Ramírez-Rodríguez, J., Rincón-García, E.A. (2014). An optimization algorithm inspired by musical composition, *Artificial Intelligence Review*, 41(3), 301-315. doi: 10.1007/s10462-011-9309-8
- [23] Kennedy. J., Eberhart, R.C. (1995). *Particle swarm optimization*, IEEE international conference on neural networks, Australia.
- [24] Mirjalili, S., Lewis, A. (2016). The whale optimization algorithm, *Advances in Engineering Software*, 95, 51-67. doi: 10.1016/j.advengsoft.2016.01.008
- [25] Lam. A.Y., Li, V.O. (2010). Chemical-reaction-inspired metaheuristic for optimization, *IEEE Transactions on Evolutionary Computation*, 14(3), 381-399.
- [26] Alatas, B. (2011). ACROA: artificial chemical reaction optimization algorithm for global optimization, *Expert Systems with Applications*, 38(10):13170–13180. doi: 10.1016/j.eswa.2011.04.126
- [27] Holland J.H., Goldberg, D.E. (1989). *Genetic algorithms in search, optimization and machine learning*, AddisonWesley Longman Publishing, Boston.
- [28] Kashan, A.H. (2011). An efficient algorithm for constrained global optimization and application to mechanical engineering design: league championship algorithm (LCA), *Computer-Aided Design*, 43(12), 1769-1792. doi: 10.1016/j.cad.2011.07.003
- [29] Khaji, E. (2014). Soccer League Optimization: A heuristic Algorithm Inspired by the Football System in European Countries. *arXiv preprint arXiv:1406.4462*.
- [30] Mirjalili, S. (2016). SCA: a Sine cosine algorithm for solving optimization problems, *Knowledge-Based Systems*, 96, 120-133. doi: 10.1016/j.knosys.2015.12.022

- [31] Allcott, H., Gentzkow, M. (2017). Social media and fake news in the 2016 election, *Journal of Economic Perspectives*, 31(2). doi: 10.1257/jep.31.2.211
- [32] Shu, K., Sliva, A., Wang, S., Tang, J., Liu, H. (2017). Fake news detection on social media: A data mining perspective, *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36.
- [33] Zhou, X., Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities, *arXiv preprint arXiv:1812.00315*.
- [34] Fallis, D. (2009). *A conceptual analysis of disinformation*, *iConference*, North Carolina, ABD.
- [35] Situngkir, H. (2011). Spread of hoax in social media, *BFI Working Papers WP-5*. doi: 10.2139/ssrn.1831202
- [36] Stevens, E. M., McIntyre, K. (2019). The layers of the onion: the impact of satirical news on affect and online sharing behaviors, *Electronic News*, 13(2), 78-92. doi: 10.1177/1931243119850264
- [37] Volkova, S., Jang, J. Y. (2018). *Misleading or falsification: Inferring deceptive strategies and types in online news and social media*, *2018 World Wide Web Conference*, USA.
- [38] Volkova, S., Shaffer, K., J. Y. Jang, J. Y., Hodas, N. (2017). *Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter*, *55th Annual Meeting of the Association for Computational Linguistics*, Canada.
- [39] Howard, P. N., Bolsover, G., Kollanyi, B., Bradshaw, S., & Neudert, L. M. (2017). Junk news and bots during the US election: What were Michigan voters sharing over Twitter, *Comprop Data Memo 2017*.
- [40] Sharma, K., Qian, F., Jiang, H., Ruchansky, N., Zhang, M., Liu, Y. (2019). Combating fake news: A survey on identification and mitigation techniques, *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(3), 21. doi: 10.1145/3305260
- [41] Roozenbeek, J., Linden, S. (2018). The fake news game: actively inoculating against the risk of misinformation, *Journal of Risk Research*, 22(5), 570-580. doi: 10.1080/13669877.2018.1443491
- [42] Ozbay, F. A., Alatas, B. (2019). A novel approach for detection of fake news on social media using metaheuristic optimization algorithms, *Elektronika ir Elektrotechnika*, 25(4), 62-67. doi: 10.5755/j01.eie.25.4.23972
- [43] Inside Russia's Social Media War on America, <https://time.com/4783932/inside-russia-social-media-war-america/>, Erişim: 28 Kasım 2019.
- [44] 'Barış Pınarı Harekatı' aleyhine sahte fotoğraflarla manipülasyon çabası, <https://www.cnnturk.com/turkiye/baris-pinari-harekat-i-aleyhine-sahte-fotograflarla-manipulasyon-cabasi>, Erişim: 28 Kasım 2019.

- [45] Ahmed, H. (2017). *Detecting opinion spam and fake news using n-gram analysis and semantic similarity*, PhD Thesis, University of Victoria, Department of Electrical and Computer Engineering.
- [46] Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., Menczer, F. (2018). The spread of low-credibility content by social bots, *Nature communications*, 9(1), 4787. doi: 10.1038/s41467-018-06930-7
- [47] Kumar S., Shah, N. (2018). False information on web and social media: A survey, *arXiv preprint arXiv:1804.08559*. doi: 10.1145/nnnnnnn.nnnnnnn
- [48] Conroy, N. J., Rubin, V. L., Chen, Y. (2015). *Automatic deception detection: Methods for finding fake news*, *Second Workshop on Computational Approaches to Deception Detection*, California.
- [49] Pierri, F., Ceri, S. (2019). False News On Social Media: A Data-Driven Survey, *arXiv preprint arXiv:1902.07539*.
- [50] Pérez-Rosas, V., Kleinberg, B., Lefevre, A., Mihalcea, R. (2017). Automatic detection of fake news, *arXiv preprint arXiv:1708.07104*.
- [51] Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news, *arXiv preprint arXiv:1702.05638*. doi: 10.18653/v1/P18-1022
- [52] Rubin, V., Conroy, N., Chen, Y., Cornwell, S. (2016). *Fake news or truth? using satirical cues to detect potentially misleading news*, *Second Workshop on Computational Approaches to Deception Detection*, California.
- [53] Kim, J., Tabibian, B., Oh, A., Schölkopf, B., Gomez-Rodriguez, M. (2018). *Leveraging the crowd to detect and reduce the spread of fake news and misinformation*, *Eleventh ACM International Conference on Web Search and Data Mining*, California.
- [54] Tschitschek, S., Singla, A., Gomez Rodriguez, M., Merchant, A., Krause, A. (2018). *Fake news detection in social networks via crowd signals*, *Web Conference 2018*, Lyon, France.
- [55] Ahmed, H., Traore, I., Saad, S. (2017). *Detection of online fake news using N-gram analysis and machine learning techniques*, *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, Canada.
- [56] Zhang, L., Wang, S., Liu, B. (2018). Deep learning for sentiment analysis: A survey, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253. doi: 10.1002/widm.1253
- [57] Bhatt, G., Sharma, A., Sharma, S., Nagpal, A., Raman, B., Mittal, A. (2017). On the benefit of combining neural, statistical and external features for fake news identification, *arXiv preprint arXiv:1712.03935*.
- [58] Singhanian, S., Fernandez, N., Rao, S. (2017). *3han: A deep neural network for fake news detection*, *International Conference on Neural Information Processing*, China.

- [59] Fang, Y., Gao, J., Huang, C., Peng, H., Wu, R. (2019). Self multi-head attention-based convolutional neural networks for fake news detection, *PloS one*, 14(9). doi: 10.1371/journal.pone.0222713
- [60] Wang, W. Y. (2017). Liar, liar pants on fire: A new benchmark dataset for fake news detection, *arXiv preprint arXiv:1705.00648* doi: 10.18653/v1/P17-2067
- [61] Qian, F., Gong, C., Sharma, K., Liu, Y. (2018). *Neural user response generator: fake news detection with collective user intelligence*, *International Joint Conference on Artificial Intelligence*, Stockholm, Sweden.
- [62] Girgis, S., Amer, E., Gadallah, M. (2018). *Deep learning algorithms for detecting fake news in online text*, *13th International Conference on Computer Engineering and Systems (ICCES)*, Cairo, Egypt.
- [63] Monti, F., Frasca, F., Eynard, D., Mannion, D., Bronstein, M. M. (2019). Fake news detection on social media using geometric deep learning, *arXiv preprint arXiv:1902.06673*.
- [64] Ruchansky, N., Seo, S., Liu, Y. (2017). *Csi: A hybrid deep model for fake news detection*, *2017 ACM on Conference on Information and Knowledge Management*, Singapore.
- [65] Farajtabar, M., Yang, J., Ye, X., Xu, H., Trivedi, R., Khalil, E., Li, S., Song, L., Zha, H. (2017). *Fake news mitigation via point process based intervention*, *34th International Conference on Machine Learning*, Sydney, Australia
- [66] Dong, M., Yao, L., Wang, X., Benatallah, B., Sheng, Q. Z., Huang, H. (2018). *DUAL: a deep unified attention model with latent relation representations for fake news detection*, *International Conference on Web Information Systems Engineering*, Dubai, UAE.
- [67] Shu, K., Wang, S., Liu, H. (2019). *Beyond news contents: The role of social context for fake news detection*, In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, Australia.
- [68] Fairbanks, J., Fitch, N., Knauf, N., Briscoe, E. (2018). *Credibility Assessment in the News: Do we need to read*, *MIS2, Marina Del Rey, CA, USA*.
- [69] Hosseinimotlagh, S., Papalexakis, E. E. (2018). *Unsupervised content-based identification of fake news articles with tensor decomposition ensembles*, *MIS2, Marina Del Rey, CA, USA*.
- [70] Jin, Z., Cao, J., Zhang, Y., Zhou, J., Tian, Q. (2016). Novel visual and statistical image features for microblogs news verification, *IEEE transactions on multimedia*, 19(3), 598-608. doi: 10.1109/TMM.2016.2617078
- [71] Huh, M., Liu, A., Owens, A., Efros, A. A. (2018). *Fighting fake news: Image splice detection via learned self-consistency*, *European Conference on Computer Vision (ECCV)*, Munich, Germany.
- [72] Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks, *arXiv preprint arXiv:1704.07506*.

- [73] Yang, S., Shu, K., Wang, S., Gu, R., Wu, F., Liu, H. (2019). *Unsupervised fake news detection on social media: A generative approach*, 33rd AAAI Conference on Artificial Intelligence, Hawaii, USA.
- [74] Volkova, S., Shaffer, K., Jang, J. Y., Hodas, N. (2017). *Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on twitter*, 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada. doi: 10.18653/v1/p17-2102
- [75] Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., Shu, L., Gao, J. (2018). *Eann: Event adversarial neural networks for multi-modal fake news detection*, 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, United Kingdom.
- [76] Liu, Y., Wu, Y. F. B. (2018). *Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks*, 32nd AAAI Conference on Artificial Intelligence, Louisiana, USA.
- [77] Mirjalili, S., Mirjalili, S. M., Lewis, A. (2014). Grey wolf optimizer, *Advances in Engineering Software*, 69, 46-61. doi: 10.1016/j.advengsoft.2013.12.007
- [78] Mech, L. D. (1999). Alpha status, dominance, and division of labor in wolf packs, *Canadian Journal of Zoology*, 77(8), 1196-1203. doi: 10.1139/z99-099
- [79] Muro, C., Escobedo, R., Spector, L., Coppinger, R. P. (2011). Wolf-pack (*Canis lupus*) hunting strategies emerge from simple rules in computational simulations, *Behavioural Processes*, 88(3), 192-197. doi: 10.1016/j.beproc.2011.09.006
- [80] Madin, L. P. (1990). Aspects of jet propulsion in salps, *Canadian Journal of Zoology*, 68(4), 765-777. doi: 10.1139/z90-111
- [81] Anderson, P. A., Bone, Q. (1980). Communication between individuals in salp chains II. Physiology, *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 210(1181), 559-574. doi: 10.1098/rspb.1980.0153 doi: 10.1098/rspb.1980.0153
- [82] Andersen, V., Nival, P. (1986). A model of the population dynamics of salps in coastal waters of the Ligurian Sea, *Journal of Plankton Research*, 8(6), 1091-1110. doi: 10.1093/plankt/8.6.1091
- [83] Henschke, N., Smith, J. A., Everett, J. D., Suthers, I. M. (2015). Population drivers of a *Thalia democratica* swarm: insights from population modelling, *Journal of Plankton Research*, 37(5), 1074-1087. doi: 10.1093/plankt/fbv024
- [84] Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., Mirjalili, S. M. (2017). Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems, *Advances in Engineering Software*, 114, 163-191. doi: 10.1016/j.advengsoft.2017.07.002
- [85] Yang, C., Gao, W., Liu, N., Song, C. (2015). Low-discrepancy sequence initialized particle swarm optimization algorithm with high-order nonlinear time-varying inertia weight, *Applied Soft Computing*, 29, 386-394. doi: 10.1016/j.asoc.2015.01.004

- [86] Kentzoglanakis, K., Poole, M. (2009). *Particle swarm optimization with an oscillating inertia weight*, 11th Annual conference on Genetic and evolutionary computation (GECCO), Canada.
- [87] Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., Al Najada, H. (2015). Survey of review spam detection using machine learning techniques, *Journal of Big Data*, 2(1), 23. doi: 10.1186/s40537-015-0029-9
- [88] Shree, S. B., Sheshadri, H. S. (2014). *An initial investigation in the diagnosis of Alzheimer's disease using various classification techniques*, 2014 IEEE International Conference on Computational Intelligence and Computing Research, Tamilnadu, India.
- [89] Platt, J. C. (1998). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*, Adv. Technical report no. 98-14, Microsoft Research.
- [90] Cohen, W. W. (1995). *Fast effective rule induction*, Twelfth International Conference on Machine Learning, California.
- [91] Buddhinath, G., Derry, D. (2006). *A Simple Enhancement to One Rule Classification*, Department of Computer Science & Software Engineering University of Melbourne, Australia.
- [92] Ruder, S. (2016). An overview of gradient descent optimization algorithms, *arXiv preprint arXiv:1609.04747*.
- [93] Breiman, L. (1996). Bagging predictors, *Machine Learning*, 24,123-140. doi: 10.1023/A:1018054314350
- [94] Tan, C., Chen, H., Xia, C. (2009). The prediction of cardiovascular disease based on trace element contents in hair and a classifier of boosting decision stumps, *Biological Trace Element Research*, 129(1-3), 9-19. doi: 10.1007/s12011-008-8279-4
- [95] Venkatesh, A., Jacob, S. G. (2016). Prediction of credit-card defaulters: a comparative study on performance of classifiers, *International Journal of Computer Applications*, 145(7). doi: 10.5120/ijca2016910702
- [96] Joshuva, A., Sugumaran, V. (2019). Selection of a meta classifier-data model for classifying wind turbine blade fault conditions using histogram features and vibration signals: a data-mining study, *Progress in Industrial Ecology, an International Journal*, 13(3), 232-251.
- [97] Karabatak, M., Mustafa, T. (2018). *Performance comparison of classifiers on reduced phishing website dataset*, 6th International Symposium on Digital Forensic and Security (ISDFS), Ankara, Türkiye.
- [98] Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., Ma, J. (2017). A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility, *Catena*, 151,147–160. doi: 10.1016/j.catena.2016.11.032
- [99] Furxhi, I., Murphy, F., Mullins, M., & Poland, C. A. (2019). Machine learning prediction of nanoparticle in vitro toxicity: A comparative study of classifiers and ensemble-classifiers

- using the Copeland Index, *Toxicology letters*, 312, 157-166. doi: 10.1016/j.toxlet.2019.05.016
- [100] Soni, J., Ansari, U., Sharma, D., & Soni, S. (2011). Predictive data mining for medical diagnosis: An overview of heart disease prediction, *International Journal of Computer Applications*, 17(8), 43-48. doi: 10.5120/2237-2860
- [101] Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms, *Physica A: Statistical Mechanics and its Applications*, 540. doi: 10.1016/j.physa.2019.123174
- [102] Gaines, B. R., Compton, P. (1995). Induction of ripple-down rules applied to modeling large databases, *Journal of Intelligent Information Systems*, 5, 211-228. doi: 10.1007/BF00962234
- [103] Arvanitakis, K., Avlonitis, M. (2016). *Identifying asperity patterns via machine learning algorithms*, *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Thessaloniki, Greece.
- [104] Ramchoun, H., Idrissi, M. A. J., Ghanou, Y., Ettaouil, M. (2016). Multilayer perceptron: architecture optimization and training, *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(1), 26-30. doi: 10.9781/ijimai.2016.415
- [105] Ben-David, A., Sterling, L., & Pao, Y. H. (1989). Learning and classification of monotonic ordinal concepts, *Computational Intelligence*, 5(1), 45-49. doi: 10.1111/j.1467-8640.1989.tb00314.x
- [106] Breiman, L. (2017). *Classification and regression trees*, Wadsworth.
- [107] Kalmegh, S. (2015). Analysis of weka data mining algorithm reptree, simple cart and randomtree for classification of indian news, *International Journal of Innovative Science, Engineering & Technology*, 2(2), 438-446.
- [108] Shafi, S., Hassan, S. M., Arshaq, A., Khan, M. J., & Shamail, S. (2008). *Software quality prediction techniques: A comparative analysis*, *4th International Conference on Emerging Technologies*, Bangkok, Thailand.
- [109] Coşkun, C., Baykal, A. (2011). *An application for comparison of data mining classification algorithms*, Karabük, Türkiye.
- [110] Quinlan, J. (1987). Simplifying decision trees, *International Journal of ManMachine Studies*, 27(3), 221–234. doi: 10.1006/ijhc.1987.0321
- [111] Dogan, N., Tanrikulu, Z. (2013). A comparative analysis of classification algorithms in data mining for accuracy, speed and robustness, *Information Technology and Management*, 14(2), 105-124. doi: 10.1007/s10799-012-0135-8
- [112] Cawley, G.C., Talbot, N.L. (2008). Efficient approximate leave-one-out cross-validation for kernel logistic regression, *Machine Learning*, 71(2-3), 243-264. doi: 10.1007/s10994-008-5055-9

- [113] Hühn, J., & Hüllermeier, E. (2009). FURIA: an algorithm for unordered fuzzy rule induction, *Data Mining and Knowledge Discovery*, 19(3), 293-319. doi: 10.1007/s10618-009-0131-8
- [114] Pal, M. (2005). Random forest classifier for remote sensing classification, *International Journal of Remote Sensing*, 26(1), 217-222. doi: 10.1080/01431160412331269698
- [115] Amini, M. R., Gallinari, P. (2002). *Semi-supervised logistic regression*, 15th European Conference on Artificial Intelligence, Lyon, France.
- [116] Bao, Y., Wu, H., Liu, X. (2016). *From prediction to action: A closed-loop approach for data-guided network resource allocation*, 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Fransisco, USA.
- [117] Parsania, V., Bhalodiya, N., Jani, N. N. (2014). Applying Naïve bayes, BayesNet, PART, JRip and OneR algorithms on hypothyroid database for comparative analysis, *International Journal of Darshan Institute on Engineering Research and Emerging Technology*, 3(1),60-64.
- [118] Sikora, R. (2015). *A modified stacking ensemble machine learning algorithm using genetic algorithms*, Handbook of Research on Organizational Transformations through Big Data Analytics, IGI Global.
- [119] Allahyari, M., Pouriye, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., Kochut, K., (2017). A brief survey of text mining: Classification, clustering and extraction techniques, *arXiv preprint arXiv:1707.02919*.
- [120] Vijayarani, S., Ilamathi, M. J., Nithya, M. (2015). Preprocessing techniques for text mining-an overview, *International Journal of Computer Science & Communication Networks*, 5(1), 7-16.
- [121] Cheng, C. H., Chen, H. H. (2019). Sentimental text mining based on an additional features method for text classification. *PloS one*, 14(6). doi: 10.1371/journal.pone.0217591
- [122] Salton, G., Wong A., Yang, C.S. (1975). A vector space model for automatic indexing, *Communications of the ACM*, 18(11), 613–620. doi: 10.1145/361219.361220
- [123] Sarkar, D., Jana, P. (2019). Analyzing user activities using vector space model in online social networks, *arXiv preprint arXiv:1910.0569*.
- [124] This analysis shows how fake election news stories outperformed real news on facebook, https://www.buzzfeed.com/craigsilverman/viralfake-election-news-outperformed-real-news-onfacebook?utm_term=.gl5qYwrP#.gt1ygZDN, Erişim 14 Mayıs 2018.
- [125] Horne, B. D., Adali, S. (2017). *This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news*, Eleventh International AAAI Conference on Web and Social Media, Canada.
- [126] Zimdars, M. (2016). *False, misleading, clickbaity, and satirical news sources*, Google Docs.

- [127] Engel, P. (2014). *Here are the most-and least-trusted news outlets in America*, Business Insider.
- [128] Niwattanakul, S., Singthongchai, J., Naenudorn, E., Wanapu, S. (2013). *Using of Jaccard coefficient for keywords similarity*, *International multiconference of engineers and computer scientists*, Hong Kong.



ÖZGEÇMİŞ

Feyza ALTUNBEY ÖZBAY

KİŞİSEL BİLGİLER

Doğum Yeri	: Elazığ
Doğum Yılı	: 1991
Uyruğu	: T.C.
Adres	: Fırat Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği, Elazığ
E-posta	: faltunbey@firat.edu.tr
Yabancı Dil	: İngilizce

EĞİTİM BİLGİLERİ

Yüksek Lisans	: “Çok amaçlı sosyal tabanlı metasezgisel optimizasyon algoritmaları ile sosyal ağlarda örtüşen topluluk keşfi”, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, Yazılım Mühendisliği Anabilim Dalı, 2015 Danışman: Prof. Dr. Bilal ALATAŞ
Lisans	: Fırat Üniversitesi, Bilgisayar Mühendisliği, 2013
Lise	: Elazığ Anadolu Lisesi, Elazığ, 2009

ARAŞTIRMA DENEYİMİ

Programlama Dilleri	: C, C++, C#, Java, Lisp, Python, PHP, Prolog
Paket Programlar	: MATLAB, Proteus, MPLab, GNS3.
İşletim Sistemleri	: Windows, Linux.

İŞ DENEYİMİ

2013 - ...	: Araştırma Görevlisi, Fırat Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği Bölümü
------------	---

AKADEMİK FAALİYETLER

Makaleler:

- Altunbey Ozbay, F., Alataş, B., (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and its Applications, 540.
- Altunbey Ozbay, F., Alataş, B., (2019). A Novel Approach for Detection of Fake News on Social Media Using Metaheuristic Optimization Algorithms. Elektronika Ir Elektrotechnika, 25(4),62-67.
- Altunbey Ozbay, F., Alataş, B., (2017). Discovery of Multi-Objective Overlapping Communities within Social Networks using a Socially Inspired Metaheuristic Algorithm. International Journal of Computer Networks and Applications (IJCNA), 4(6),148-158.

4. Altunbey, F., Alataş, B., (2015). Overlapping Community Detection in Social Networks Using Parliamentary Optimization Algorithm. *International Journal of Computer Networks and Applications*, 2(1),12-19.
5. Altunbey, F., Alataş, B., (2015) Review of Social-Based Artificial Intelligence Optimization Algorithms for Social Network Analysis. *International Journal of Pure and Applied Sciences*, 1(1),33-52.
6. Altunbey Ozbay, F., Alataş, B., (2016). Review of Computational Intelligence Method Inspired from Behavior of Water. *Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 16(1),137-147.
7. Altunbey Ozbay, F., Alataş B., Çevrimiçi sosyal medyada sahte haber tespiti, *DÜMF Mühendislik Dergisi*.

Bildiriler:

1. Altunbey Ozbay, F., Alataş, B., (2019). A Novel Approach for Detection of Fake News on Social Media Using Metaheuristic Optimization Algorithms. *Electronics* 2019.
2. Altunbey Ozbay, F., Alataş, B., (2016). Müzik Tabanlı Hesapsal Zekâ Algoritmalarının İncelenmesi. *1st International Conference on Engineering Technology and Applied Sciences*, 663-669.
3. Altunbey Ozbay, F., Alataş, B., (2016). A New Approach to Swarm based Computational Intelligence Whale Optimization. *International Conference on Engineering and Natural Sciences ICENS 2016*,653-659.
4. Altunbey Ozbay, F., Alataş, B., (2016). A Simple and Global Physics based Metaheuristic Method Water Evaporation Optimization. *International Conference on Engineering and Natural Sciences ICENS 2016*,660-664.
5. Altunbey Ozbay, F., Alataş, B., (2016). A New Approach to Artificial Intelligence Optimization Ant Lion Algorithm. *International Conference on Mathematics and Mathematics Education ICMME 2016*, 241-242.
6. Altunbey Ozbay, F., Alataş, B., (2017). Performance Analysis of Stochastic Fractal Search Algorithm for Global Optimization within Uni-modal and Multi-modal Benchmark Functions. *International Conference on Advances and Innovations in Engineering (ICAIE)*,944-949.
7. Altunbey Ozbay, F., Alataş, B., (2017). Evaporation Rate Based Water Cycle Algorithm for Engineering Design Optimization. *International Conference on Advances and Innovations in Engineering (ICAIE)*, 950-954.