



**T.C.
HARRAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**METASEZGİSEL ALGORİTMALARLA ÖZNİTELİK SEÇİMİ
YÖNTEMLERİNİN UYGULAMASI VE GELİŞTİRİLMESİ**

İSMAİL AÇAR

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

**Şanlıurfa
2026**



**T.C.
HARRAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ**

YÜKSEK LİSANS TEZİ

**METASEZGİSEL ALGORİTMALARLA ÖZNİTELİK SEÇİMİ
YÖNTEMLERİNİN UYGULAMASI VE GELİŞTİRİLMESİ**

İSMAİL AÇAR

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
Tez Danışmanı: Prof. Dr. İBRAHİM BERKAN AYDİLEK**

**Şanlıurfa
2026**



**T.C.
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
YÜKSEK LİSANS TEZ ONAY SAYFASI**

Prof. Dr. İbrahim Berkan AYDİLEK danışmanlığında, İsmail AÇAR'ın hazırladığı “Metasezgisel Algoritmalarla Öznitelik Seçimi Yöntemlerinin Uygulaması ve Geliştirilmesi” konulu bu çalışma 12/01/2026 tarihinde aşağıdaki jüri tarafından Harran Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

☒ Oybirliğiyle / ☐ Oy çokluğu ile

İmza

Danışman	: Prof. Dr. İbrahim Berkan AYDİLEK	e-imzalıdır
Üye	: Doç. Dr. Mehmet Bilal ER	e-imzalıdır
Üye	: Dr. Öğr. Üyesi İsmail İLHAN	e-imzalıdır

Bu tezin Bilgisayar Mühendisliği Ana Bilim Dalında yapıldığını ve enstitümüz kurallarına göre düzenlendiğini onaylarım.

Dr. Öğr. Üyesi Sabri AKIN
Enstitü Müdürü

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

	T.C. HARRAN ÜNİVERSİTESİ TEZLİ YÜKSEK LİSANS ORİJİNALLİK RAPORU VE BEYAN FORMU	Doküman No: FRM-0072
		Revizyon No: 03
		Yayın Tarihi: 10.09.2020
		Revizyon Tarihi: 30.01.2024
		Sayfa No: 1 / 1

ÖĞRENCİNİN	
Öğrenci No	235121011
Adı / Soyadı	İsmail AÇAR
Anabilim Dalı/Programı	Bilgisayar Mühendisliği Anabilim Dalı / Tezli Yüksek Lisans
Tez Konusu	Metasezgisel Algoritmalarla Öznitelik Seçimi Yöntemlerinin Uygulaması ve Geliştirilmesi

FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Yukarıda başlığı belirtilen Metasezgisel Algoritmalarla Öznitelik Seçimi Yöntemlerinin Uygulaması ve Geliştirilmesi çalışmamın a) Kapak sayfası, b) Giriş, c) Ana bölümler ve d) Sonuç kısımlarından oluşan toplam 65 sayfalık kısmına ilişkin, 10/03/2026 tarihinde danışmanım tarafından **Turnitin** adlı intihal tespit programından aşağıda belirtilen filtrelemeler uygulanarak alınmış olan orijinallik raporuna göre, benzerlik oranı %14'tür. Ayrıca Metasezgisel Algoritmalarla Öznitelik Seçimi Yöntemlerinin Uygulaması ve Geliştirilmesi hiçbir blok kopyalama içermemektedir.

Uygulanan filtrelemeler:

- 1- Kabul/Onay ve Bildirim sayfaları hariç,
- 2- Kaynakça hariç
- 3- Alıntılar hariç/dâhil
- 4- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

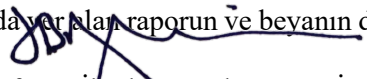
Yükseköğretim Kurulu Bilimsel Araştırma ve Yayın Etiği Yönergesi, Harran Üniversitesi Bilimsel Araştırma ve Yayın Etiği Yönergesini ve Harran Üniversitesi Lisansüstü Eğitim Enstitüsü tez çalışması Orijinallik Raporu alınması ve kullanılması Uygulama Esasları'nı inceledim ve bu Uygulama Esasları'nda belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim. Etik ihlal tespiti halinde, Enstitü yönetim kurulunca, diplomamın iptal edilmesini kabul ediyorum.

Gereğini saygılarımla arz ederim.

10/03/2026
Hazırlayan Öğrencinin



Yukarıda yer alan raporun ve beyanın doğruluğunu onaylarım. 10/03/2026


Prof. Dr. İbrahim Berkan AYDİLEK
Danışmanın Unvanı-Adı-Soyadı
(İmzası)

TEŐEKKÖR

Bu tez alıřmasının konusunun belirlenmesinde ve uygulanmasında deęerli bilgi, tecrübe ve yönlendirmeleriyle bana her zaman yol gösteren kıymetli danışmanım Sayın Prof. Dr. İbrahim Berkan AYDİLEK hocama sonsuz teşekkür ederim.

Ayrıca, hayatımın her kademesinde yanımda olan, desteklerini esirgemeyen deęerli anneme, babama ve kardeşlerime; yol boyunca karşılaşılan tüm olumsuzluklara rağmen asla pes etmeden, alıřma disiplinimden ödün vermeden yola devam ettięim için kendime gönülden teşekkür ediyorum.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
ŞEKİLLER DİZİNİ	iii
ÇİZELGELER DİZİNİ	iv
KISALTMALAR	v
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR	3
2.1. Metasezgisel Yöntemlere Dayalı Çalışmalar	3
2.2. Geleneksel Yöntemlere Dayalı Çalışmalar	8
3. GEREÇ VE YÖNTEM	13
3.1. Geleneksel Makine Öğrenmesi ve Metasezgisel Yöntemler Kullanılarak Yapılan Çalışma	13
3.1.1. Kullanılan Veri kümeleri	13
3.1.2. Sınıflandırma Yöntemleri	14
3.1.2.1. Karar Ağacı	14
3.1.2.2. Rastgele Orman Algoritması	15
3.1.2.3. Gradyan Artırma Algoritması	17
3.1.2.4. Hafif Gradyan Artırma Algoritması	17
3.1.2.5. Aşırı Gradyan Artırma Algoritması	18
3.1.3. Kullanılan Öz nitelik Seçimi Yöntemleri	19
3.1.3.1. Geleneksel Makine Öğrenmesine Dayalı Öz nitelik Seçimi Yöntemleri	19
3.1.3.1.1. Korelasyon Analizi	19
3.1.3.1.2. Hipotez Testi	19
3.1.3.1.3. Lasso Yöntemi	20
3.1.3.2. Metasezgisel Yöntemlere Dayalı Öz nitelik Seçimi Yöntemleri	20
3.1.3.2.1. Parçacık Sürü Optimizasyon Algoritması	20
3.1.3.2.2. Genetik Algoritma	24
3.1.3.2.2.1. Parametre kodlama	25
3.1.3.2.2.2. Başlangıç Popülasyonu	26
3.1.3.2.2.3. Uygunluk Fonksiyonu	26
3.1.3.2.2.4. Kromozom Seçimi	26
3.1.3.2.2.5. Çaprazlama ve Mutasyon	27
3.1.3.2.2.6. Sonlandırma	27
3.2. Önerilen SVM-DLOFS Kullanılarak Yapılan Çalışmalar	28
3.2.1. Kullanılan Veri Kümeleri	28
3.2.2. Draco Lizard Optimizer	29
3.2.3. Öz nitelik Seçimi için Önerilen SVM-DLOFS Sınıflandırıcısı	29
3.3. Sınıflandırma Metrikleri	31
4. BULGULAR	34
4.1. Geleneksel Makine Öğrenmesi ve Metasezgisel Yöntemler Kullanılarak Elde Edilen Bulgular	34
4.1.1. Sonar Veri Kümesi	35
4.1.2. Parkinson Veri Kümesi	39
4.1.3. Breast Cancer Veri Kümesi	44
4.2. Önerilen SVM-DLOFS Kullanılarak Elde Edilen Bulgular	51
5. TARTIŞMA	55
6. SONUÇLAR	57
7. ÖNERİLER	59
KAYNAKLAR	60
ÖZGEÇMİŞ	65

ÖZET

YÜKSEK LİSANS TEZİ

METASEZGİSEL ALGORİTMALARLA ÖZNETELİK SEÇİMİ YÖNTEMLERİNİN UYGULAMASI VE GELİŞTİRİLMESİ

İSMAİL AÇAR

HARRAN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Tez Danışman: Prof. Dr. İBRAHİM BERKAN AYDİLEK
Yıl: 2026, Sayfa : 65

Öznitelik seçimi veri kümelerinde gereksiz ve alakasız öznitelikleri elimine ederek makine öğrenmesi problemlerinde önemli bir rol oynamaktadır. Öznitelik seçimi süreci, modelin doğruluğunu artırmanın yanı sıra eğitim süresini kısaltarak modelin yorumlanabilirliğini de artırmaktadır. Öznitelik seçimi yüksek boyutlu veri kümelerine uygulandığında aşırı öğrenme riskini azaltarak daha genellenebilir sonuçlar elde edilebilmektedir. Bu tez kapsamında iki farklı çalışmadan elde edilen deneysel çalışmalar sunulmuştur. İlk olarak UCI veri deposundan elde edilen 3 farklı veri kümesi kullanılarak 5 farklı makine öğrenmesi algoritması 10 katlı çapraz doğrulama yöntemi uygulanarak algoritmaların başarımları ve çalışma süreleri ölçülmüştür. Öznitelik seçimi yöntemleri olarak, meta-sezgisel yaklaşımlardan Genetik Algoritma (GA) ve Parçacık Sürü Optimizasyonu (PSO); geleneksel yöntemler arasında ise Korelasyon Analizi, Lasso ve Hipotez Testi tercih edilmiştir. Yapılan deneysel çalışmalar göz önüne alındığında GA ve PSO tabanlı öznitelik seçiminin, geleneksel yöntemlere kıyasla hem sınıflandırma başarımında hem de işlem süresinde daha etkin sonuçlar verdiği elde edilen deneysel verilerle ortaya konmuştur. İkinci çalışma olarak 2025 yılında ortaya çıkmış olan güncel metasezgisel algoritmalar olan Draco Kertenkele Optimizasyon (DLO) algoritması, destek vektör makinesi ile birlikte kullanılarak öznitelik seçimi problemi için SVM-DLOFS yöntemi önerilmiştir. Önerilen yöntem UCI veri deposundan elde edilen 6 farklı veri kümesi üzerinde deneysel çalışmalar yapılmış ve SVM algoritması ile 5 katlı çapraz doğrulama yöntemi kullanılarak karşılaştırılmıştır. Deneysel çalışmalar seçilen öznitelik sayısı, en iyi skor, en kötü skor, ortalama skor ve standart sapma değerleriyle analiz edilmiştir. Elde edilen deneysel çalışmalar incelendiğinde önerilen SVM-DLOFS yönteminin öznitelik seçiminde daha başarılı sonuçlar verdiği gözlemlenmiştir.

ANAHTAR KELİMELELER: Metasezgisel algoritmalar, Draco kertenkele optimizasyon, Genetik algoritma, Parçacık sürü optimizasyon algoritması, Öznitelik seçimi

ABSTRACT

MASTER THESIS

APPLICATION AND DEVELOPMENT OF FEATURE SELECTION METHODS USING METAHEURISTIC ALGORITHMS

İSMAİL AÇAR

**HARRAN UNIVERSITY
INSTITUTE OF GRADUATE EDUCATION
DEPARTMENT OF COMPUTER ENGINEERING**

**Thesis Supervisor: Prof. Dr. İBRAHİM BERKAN AYDİLEK
Year: 2026, Page : 65**

Feature selection plays an important role in machine learning problems by eliminating unnecessary and irrelevant features in datasets. In addition to improving the accuracy of the model, the feature selection process also shortens the training time and increases the interpretability of the model. When applied to high-dimensional datasets, feature selection reduces the risk of overfitting and provides more generalizable results. Within the scope of this thesis, two different experimental studies have been presented. First, three different datasets obtained from the UCI repository were used, and the performance and execution times of five different machine learning algorithms were measured by applying 10-fold cross-validation. As feature selection methods, Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) were chosen among meta-heuristic approaches, while Correlation Analysis, Lasso, and Hypothesis Testing were preferred among traditional methods. Considering the experimental studies, it has been demonstrated with the obtained results that GA- and PSO-based feature selection produced more effective outcomes in terms of both classification performance and computational time compared to traditional methods. As the second study, the Draco Lizard Optimization (DLO) algorithm, which is one of the recently developed meta-heuristic algorithms introduced in 2025, was employed together with the Support Vector Machine to propose the SVM-DLOFS method for the feature selection problem. The proposed method was tested on six different datasets obtained from the UCI repository, and compared with the SVM algorithm using 5-fold cross-validation. The experimental studies were analyzed with respect to the number of selected features, best score, worst score, average score, and standard deviation values. When the obtained experimental results were examined, it was observed that the proposed SVM-DLOFS method yielded more successful results in feature selection.

KEYWORDS: Meta-heuristic algorithms, Draco lizard optimization, Genetic algorithm, Particle swarm optimization algorithm, Feature selection

ŞEKİLLER DİZİNİ

Şekil 3.1.	Karar Ağacı Algoritmasının Bölünmesi	15
Şekil 3.2.	Rastgele Orman Algoritmasının Çalışma Mantığı	16
Şekil 3.3.	Parçacık Sürü Optimizasyon Algoritması Akış Şeması	22
Şekil 3.4.	Genetik Algoritma Optimizasyon Algoritması Akış Şeması	25
Şekil 3.5.	SVM-DLOFS Akış Şeması	30
Şekil 3.6.	Karmaşıklık matrisi	32
Şekil 4.1.	Sonar Veri kümesi ortalama doğruluk sonuçları	39
Şekil 4.2.	Parkinson Veri kümesi ortalama doğruluk sonuçları	44
Şekil 4.3.	Breast Cancer Veri kümesi ortalama doğruluk sonuçları	49
Şekil 4.4.	Hafif Gradyan Artırma Makineleri ile 3 veri kümesinin ortalama doğruluk, duyarlılık, kesinlik, F1- Skor sonuçları	50
Şekil 4.5.	SVM-DLOFS doğruluk kutu grafiği	54

ÇİZELGELER DİZİNİ

Çizelge 3.1. Kullanılan veri kümeleri	14
Çizelge 3.2. Kullanılan veri kümeleri	28
Çizelge 4.1. Sonar Veri kümesi Deney sonuçları (10 cv doğruluk sonuçları/(çalışma süresi (ms))	35
Çizelge 4.2. Sonar Veri kümesi Deney sonuçları (10 cv kesinlik sonuçları/(çalışma süresi (ms))	36
Çizelge 4.3. Sonar Veri kümesi Deney sonuçları (10 cv duyarlılık sonuçları/(çalışma süresi (ms))	37
Çizelge 4.4. Sonar Veri kümesi Deney sonuçları (10 cv f1-skor sonuçları/(çalışma süresi (ms))	38
Çizelge 4.5. Parkinson Veri kümesi Deney sonuçları (10 cv doğruluk sonuçları/(çalışma süresi (ms))	40
Çizelge 4.6. Parkinson Veri kümesi Deney sonuçları (10 cv kesinlik sonuçları/(çalışma süresi (ms))	41
Çizelge 4.7. Parkinson Veri kümesi Deney sonuçları (10 cv duyarlılık sonuçları/(çalışma süresi (ms))	42
Çizelge 4.8. Parkinson Veri kümesi Deney sonuçları (10 cv f1-skor sonuçları/(çalışma süresi (ms))	43
Çizelge 4.9. Breast Cancer Veri kümesi Deney sonuçları (10 cv doğruluk sonuçları/(çalışma süresi (ms))	45
Çizelge 4.10. Breast Cancer Veri kümesi Deney sonuçları (10 cv kesinlik sonuçları/(çalışma süresi (ms))	46
Çizelge 4.11. Breast Cancer Veri kümesi Deney sonuçları (10 cv duyarlılık sonuçları/(çalışma süresi (ms))	47
Çizelge 4.12. Breast Cancer Veri kümesi Deney sonuçları (10 cv f1-skor sonuçları/(çalışma süresi (ms))	48
Çizelge 4.13. SVM-DLOFS ve SVM deney sonuçları	53

KISALTMALAR

ACO	Karınca Kolonisi Optimizasyonu
AUC	Eğri Altındaki Alan
BDB	Birini Dışarıda Bırak
DLO	Draco Lizard Optimizer
DLOFS	Draco Lizard Optimizer Feature Selection
FN	Yanlış Negatif
FP	Yanlış Pozitif
GA	Genetik Algoritma
GBM	Gradyan Artırma Makineleri
KA	Karar Ağacı
KNN	K-En Yakın Komşu
LightGBM	Hafif Gradyan Artırma Makineleri
LR	Lojistik Regresyon
MÖ	Makine Öğrenmesi
NB	Naïve Bayes
PSO	Parçacık Sürü Optimizasyon
RO	Rastgele Orman
SVM	Destek Vektör Makineleri
TBA	Temel Bileşen Analizi
TN	Gerçek Negatif
TP	Gerçek Pozitif
XGBoost	Aşırı Gradyan Artırma Makineleri
YSA	Yapay Sinir Ağları

1. GİRİŞ

Veri bilimi ve makine öğrenmesi alanlarında, veri kümelerinin boyutunun artması, mevcut donanım kaynaklarının daha yoğun kullanılmasına ve buna bağlı olarak işlem hızında azalmaya yol açmaktadır. Özellikle makine öğrenmesi uygulamalarında, çoğu zaman çok sayıda öznitelik içeren veri kümeleriyle karşılaşilmektedir. Bu durum, hem hesaplama maliyetini artırmakta hem de modelin performansını olumsuz etkileyebilmektedir. Bu nedenle, öznitelik seçimi süreci, makine öğrenmesi algoritmalarının başarımını doğrudan etkileyen kritik bir adımdır. Öznitelik seçimi, bir optimizasyon problemi olarak ele alınabilir; modelin başarısı için en anlamlı özniteliklerin belirlenmesi ve gereksiz olanların elenmesi, algoritmanın doğruluk oranını ve genellenebilirliğini artırmaktadır. Ayrıca, gereksiz özniteliklerin veri setinden çıkarılması, aşırı öğrenme riskini azaltmakta ve daha güvenilir, genellenebilir sonuçların elde edilmesine olanak sağlamaktadır. Bu bağlamda, öznitelik seçimi yöntemlerinin geliştirilmesi ve uygulanması, makine öğrenmesi modellerinin etkinliği açısından büyük önem taşımaktadır (Açar ve Aydılek, 2024).

Son yıllarda, makine öğrenmesi uygulamalarında hem örnek sayısı hem de boyut açısından verinin hacmi olağanüstü bir şekilde artış göstermiştir. Verinin bu denli hızlı artışı, etkili ve verimli veri yönetimi açısından önemli zorluklar doğurmaktadır. Farklı türdeki verilerden otomatik olarak bilgi keşfetmek amacıyla veri madenciliği ve makine öğrenmesi tekniklerinin uygulanması giderek daha fazla önem kazanmaktadır. Ancak, yüksek boyutlu veriler üzerinde bu algoritmalar kullanıldığında, “boyutlanma laneti” olarak bilinen temel bir sorun ortaya çıkmaktadır. Bu olgu, yüksek boyutlu uzayda verinin seyrekleşmesiyle birlikte, düşük boyutlu uzaylar için tasarlanmış algoritmaların performansının olumsuz etkilenmesine yol açmaktadır. Ayrıca, çok sayıda öznitelik içeren veri kümelerinde, öğrenme modelleri aşırı öğrenmeye daha yatkın hale gelmekte ve bu durum, modelin yeni veriler üzerindeki başarımını düşürebilmektedir. Yüksek boyutlu veriler, aynı zamanda bellek ve işlem maliyetlerini de önemli ölçüde artırmaktadır. Bu sorunların üstesinden gelmek için boyut indirgeme yöntemleri oldukça etkili araçlar olarak öne çıkmaktadır (Li vd.,2017).

Boyut indirgeme temelde öznitelik çıkarımı ve öznitelik seçimi olmak üzere iki ana başlık altında incelenebilir. Öznitelik çıkarımı, orijinal yüksek boyutlu öznitelikleri daha düşük boyutlu yeni bir öznitelik uzayına yansıtırken, öznitelik seçimi ise model oluşturma sürecinde doğrudan en anlamlı özniteliklerin alt kümesini belirlemeye odaklanır. Her iki yaklaşım da öğrenme performansını

artırmak, hesaplama verimliliğini yükseltmek, bellek gereksinimini azaltmak ve daha iyi genellenebilir modeller oluşturmak açısından avantajlar sunmaktadır. Özellikle öznitelik seçimi, orijinal özniteliklerin fiziksel anlamını koruduğu ve modelin yorumlanabilirliğini artırdığı için metin madenciliği ve genetik analiz gibi birçok uygulamada tercih edilmektedir. Ayrıca, gerçek dünya verilerinde çoğu zaman gereksiz, tekrarlı ve gürültülü öznitelikler bulunmakta; bu özniteliklerin elenmesi, hem depolama ve işlem maliyetlerini azaltmakta hem de bilgi kaybı olmadan model başarımını korumaktadır (Li vd.,2017).

Öznitelik seçimi stratejileri ise genel olarak üç ana başlık altında toplanır: sarmalayıcı, filtre ve gömülü yöntemler. Sarmalayıcı yöntemler, belirli bir öğrenme algoritmasının başarımına göre öznitelik alt kümelerini değerlendirir ve en iyi performans sağlayan öznitelik setini seçer. Ancak, yüksek boyutlu veri kümelerinde arama uzayının büyüklüğü nedeniyle pratikte kullanımları sınırlıdır. Filtre yöntemleri ise, herhangi bir öğrenme algoritmasından bağımsız olarak, özniteliklerin veri üzerindeki istatistiksel özelliklerine göre önem derecesi belirler ve genellikle daha hızlıdır. Gömülü yöntemler ise, öznitelik seçimini model öğrenme sürecine entegre ederek hem filtre hem de sarmalayıcı yöntemlerin avantajlarını bir arada sunar. Son yıllarda, bu üç yöntemin bir arada kullanıldığı hibrit yaklaşımlar da geliştirilmiş ve öznitelik seçiminin kararlılığını artırmak amacıyla uygulanmaya başlanmıştır (Venkatesh ve Anuradha, 2019).

Yüksek boyutlu verilerle çalışırken öznitelik seçimi ve boyut indirgeme yöntemleri, hem model başarımını artırmak hem de hesaplama maliyetlerini azaltmak açısından vazgeçilmez araçlar olarak öne çıkmaktadır. Bu yöntemlerin doğru ve etkin bir şekilde uygulanması, makine öğrenmesi ve veri madenciliği uygulamalarında daha güvenilir ve yorumlanabilir sonuçlar elde edilmesini sağlamaktadır (Li vd.,2017).

2. ÖNCEKİ ÇALIŞMALAR

Bu bölümde öznitelik seçimi yöntemlerinin kullanımı ile ilgili literatürde yer alan bazı çalışmalar ele alınmış ve kısaca bahsedilmiştir. Önceki çalışmalar bölümü iki başlık altında ele alınmıştır. Metasezgisel yöntemlere dayalı çalışmalar ve geleneksel yöntemlere dayalı çalışmalar olarak ilgili alanlarda yapılmış çalışmalardan bahsedilmiştir.

2.1. Metasezgisel Yöntemlere Dayalı Çalışmalar

Bu bölümde Metasezgisel yöntemlere dayalı öznitelik seçimi ile ilgili yapılan çalışmalardan bahsedilmiştir.

Aydilek ve Açar (2025), tarafından yapılan çalışmada yeni bir metasezgisel algoritma olan draco lizard optimizer (DLO) yöntemi ve destek vektör makineleri algoritması kullanılarak yeni bir sınıflandırma modeli önerilmiştir. Çalışma kapsamında 5 farklı veri kümesi kullanılmış ve destek vektör makineleri algoritması ile dört farklı metrik ile 10 defa çalıştırılmış ve en iyi, en kötü, ortalama, standart sapma ve ortalama öznitelik sayısı kriterleri baz alınarak karşılaştırılmıştır. Elde edilen deney sonuçları göz önüne alındığında önerilen yöntemin destek vektör makineleri algoritmasına göre daha başarılı sonuçlar verdiği gözlemlenmiştir.

Kılıçaslan vd. (2019), tarafından yapılan çalışmada 102 hastaya ait 10509 öznitelikten oluşan mikrodizi veri seti kullanılmıştır. Boyut indirgeme amacıyla temel bileşenler analizi (TBA) ile parçacık sürü optimizasyonu (PSO) uygulanmıştır. Böylelikle elde edilen indirgenmiş veri kümeleri, sınıflandırma aşamasında destek vektör makineleri ve k-en yakın komşuluk algoritmalarına giriş veri seti olarak sunulmuş ve modellerin sınıflandırma başarımları karşılaştırmalı olarak değerlendirilmiştir. Sonuçlar, PSO tabanlı boyut indirgeme yaklaşımının prostat kanserine ilişkin etkin genlerin tespitinde başarılı olduğunu göstermiş ve 50 öznitelik kullanıldığında %95,77 doğruluk ile en yüksek sınıflandırma performansı elde edilmiştir.

Alp ve Soygazi (2024), tarafından yapılan çalışmada diyabet hastalığının tespiti için 5 farklı makine öğrenmesi kullanılarak benzetilmiş tavlama algoritması ve genetik algoritma ile öznitelik seçimi yapılarak öznitelik seçimindeki sonuçları karşılaştırılmıştır. Her bir sınıflandırma işleminin başarısı doğruluk, kesinlik, geri çağırma ve F1 skor performans metrikleri ile raporlanmıştır. Önerilen yöntemler arasından en yüksek doğruluk değerinin 86% olarak LightGBM+GA modelinden elde edildiği tespit edilmiş ve öznitelik seçiminin başta LightGBM ve XGBoost

olmak üzere tüm modellerde sonucu iyileştirdiği tespit edilmiştir.

Gümüşçü vd. (2017), tarafından yapılan çalışmada Parkinson hastalığının teşhisinde kullanılmak amacıyla ses analizi işlemi yapılmış ve genetik algoritma kullanılarak öznelik seçimi işlemi yapılmıştır. Öznelik seçimi işlemi sonucunda elde edilen özneliklerin başarımları DVM sınıflandırıcısı ile çapraz doğrulama yöntemi kullanılarak ölçülmüştür. Elde edilen deneysel çalışmalar göz önüne alındığında DVM sınıflandırıcısı ve birini dışarıda bırak (BDB) çapraz doğrulama yöntemleri kullanılan diğer çalışmalara göre daha başarılı sonuçlar elde ettiği görülmüştür.

Candan vd. (2019), tarafından yapılan çalışmada Kanseri verileri üzerinde GA ile öznelik seçimi yapılarak 5 farklı makine öğrenmesi sınıflandırma algoritması üzerinde deneysel çalışmalar yapılmıştır. Elde edilen çalışmalar gözlemlendiğinde GA ile öznelik seçimi yapılırken başarımlarının arttığı görülmüştür. En başarılı sonucu destek vektör makineleri algoritmaları ile elde edilmiştir.

Vatansever vd. (2021), tarafından yürütülen çalışmada, Kaggle platformundan temin edilen kalp veri seti kullanılmıştır. Araştırmada söz konusu veri seti üzerinde KNN, LR, KA, RO, NB, DVM ve GA algoritmalarını içeren üç farklı grup deney tasarlanmış ve bu kapsamda toplam 28 deney gerçekleştirilmiştir. Elde edilen sınıflandırma performansları sırasıyla LR için %90,16; KNN için %81,96; DVM için %83,60; NB için %85,24; KA için %81,96 ve RO için %83,60 olarak raporlanmıştır. Hata oranları değerlendirildiğinde, en başarılı performansın GA tabanlı tahmin yöntemi ile %93,44 doğruluk oranına ulaşarak elde edildiği görülmektedir. GA yaklaşımıyla gerçekleştirilen tahminlerin diğer yöntemlere kıyasla daha yüksek doğruluk sağladığı gözlemlenmiştir. Bu bulgular, GA yaklaşımı ve meta-sezgisel (MÖ) yöntemlerde öznelik seçimi uygulanmasının, kalp hastalığının daha yüksek doğruluk ve başarımla öngörülebileceğini ortaya koymaktadır.

Şengür (2021), tarafından yapılan çalışmada Kolombiya'daki devlet okullarında görev yapan öğretmenlerden elde edilen ve 123 psikososyal öznelikten oluşan 4 sınıflı bir veri seti kullanılmıştır. Bu çalışmada Öznelik seçim yöntemi olarak Parçacık Sürü Optimizasyon algoritması kullanılmış ve Sınıflandırma işlemleri ise 5 katlı çapraz doğrulama yöntemiyle değerlendirilmiştir. Sınıflandırıcı olarak karar ağacı (KA), destek vektör makineleri (DVM), k-en yakın komşu (KNN) ve yapay sinir ağları (YSA) algoritmaları tercih edilmiştir. Elde edilen sonuçlar, en

yüksek doğruluk oranının DVM sınıflandırıcısı tarafından üretildiğini; en belirgin iyileşmenin ise KNN sınıflandırıcısında gözlemlendiği görülmüştür.

Ekrem vd. (2020), tarafından yapılan çalışmada kaggle internet sitesinden alınan kalp veri kümesi kullanılarak parçacık sürü optimizasyon algoritması ile öznitelik seçimi yapıp rastgele orman algoritması ile sınıflandırma işlemi gerçekleştirilmiştir. Sınıflandırma işleminin başarısı doğruluk,özgüllük,duyarlılık,kesinlik,f1-skor ve roc eğrisi ile ölçülmüştür. Elde edilen sonuçlar incelendiğinde Rastgele orman algoritmasının daha başarılı sonuçlar verdiği gözlemlenmiştir.

Deperlioğlu (2023), tarafından yapılan çalışmada çok yaygın kullanılan 5 veri kümesi üzerinde genetik algoritma tabanlı öznitelik seçimi uygulanmış ve elde edilen sonuçlar yedi ayrı sınıflandırma algoritması ile karşılaştırmalı olarak analiz edilmiştir. Elde edilen deneysel çalışmalar incelendiğinde genetik algoritma ile öznitelik seçimi sonucunda tüm veri kümelerinde başarımın arttığı gözlemlenmiştir.

Ciran ve Özbay (2024), tarafından yapılan çalışmada göz hastalıklarına ait 4 sınıftan oluşan veri kümesi üzerinde Denset201,DarkNet53 ve Resnet101 evrişimli sinir ağları ön eğitilmiş modelleri ile özellik çıkarımı yapılmış çıkarılan özellikler Parçacık sürü optimizasyon algoritması ile azaltılmıştır. Sonraki aşamada sınıflandırıcı yöntemi olarak destek vektör makineleri algoritması kullanılmış ve bayes yöntemi ile uygun hiperparametreler seçilerek sonuçlar karşılaştırılmıştır. Daha sonra 3 ön eğitilmiş model birleştirilerek önceki sonuçlarla karşılaştırılmıştır. Elde edilen sonuçlar incelendiğinde önerilen yöntemin daha başarılı sonuçlar verdiği gözlemlenmiştir.

Yağmur ve Yağmur (2022), tarafından gerçekleştirilen çalışmada, Türkiye’de bulunan bir rüzgâr türbinine ait 2018 yılı boyunca SCADA sistemi üzerinden elde edilen veri seti ile aynı konuma ilişkin NASA tarafından sunulan meteorolojik veri seti birlikte kullanılarak rüzgâr gücü tahmini yapılmıştır. Araştırmada beş farklı makine öğrenmesi yöntemi uygulanmış ve modeller üç ayrı değerlendirme metriği kullanılarak karşılaştırılmıştır. Öznitelik seçimi aşamasında genetik algoritma (GA) tercih edilmiştir. Çalışma kapsamında hem farklı makine öğrenmesi algoritmalarının tahmin performansları değerlendirilmiş hem de öznitelik seçiminin modele etkisi incelenmiştir. GA tabanlı nihai modelde değişken sayısı 47’den 9’a düşürülmüş, gereksiz öznitelikler modelden ayıklanmış ve yalnızca gerekli değişkenlerle R^2 değeri 0,98 olan yüksek doğrulukta bir tahmin modeli elde edilmiştir.

Tokmak (2023), tarafından yapılan çalışmada kötü amaçlı jar dosyalarının tespiti için rejaftada veri kümesi kullanılmıştır. ikili parçacık sürü optimizasyonu tabanlı öznitelik seçimi ile XGBoost algoritmasını bir araya getiren hibrit bir model önerilmiştir. Önerilen yöntem, 10 katlı çapraz doğrulama yaklaşımı kullanılarak doğruluk, kesinlik, F1 skoru ve duyarlılık ölçütleri üzerinden değerlendirilmiştir. Modelin performansını karşılaştırmak amacıyla AdaBoost, Gradient Boosting, Destek Vektör Makineleri, Yapay Sinir Ağları ve Naive Bayes algoritmaları ile ek analizler gerçekleştirilmiştir. Deneysel çalışmalar sonucunda önerilen yöntemin daha başarılı sonuçlar verdiği gözlemlenmiştir.

Harb ve Desuky (2014), tarafından yapılan çalışmada 3 farklı veri kümesi üzerinde Parçacık sürü optimizasyon algoritması filtre ve sarıcı yaklaşımlarla birleştirilerek öznitelik seçimi uygulanmış ve öznitelik seçimindeki başarısı genetik algoritma ile karşılaştırılmıştır. Veri kümeleri üzerinde 5 makine öğrenmesi algoritması uygulanmıştır. Elde edilen bulgular, Parçacık sürü optimizasyonu tabanlı öznitelik seçimi yaklaşımının diğer yöntemlere kıyasla daha üstün performans sergilediğini ortaya koymaktadır.

Sharawi vd. (2017), tarafından yapılan çalışmada UCI veri deposundan 16 farklı veri seti kullanılarak balina optimizasyon, genetik algoritma ve parçacık sürü optimizasyon yöntemleri verisetlerine uygulanarak öznitelik seçimi yapıp öznitelik seçim yöntemlerinin performansları karşılaştırılmıştır. Elde edilen sonuçlar incelendiğinde genel olarak balina optimizasyon yönteminin daha iyi performans sergilediği gözlemlenmiştir.

Vaishali vd. (2017), tarafından yapılan çalışmada pima kızıl derileri diyabet veri kümesi üzerinde genetik algoritma ile öznitelik seçimi yapılarak öznitelik seçiminin başarısı çok amaçlı evrimsel bulanık sınıflandırma algoritması kullanılarak başarıyı test edilmiştir. Çok amaçlı bulanık sınıflandırma modelinin başarısı naive bayes, J48 Karar ağacı aşısı, çok katmanlı algılayıcı algoritmaları ile karşılaştırılmıştır. Elde edilen sonuçlar incelendiğinde çok amaçlı bulanık sınıflandırma modelin %83,04'lük başarı oranına ulaşarak alternatif yöntemlere göre daha üstün performans sergilediği görülmüştür.

Shabbir vd. (2023), tarafından yapılan çalışmada Meme kanseri wiskonsin veri kümesi kullanılarak genetik algoritma ve PCA(Temel bileşen analizi) ile öznitelik seçimi yapılarak sonuçlar karşılaştırılmıştır. Makine öğrenmesi algoritmaları olarak Lojistik regresyon, karar ağacı ve rastgele orman algoritması

kullanılarak elde edilen sonuçlar karşılaştırılmıştır. Veri kümesinin orijinal hali, genetik algoritma uygulanarak öznitelik seçimi yapılmış hali ve temel bileşen analizi ile öznitelik seçimi yapılmış halleri kıyaslanmıştır. Rastgele orman algoritması üç veri kümesinin hepsinde diğer algoritmalara göre daha iyi performans göstermiştir. Öznitelik seçimi yöntemi olarak değerlendirildiğinde genetik algoritmanın daha iyi performans sergilediği görülmüştür.

Sakri vd. (2018), tarafından yapılan çalışmada Meme Kanseri Wisconsin Prognostik Veri kümesi üzerinde parçacık sürü optimizasyonu kullanılarak öznitelik seçimi gerçekleştirilerek naïve Bayes, REPTree ve KNN sınıflandırıcıları 10 katlı çapraz doğrulama yöntemi uygulanarak karşılaştırılmıştır. Elde edilen sonuçlar incelendiğinde parçacık sürü optimizasyon algoritması naïve bayes algoritması ile kullanıldığında veri kümesinin orijinal durumuna göre daha başarısız sonuçlar verirken, diğer iki algoritma ile öznitelik seçimi uygulandığında başarı oranının arttığı gözlemlenmiştir.

Sulaiman vd. (2024), tarafından yapılan çalışmada wind power forecasting veri kümesi kullanılarak metasezgisel algoritmalarla öznitelik seçimi yapılarak yapay sinir ağları ile entegre edilmiştir. Genetik algoritma, parçacık sürüsü optimizasyonu, karınca kolonisi optimizasyonu, öğretim-öğrenme tabanlı optimizasyon ve evrimsel çiftleşme algoritması yöntemleri ile 18 farklı öznitelikten oluşan veri kümesinden optimum öznitelikler seçildikten sonra YSA ile entegre edilerek performansları karşılaştırılmıştır. Elde edilen sonuçlar gözlemlendiğinde genetik algoritma 7,1837'lik en iyi ortalama karekök hatası (RMSE) ve 18,6313 değerindeki en düşük ortalama mutlak hata (MAE) sonucunun, yöntemin diğer yaklaşımlara kıyasla daha başarılı bir performans ortaya koyduğunu göstermektedir.

Alickovic ve Subasi (2017), tarafından yapılan çalışmada iki farklı Wisconsin Meme Kanseri veri kümesi üzerinde genetik algoritma ile öznitelik seçimi yapılarak Lojistik Regresyon, Karar Ağaçları, Rastgele Orman, Bayes Ağı, Çok Katmanlı Algılayıcı, Radyal Bazlı Fonksiyon Ağları, Destek Vektör Makinesi ve Rotasyon Orman algoritmalarının performansları karşılaştırılmıştır. Sınıflandırma performansları ROC eğrisi ve F1-skor ile ölçülmüştür. Elde edilen sonuçlar incelendiğinde Genetik algoritma tabanlı rotasyon orman algoritması ile öznitelik sayısını 32 den 14 e düşürerek %99,48 ile en başarılı sonucu verdiği gözlemlenmiştir.

Nemati vd. (2024), tarafından yapılan çalışmada anomali tespiti amacıyla

öznitelik seçimi için metasezgisel algoritmalar ve veri madenciliği tabanlı bir yaklaşım önerilmiştir. Çalışmada, farklı veri setleri üzerinde yüksek boyutlu özniteliklerin azaltılması için parçacık sürü optimizasyonu (PSO), genetik algoritma (GA) ve benzetilmiş tavlama (SA) gibi metasezgisel algoritmalar kullanılmıştır. Öznitelik seçimi sonrasında, elde edilen boyutu azaltılmış veri kümeleri üzerinde 3 farklı makine öğrenmesi modeli uygulanarak model performansları 4 farklı metrik ile değerlendirilmiştir. Elde edilen sonuçlar incelendiğinde PSO+SVM kombinasyonundan elde edilen modelin daha başarılı sonuçlar verdiği görülmüştür.

2.2. Geleneksel Yöntemlere Dayalı Çalışmalar

Bu bölümde geleneksel makine öğrenmesi yöntemleri kapsamında gerçekleştirilen öznitelik seçimi çalışmalarına ilişkin yapılan çalışmalardan bahsedilmiştir.

Emanet vd. (2021), tarafından yürütülen çalışmada, CSE-CIC-IDS2018 veri kümesi üzerinde farklı öznitelik seçimi yöntemlerinin siber saldırı tespitindeki etkisi değerlendirilmiştir. Öznitelik seçimi aşamasında Ki-Kare Testi, Spearman'ın Sıralama Korelasyon Katsayısı ve Özyinelemeli Öznitelik Eliminasyonu yöntemleri uygulanmıştır. Sınıflandırma sürecinde ise Adaptif Yükseltme, Karar Ağacı, Lojistik Regresyon, Çok Katmanlı Algılayıcı, Ekstra Ağaçlar, Pasif-Agresif ve Gradyan Artırma algoritmalarının performansları k-fold çapraz doğrulama yöntemi kullanılarak ölçülmüştür. Bulgular, Ki-Kare Testi ve Spearman'ın Sıralama Korelasyon Katsayısı yöntemlerinin veri boyutunu küçültürken işlem süresini %45 oranında azaltmasına karşın hata oranını sırasıyla %14,46 ve %10,52 artırdığını göstermektedir. Özyinelemeli Öznitelik Eliminasyonu uygulandığında ise işlem süresinin %38 oranında azaldığı ve hata oranının %2,95 düzeyinde düşürüldüğü görülmüştür.

Emhan ve Akın (2019), tarafından yapılan çalışmada NSLKDD veri kümesi kullanılarak öznitelik seçim yöntemlerinin saldırı tespit sistemleri üzerindeki etkisi ölçülmüştür. Çalışmada öznitelik sayısını azaltmak amacıyla bilgi kazancı, kazanç oranı, simetrik belirsizlik katsayısı, ki-kare, One-R ve korelasyon tabanlı öznitelik seçimi yöntemleri uygulanarak veri kümesindeki öznitelikler indirgenmiştir. Öznitelik seçimi yapıldıktan sonra sınıflandırma performansları k-en yakın komşu ve rastgele orman algoritmaları kullanılarak ölçülmüştür. Elde edilen sonuçlar gözlemlendiğinde sınıflandırıcı performansı olarak rastgele orman algoritmasının daha başarılı olduğu görülmüştür. Öznitelik seçimi yöntemi olarak bilgi kazancı, ki-kare, One-R yöntemlerinin rastgele orman algoritması ile başarılı sonuçlar verdiği

gözlemlenmiştir.

Bilen ve Özer (2022), tarafından gerçekleştirilen çalışmada, Elazığ ilinde meydana gelen suç olaylarına ilişkin veri kümesi kullanılmış ve öznitelik sayısını azaltmak amacıyla tek değişkenli öznitelik seçimi, özyinelemeli öznitelik eleme, ağaç tabanlı öznitelik seçimi ve temel bileşen analizi yöntemleri uygulanmıştır. Sınıflandırma sürecinde rastgele orman algoritması tercih edilmiştir. Öznitelik seçim yöntemleri aracılığıyla belirlenen en önemli 13 öznitelik arasından farklı alt kümeler oluşturulmuş; 6, 5 ve 4 öznitelikten oluşan bu alt kümelerin rastgele orman algoritması üzerindeki etkileri karşılaştırmalı olarak analiz edilmiştir. Bulgular, öznitelik sayısının 4'e düşürüldüğü durumda %97,24 doğruluk oranı ile en yüksek performansın elde edildiğini göstermiştir.

Kumar ve Shaikh (2017), tarafından yapılan çalışmada tıbbi verilerden oluşan Cleveland veri kümesi üzerinde öznitelik seçim yöntemleri kullanılarak relief, rastgele orman, özyinelemeli özellik seçme ve boruta öznitelik seçme yöntemleri ile öznitelik sayısı indirgenmiştir. Elde edilen sonuçlar gözlemlendiğinde boruta öznitelik seçimi yöntemi ile %98 doğruluk oranı ile en başarılı sonuçları verdiği gözlemlenmiştir.

Osanaie vd. (2019), tarafından yapılan çalışmada 41 öznitelikten oluşan NSL-KDD veri kümesi üzerinde öznitelik seçimi işlemi yapılarak öznitelik sayısı indirgenmiştir. Çalışma kapsamında öznitelik seçimi için ki-kare, relief ve kazanç oranı filtre yöntemleri birleştirilerek yeni bir öznitelik seçim yöntemi ortaya atılmıştır. Sınıflandırıcı olarak rastgele orman algoritması kullanılmıştır. Elde edilen sonuçlar gözlemlendiğinde öznitelik sayısı 14'e indirgenmiş ve başarı oranının arttığı görülmüştür.

Yılmaz ve Sümer (2021), tarafından yapılan çalışmada kalp hastalığı veri kümesinde relief özellik seçimi yöntemi kullanılarak lojistik regresyon, karar ağacı, yapay sinir ağları , k-en yakın komşu, naif bayes ve destek vektör makineleri sınıflandırıcıları kullanılarak sonuçlar relief tabanlı hibrit modelle karşılaştırılmıştır. Önerilen hibrit yöntemde reilef öznitelik seçim yöntemi destek vektör makinesi algoritması ile birleştirilmiştir. Elde edilen sonuçlar gözlemlendiğinde hibrit tabanlı yöntemin %98'lik doğruluk oranı ile diğer yöntemlere kıyasla daha iyi sonuçlar verdiği gözlemlenmiştir.

Kayaalp vd. (2018), tarafından yapılan çalışmada UCI veri deposundan kronik böbrek hastalığı veri kümesi kullanılarak relief ve kazanç oranı öznitelik

seçim yöntemleri kullanılarak öznitelik sayısı indirgenmiştir. Çalışmada sınıflandırıcı olarak bayes sınıflandırma algoritması, k-en yakın komşu (k-NN) algoritması ve destek vektör makineleri kullanılmış; bu yöntemlerle elde edilen sonuçlar karşılaştırmalı olarak değerlendirilmiştir. Bulgular, relief öznitelik seçim yöntemi uygulandıktan sonra k-en yakın komşu algoritması ile gerçekleştirilen sınıflandırmanın diğer yöntemlere kıyasla daha yüksek başarı sağladığı görülmüştür.

Atalay vd. (2018), tarafından yapılan çalışmada RAVDESS veri kümesi üzerinde otokodlayıcı, relief ve ki kare öznitelik seçim yöntemleri kullanılarak destek vektör makineleri sınıflandırıcısı ile sonuçlar karşılaştırılmıştır. Elde edilen sonuçlar karşılaştırıldığında relief öznitelik seçim yönteminin daha başarılı sonuçlar verdiği gözlemlenmiştir.

Kaynar vd. (2018), tarafından gerçekleştirilen çalışmada KDDCup99 veri kümesi üzerinde ki-kare, bilgi kazancı, kazanım Oranı, gini katsayısı, OneR, ReliefF, genetik algoritma, ileriye doğru seçme ve geriye doğru eleme gibi çeşitli öznitelik seçim yöntemleri uygulanmış ve bu yöntemlerle elde edilen sonuçlar karşılaştırmalı olarak analiz edilmiştir. Çalışma kapsamında sınıflandırıcı olarak k-en yakın komşu, destek vektör makineleri ve aşırı öğrenme algoritmaları kullanılmıştır. Yapılan deneysel çalışmalar sonucunda öznitelik seçim yöntemlerinin 3 sınıflandırma yöntemi içinde başarı oranını arttırdığı ve çalışma süresini azalttığı gözlemlenmiştir.

Gümüşçü vd. (2016), tarafından yapılan çalışmada 11 farklı mikro-dizilim veri kümesi üzerinde ki-kare, ReliefF, f-skör, korelasyon tabanlı öznitelik seçimi ve bilgi kazancı yöntemleri uygulanmıştır. Öznitelik seçiminden elde edilen alt kümeler, 10 katlı çapraz doğrulama yaklaşımı kullanılarak k-en yakın komşu (k-NN) sınıflandırıcısı ile değerlendirilmiş ve yöntemlerin performansları karşılaştırmalı olarak analiz edilmiştir. Yapılan deneysel çalışmalar sonucunda korelasyon-bazlı öznitelik seçimi yöntemi sınıflandırıcı performansı olarak en iyi sonuçları verirken, çalışma süresi bakımından ki-kare yöntemi ile daha başarılı sonuçlar elde edilmiştir.

Sağbaş. (2023), tarafından gerçekleştirilen çalışmada metin tabanlı dokuz farklı duygu analizi veri kümesi kullanılmıştır. Bu veri kümeleri üzerinde korelasyon tabanlı öznitelik seçimi, ki-kare, kazanç oranı, bilgi kazancı, OneR ve simetrik belirsizlik katsayısı yöntemleri uygulanmış ve elde edilen sonuçlar karşılaştırmalı olarak değerlendirilmiştir. Çalışma kapsamında sınıflandırıcı olarak çok terimli naive bayes algoritması kullanılmıştır. Elde edilen sonuçlar incelendiğinde 9 duygu analizi veri kümesi için ortalama %94,34 doğruluk oranlarına ulaşıldığı görülmüştür.

Hasan vd. (2025), tarafından yapılan çalışmada kötü amaçlı yazılımın tespiti için öznitelik seçimi, veri ölçekleme ve makine öğrenmesi modellerinin etkisi değerlendirilmiştir. Çalışmada, 11.598 örnek ve 139 öznitelik içeren kamuya açık bir ikili tablo sınıflandırma veri kümesi kullanılmıştır. Metodolojide, üç farklı öznitelik ölçekleme tekniği (ölçekleme yok, normalizasyon, min-maks ölçekleme) ve üç öznitelik seçme yöntemi (seçim yok, doğrusal ayırım analizi (LDA), temel bileşen analizi (PCA)) ile birlikte, geleneksel ve topluluk tabanlı toplam on iki makine öğrenimi modeliyle deneyler yapılmıştır. Model performansı 5 farklı metrik ile değerlendirilmiştir. Elde edilen sonuçlar incelendiğinde PCA ve min-max ölçekleme veya normalizasyon Lightgbm algoritmasına uygulandığında modellerin kombinasyonuna göre en başarılı sonuçları verdiği gözlemlenmiştir.

Bulut vd. (2023), tarafından yapılan çalışmada EEG tabanlı beyin bilgisayar arayüzü sistemi için 10 farklı filtre tabanlı öznitelik seçimi yöntemi 4 farklı sınıflandırma modeli uygulanarak sonuçlar değerlendirilmiştir. Çalışma kapsamında EEG kayıtları dört aşamalı bir analizden geçirilerek 50 öznitelikten oluşan yeni bir veri seti oluşturulmuştur. Elde edilen veri setine öznitelik seçim yöntemleri uygulanarak sınıflandırma sonuçları incelenmiştir. Elde edilen sonuçlar göz önüne alındığında öznitelik seçimi yapılmamış durumuna kıyasla filtre tabanlı öznitelik seçim yöntemlerinin uygulanması sonucunda başarımlar önemli ölçüde artmıştır. Eigenvector Centrality ve Infinite Feature Selection yöntemlerinin tüm sınıflandırıcılarda en başarılı sonuçlara ulaştığı gözlemlenmiştir.

Emmanuel vd. (2024), tarafından yapılan çalışmada kredi risk tahmini için rastgele orman, gradient boosting ve extreme gradient boosting algoritmaları birleştirilip kredi riski tahmini için yeni bir model önerilmiştir. Önerilen yöntem 6 farklı MÖ modeli ile bilgi kazanımı teorisine dayanan filtre tabanlı öznitelik seçim yöntemi 3 farklı kredi riski verilerinden oluşan veri kümesine uygulanarak karşılaştırılmıştır. Model performansları 3 farklı metrik ile değerlendirilmiştir. Elde edilen sonuçlar gözlemlendiğinde önerilen yöntemin diğer yöntemlere kıyasla daha iyi performans sergilediği görülmüştür.

Farsi. (2021), tarafından yapılan çalışmada 4 farklı kanser veri kümesi kullanılarak Lojistik Regresyon, Ki-kare , rastgele orman ve LightGBM yöntemleri öznitelik seçimi yapılması amacıyla kullanılmıştır. Azaltılan öznitelik kümeleri svm, knn, karar ağacı, lojistik regresyon ve rastgele orman sınıflandırıcıları kullanılarak karşılaştırılmıştır. Kullanılan modellerin performansları 5 farklı değerlendirme metriği kullanılarak ölçülmüştür. Elde edilen sonuçlar incelendiğinde öznitelik

seçimi uygulandıktan sonra genel olarak başarımlar oranlarının arttığı ve çalışma sürelerinin azaldığı gözlemlenmiştir.

3. GEREÇ VE YÖNTEM

Bu bölümde tez kapsamında kullanılan veri kümeleri, kullanılan makine öğrenmesi algoritmaları ,kullanılan öznitelik seçim yöntemleri hakkında detaylı bilgi verilmiştir.

Tez kapsamında kullanılan yöntemler iki farklı çalışma için bağımsız bir şekilde alt başlıklar halinde sunulmuştur.

3.1. Geleneksel Makine Öğrenmesi ve Metasezgisel Yöntemler Kullanılarak Yapılan Çalışma

3.1.1. Kullanılan Veri kümeleri

Bu çalışmada, deneylerin yürütülmesi amacıyla UCI Machine Learning Repository’den temin edilmiş üç farklı veri kümesi kullanılmıştır. Kullanılan veri kümelerine ilişkin temel bilgiler Çizelge 3.1’de özetlenmiştir. Seçilen her bir veri kümesi, sınıflandırma problemlerine uygun yapısı ve alanlarında sağladığı temsil gücü nedeniyle çalışmaya dahil edilmiştir; böylece önerilen metodolojinin genelleştirilebilirliği ve performans tutarlılığı farklı veri dağılımları ve boyutlarda sınanmıştır (UCI MACHINE LEARNING REPOSITORY, 2025).

Sonar: Sonar veri kümesi 61 öznitelikten ve 208 gözlemden oluşmaktadır. Bu veri kümesi, sonar sinyallerinden elde edilen ölçümler kullanılarak denizaltı kaynaklı nesnelerin sınıflandırılmasını hedeflemektedir. Veri kümesindeki örnekler iki sınıfa ayrılmış olup, sınıflandırma görevi kaya ve maden cisim ayırımını gerçekleştirmektedir. Özellik sayısının göreceli olarak yüksek olması, boyutsal yapı ve olası çoklu korelasyonları nedeniyle öznitelik seçimi ve model düzenleme tekniklerinin önemini artırmaktadır (Gorman ve Sejnowski, 1988).

Parkinson: Parkinson veri kümesi 23 öznitelikten oluşmakta olup toplam 195 gözlem içermektedir. Parkinson hastalığı, beyinde dopamin üreten nöronların azalması sonucunda gelişen nörodejeneratif bir bozukluktur ve erken tanı hastalık yönetimi açısından kritik öneme sahiptir. Veri setinde yer alan 23 özellik, hastalığın teşhisini kolaylaştırmak üzere hasta ve sağlıklı bireylerden elde edilen ölçümler ve biyometrik göstergeleri yansıtmaktadır. Veri kümesi hasta ve sağlıklı olmak üzere iki sınıflı etiketten oluşmaktadır (Little vd., 2007).

Breast Cancer: Breast cancer veri kümesi 32 öznitelikten ve 569 gözlemden oluşmaktadır. Meme kanseri, özellikle belirli yaştan sonra meme dokusundaki

hücrel proliferasyonun kontrolsüz hale gelmesiyle karakterize edilen önemli bir onkolojik hastalıktır. Veri kümesi meme kanseri hastalığını etkileyen 31 tane özellikten ve hedef değişken olarak iki sınıflı etiketten oluşmaktadır (Wolberg ve Mangasarian, 1990).

Çizelge 3.1. Kullanılan veri kümeleri

Veri kümeleri	Öznitelik Sayısı	Gözlem Sayısı
Sonar	61	208
Parkinson	23	195
Breast Cancer	32	569

3.1.2. Sınıflandırma Yöntemleri

Bu bölümde tez kapsamında öznitelik seçim yöntemlerinin sınıflandırıcı performanslarını ölçmek için kullanılan ağaç tabanlı yöntemler olan karar ağacı (KA), rastgele orman (RO), gradyan artırma makineleri (GBM), hafif gradyan artırma (Lightgbm) ve aşırı gradyan artırma (Xgboost) algoritmaları hakkında detaylı bilgi verilmiştir.

3.1.2.1. Karar Ağacı

Karar ağaçları algoritması, kural tabanlı bir makine öğrenmesi yaklaşımı olup hem sınıflandırma hem de regresyon problemlerinin çözümünde kullanılmaktadır. Algoritma veriyi kök düğümden başlayıp dal ve yaprak düğümlere doğru hiyerarşik bir biçimde bölme işlemini gerçekleştirir. Bu bölünme işleminde amaç, verinin alt kümelerinin homojenliğini artırmaktır. Her bir iç düğüm, belirli bir özniteliğe göre veri setini iki veya daha fazla alt kümeye ayırırken; yaprak düğümler sınıf etiketlerini veya tahmin değerlerini temsil eder (Charbuty ve Abdulazeez, 2021).



Şekil 3.1. Karar Ağacı Algoritmasının Bölünmesi

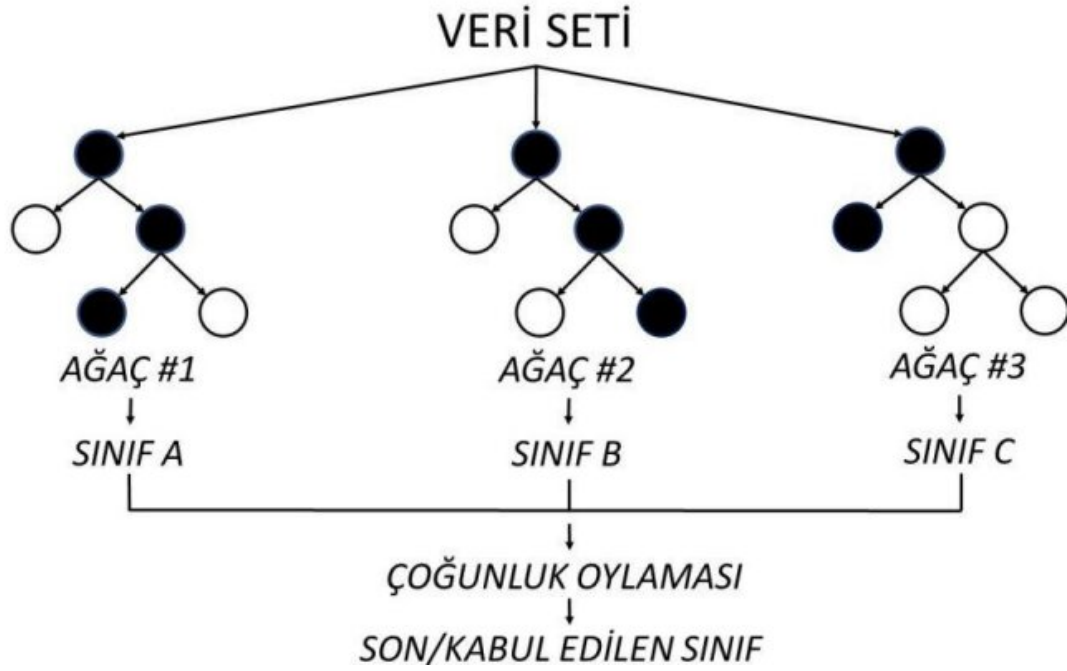
Karar ağaçlarında bölünme kararının hangi öznelik üzerinden yapılacağı, bilgi kazancı (information gain), Gini saflık ölçütü (Gini impurity) veya ki-kare istatistiği gibi kriterler kullanılarak belirlenir. Bu ölçütler, düğüm bölünmesinin veri kümesindeki belirsizliği ne ölçüde azalttığını hesaplar. Örneğin, bilgi kazancı ölçütü entropi temelli bir değerlendirme yaparak, en yüksek bilgi kazancını sağlayan özneliği bölünme noktası olarak seçer (Charbuty ve Abdulazeez, 2021). Karar ağaçları, eğitim süresinin görece kısa olması, veri normalizasyonu veya ölçeklendirme gerektirmemesi, kategorik ve sayısal verilerle doğrudan çalışabilmesi gibi avantajlara sahiptir. Ayrıca, modelin yapısı ağaç biçiminde görselleştirilebildiği için karar verme süreci şeffaf ve yorumlanabilir niteliktedir. Ancak, bu yöntem aşırı öğrenme problemine yatkındır. Aşırı öğrenme, modelin eğitim verisine fazla uyum sağlayarak yeni verilerde düşük genelleme performansı göstermesi durumudur. Bu problemi azaltmak amacıyla budama teknikleri, maksimum derinlik sınırlaması, minimum yaprak düğüm örnek sayısı belirleme gibi yöntemler kullanılmaktadır. Ayrıca, karar ağaçlarının tek başına kullanılmasından kaynaklanan bu dezavantaj, Rastgele Orman (Random Forest) ve Gradient Boosting gibi çok sayıda ağacın bir arada kullanıldığı topluluk (ensemble) öğrenme yöntemleri ile önemli ölçüde giderilebilmektedir.

3.1.2.2. Rastgele Orman Algoritması

Rastgele orman algoritması, topluluk öğrenmesi yaklaşımına dayanan ve hem sınıflandırma hem de regresyon problemlerinde yaygın biçimde kullanılan güçlü bir makine öğrenmesi yöntemidir. Rastgele orman algoritması aşırı öğrenme probleminin önüne geçilmesi için çok sayıda karar ağacının çıktılarını birleştirilerek daha genellenebilir sonuçlar vermektedir (Breiman , 2001). Rastgele Orman algoritmasının temeli, bagging metodolojisine dayanmaktadır. Bagging, eğitim veri

kümesinden tekrar seçmeli örnekleme yöntemi ile farklı alt kümeler oluşturur ve her alt kümede ayrı bir karar ağacı eğitir. Ağaçlar paralel bir şekilde ayrı ayrı eğitilir ve sonunda çıktıları birleştirilir.

Rastgele orman, bu yönteme ek olarak her ağacın düğüm bölünmelerinde kullanılacak özniteliklerin rastgele bir alt kümesinden seçim yapılmasını sağlar. Böylece ağaçlar arasında çeşitlilik artırılır ve aşırı öğrenme riski azaltılır (Salman vd., 2024).



Şekil 3.2. Rastgele Orman Algoritmasının Çalışma Mantığı

Rastgele orman algoritmasının çalışma adımları şu şekilde özetlenebilir:

- Eğitim veri kümesinden, tekrar seçmeli örnekleme yöntemi ile farklı alt kümelerin oluşturulması,
- Her alt kümeden bir karar ağacının oluşturulması,
- Ağaçların çoğunluk oylaması yapılarak tahminlerinin birleştirilmesi adımlarından oluşur.

Rastgele orman algoritmasının başarılı sonuçlar vermesi için ağaç sayısı ($n_{\text{estimators}}$), maksimum öznitelik sayısı (max_features), ve maksimum derinlik (max_dept) hiperparametrelerinin uygun bir şekilde ayarlanması gerekir.

3.1.2.3. Gradyan Artırma Algoritması

Gradyan artırma algoritması, sıralı öğrenme yaklaşımını temel alan ve hem sınıflandırma hem de regresyon problemlerinde etkili sonuçlar üreten güçlü bir topluluk öğrenme yöntemidir. Algoritma, zayıf öğrenicileri (genellikle sığ karar ağaçları) sıralı olarak birleştirerek, her adımda bir önceki modelin hatalarını düzeltmeye odaklanır ve böylece güçlü bir tahmin modeli oluşturur (Friedman, 2001).

Gradyan Artırma'nın temeli, boosting metodolojisine dayanmaktadır. Boosting, zayıf öğrenicileri sıralı bir şekilde eğiterek her yeni modelin bir önceki modelin yaptığı hataları düzeltmesini sağlar. Rastgele orman'dan farklı olarak, modeller paralel değil ardışık olarak eğitilir. Her iterasyonda, mevcut modelin tahmin hatalarına göre hesaplanan gradyanlar kullanılarak yeni bir zayıf öğrenici eklenir. Bu süreç, belirlenen iterasyon sayısına ulaşılan veya hata belirli bir eşiğin altına düşene kadar devam eder (Chen ve Guestrin, 2016).

Gradyan Artırma algoritmasının çalışma adımları şu şekilde özetlenebilir:

- a. ilk tahmin modelinin (genellikle sabit bir değer) oluşturulması,
- b. Her iterasyonda mevcut modelin hatalarının hesaplanması ve gradyanların belirlenmesi,
- c. Hatalara odaklanan yeni bir zayıf öğrenicinin eğitilmesi,
- d. Yeni modelin mevcut topluluk modeline eklenmesi,
- e. Belirlenen durdurma kriterine ulaşılan kadar bu sürecin tekrarlanması.

Gradyan Artırma algoritmasının yüksek performans göstermesi, öğrenme oranı (learning_rate), tahmin edici sayısı (n_estimators), maksimum ağaç derinliği (max_depth) ve alt örnekleme oranı (subsample) gibi hiperparametrelerin uygun biçimde ayarlanmasına bağlıdır.

3.1.2.4. Hafif Gradyan Artırma Algoritması

Hafif gradyan artırma algoritması, gradyan artırma algoritması üzerinde inşa edilmiş , özellikle büyük ölçekli veri kümeleri ve öznitelik sayısının çok olduğu durumlarda hızlı ve verimli çalışmasıyla öne çıkan bir makine öğrenmesi yöntemidir. LightGBM, temel olarak ağaç tabanlı öğrenme yapısını kullanır ve Gradyan

Artırmalı Karar Ağaçları (GBDT – Gradient Boosted Decision Trees) algoritmasının optimize edilmiş bir versiyonudur (Ke vd., 2017).

LightGBM'in en önemli farkı, histogram tabanlı karar ağaçları ve yaprak odaklı büyüme stratejisi kullanmasıdır. Histogram tabanlı yaklaşım, sürekli öznitelikleri belirli aralıklara ayırarak veri boyutunu küçültür, bellek kullanımını azaltır ve işlem hızını artırır. Yaprak odaklı strateji ise, ağaç büyümesini seviye bazlı değil en yüksek bilgi kazancı sağlayan yaprak üzerinden derinlemesine genişletir. Bu sayede model doğruluğu artırılır; ancak dikkat edilmezse aşırı öğrenme riski ortaya çıkabilmektedir (Ke vd., 2017).

LightGBM Algoritmasının Çalışma Adımları şu şekilde özetlenebilir:

- a. Veri, histogram tabanlı yapılara dönüştürülür.
- b. İlk model oluşturulduktan sonra her iterasyonda gradyan hataları hesaplanır.
- c. En fazla bilgi kazancı sağlayan yaprak seçilir ve bölünür.
- d. Yeni dallar topluluk modeline eklenir.
- e. Belirlenen durdurma kriterine ulaşılan kadar süreç devam eder.

3.1.2.5. Aşırı Gradyan Artırma Algoritması

Aşırı gradyan artırma algoritması, gradyan artırma algoritmasının ölçeklenebilir ve yüksek performanslı bir versiyonudur. Özellikle büyük veri kümelerinde hızlı ve doğru sonuçlar vermesinden dolayı günümüzde sıkça kullanılan bir topluluk öğrenme yöntemi algoritmasıdır. Xgboost algoritmasının temelinde boosting metodolojisi yöntemi yer almaktadır. Xgboost, zayıf öğrenicilerin ardışık olarak eğitilmesi prensibine göre çalışmaktadır. Ancak klasik gradyan artırma yöntemlerinden farklı olarak Xgboost, ikinci dereceden gradyan bilgilerini de kullanarak hata fonksiyonunun daha hassas şekilde optimize edilmesini sağlar. Ayrıca aşırı öğrenmenin önüne geçmek amacıyla L1 (Lasso) ve L2 (Ridge) düzenleme terimlerini içeren düzenlenmiş bir hedef fonksiyonu kullanır (Chen ve Guestrin, 2017).

XGBoost algoritması, özellikle geniş ölçekli veri kümelerinde yüksek doğruluk ve hız açısından oldukça etkili bir performans sergilemektedir. Bu yöntem,

topluluk öğrenme tekniklerinin sağladığı avantajlar sayesinde, genellikle aşırı öğrenme (overfitting) sorununa karşı da dayanıklı bir yapı sunmaktadır. Bununla birlikte, algoritmanın başarısı büyük ölçüde doğru hiperparametre seçimine bağlıdır. Hiperparametrelerin sayıca fazla olması ve bu parametrelerin birbirleriyle etkileşimli çalışması, modelin ayarlanma sürecini zaman alıcı ve karmaşık hale getirebilmektedir.

3.1.3. Kullanılan Öznitelik Seçimi Yöntemleri

Bu bölümde tez kapsamında ele alınan geleneksel makine öğrenmesi yöntemleri ve metasezgisel yöntemleri ile öznitelik seçimi hakkında bilgi verilmiştir. Geleneksel makine öğrenmesi yöntemlerinden korelasyon analizi, hipotez testi, lasso yöntemleri ve metasezgisel yöntemlerden parçacık sürü optimizasyon ve genetik algoritma hakkında bilgi verilmiştir.

3.1.3.1. Geleneksel Makine Öğrenmesine Dayalı Öznitelik Seçimi Yöntemleri

3.1.3.1.1. Korelasyon Analizi

Korelasyon analizi, iki değişken arasındaki doğrusal ilişkinin yönünü ve şiddetini belirlemeye yönelik bir istatistiksel yöntemdir. Analiz sonucunda elde edilen korelasyon katsayısı -1 ile $+1$ arasında değişmekte olup, katsayının $+1$ 'e yakın olması değişkenler arasında güçlü ve pozitif bir ilişkiyi, -1 'e yakın olması ise güçlü ve negatif bir ilişkiyi ifade etmektedir.

Katsayının 0 'a yakın olduğu durumlarda ise değişkenler arasında anlamlı bir doğrusal ilişki bulunmadığı sonucuna varılır. Bu yöntem, özellikle makine öğrenmesi ve veri madenciliği çalışmalarında öznitelik seçimi amacıyla yaygın olarak kullanılmaktadır. Korelasyon analizi sayesinde birbirine yüksek derecede bağlı öznitelikler tespit edilip modelden çıkarılabilir, böylece hem modelin daha yalın hem de daha yorumlanabilir hale gelmesi sağlanır. Ayrıca korelasyon analizinde Pearson, Spearman ve Kendall gibi farklı ölçüm yöntemleri de kullanılabilir; yöntem seçimi veri tipine ve dağılım özelliklerine bağlı olarak değişmektedir (Chen ve Wasikowski, 2008).

3.1.3.1.2. Hipotez Testi

Hipotez testi, özniteliklerin modeldeki katkı düzeyini ve istatistiksel açıdan anlamlılığını değerlendirmek için kullanılan temel bir yöntemdir. İstatistiksel hipotez testinde genellikle sıfır hipotezi (H_0) ve alternatif hipotez (H_1) oluşturulur. Sıfır hipotezi, öznitelik katsayılarının modele anlamlı katkı sağlamadığını, alternatif

hipotez ise özniteliklerin anlamlı olduğunu ifade eder. Bu testler sırasında p-değeri hesaplanır ve bu değer, önceden belirlenen anlamlılık düzeyi (genellikle 0,01 veya 0,05) ile karşılaştırılır. Eğer p-değeri anlamlılık düzeyinden küçükse H_0 reddedilir ve ilgili öznitelik model için istatistiksel olarak anlamlı kabul edilir. Aksi durumda öznitelik modelden çıkarılabilir. Bu yaklaşım, özellikle çok sayıda değişken içeren veri kümelerinde modelin gereksiz özniteliklerden arındırılması ve yalnızca anlamlı olanların tutulması açısından önem taşımaktadır. Böylece modelin hem performansı hem de genelleme gücü artırılmış olur (Kartal, 1998).

3.1.3.1.3. Lasso Yöntemi

Lasso yöntemi, regresyon temelli modellerde aşırı öğrenmeyi önlemek ve öznitelik seçimi yapmak amacıyla geliştirilmiş bir düzenleme tekniğidir. Bu yöntem, model katsayılarına L1 normu cezası ekleyerek bazı katsayıların değerlerini sıfıra kadar indirir. Böylece gereksiz veya düşük etkili öznitelikler modelden otomatik olarak çıkarılmış olur. Lasso'nun en önemli avantajı, hem değişken seçim sürecini gerçekleştirmesi hem de modelin daha sade ve yorumlanabilir hale gelmesini sağlamasıdır. Ayrıca, çok sayıda öznitelik içeren yüksek boyutlu veri kümelerinde Lasso, klasik regresyon yöntemlerine göre daha etkili sonuçlar verebilmektedir. Bununla birlikte, yüksek korelasyonlu öznitelikler arasında seçim yaparken yalnızca birini modelde tutma eğiliminde olduğu için, bu durum araştırmacının dikkat etmesi gereken bir noktadır. Lasso yöntemi günümüzde biyoinformatik, finans, sosyal bilimler ve makine öğrenmesi uygulamaları gibi birçok alanda öznitelik seçimi ve aşırı öğrenmenin önlenmesi amacıyla yaygın olarak kullanılmaktadır (Ranstam ve Cook, 2018).

3.1.3.2. Metasezgisel Yöntemlere Dayalı Öznitelik Seçimi Yöntemleri

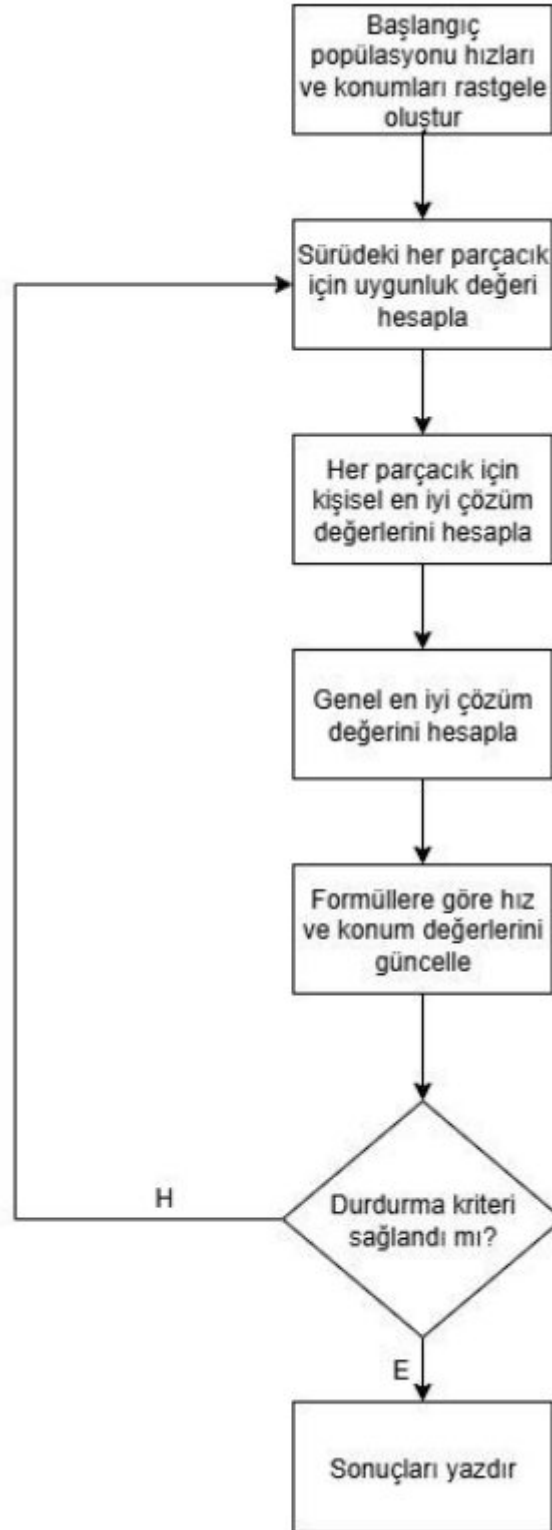
3.1.3.2.1. Parçacık Sürü Optimizasyon Algoritması

Parçacık sürü optimizasyon algoritması (PSO), kuş ve balık gibi canlıların sürü hâlinde hareket etme davranışlarından esinlenerek Kennedy ve Eberhart tarafından 1995 yılında geliştirilen bir optimizasyon algoritmasıdır. Kuş ve balık gibi birçok hayvan sürülerinin yiyecek bulma ve avcılardan gelebilecek tehlikelerden kaçınmalarını taklit eden sürü zekası temelli bir yöntemdir (Kennedy ve Eberhart, 1995).

Doğadan esinlenen bu yöntem, hayvanların toplu halde hareket ederken sergiledikleri davranışların incelenmesine dayanmaktadır. Çoğu zaman rastlantısal gibi görünen bu hareketler, yiyecek bulma, korunma veya yön belirleme gibi

amaçlara daha kısa sürede ulaşılmasını sağlamaktadır. Bu davranış biçiminden esinlenen PSO algoritması, bireyler arasındaki sosyal bilgi paylaşımını temel alır ve problemi çözmek için bu kolektif etkileşimden yararlanır. Algoritmada her birey parçacık olarak adlandırılır; parçacıklarının tümünün bir bütün olarak birleşimi ise sürü olarak adlandırılmaktadır. Böylelikle parçacıklar, sürünün keşfettiği en iyi pozisyona yaklaşma eğilimi gösterir. Hareket hızı ve yönelimler belirli ölçüde rastgelelik içerdiği için, parçacıklar genellikle önceki konumlarına kıyasla daha iyi konumlara ilerler. Bu süreç, hedeflenen sonuca ulaşılan kadar tekrarlanır (Özsağlam ve Çunkaş, 2008).

PSO algoritması, genetik algoritmalara benzer biçimde aramayı iterasyonlar boyunca sürdürür. Ancak, genetik algoritmalarından farklı olarak çaprazlama ve mutasyon gibi işlemler yerine bireyler arasındaki bilgi aktarımı ve uyum mekanizmasından yararlanır. Geniş bir uygulama alanına sahip olan PSO, sipariş miktarının belirlenmesi, çizelgeleme problemleri, enerji sistemlerinde güç ve voltaj kontrolü, motor parametrelerinin optimizasyonu ile tedarik zincirindeki seçim ve sıralama problemleri gibi birçok karmaşık optimizasyon probleminde etkin bir şekilde uygulanmıştır. Bu bağlamda, PSO özellikle yüksek boyutlu çözüm uzaylarında etkinliği, uygulama kolaylığı ve hızlı yakınsama özellikleri nedeniyle literatürde yaygın olarak tercih edilen metasezgisel yöntemlerden biridir.



Şekil 3.3. Parçacık Sürü Optimizasyon Algoritması Akış Şeması

Parçacık sürü optimizasyon algoritması sırası ile aşağıdaki adımlardan oluşur:

1. Başlangıçta parçacıkların konum ve hızları rastgele belirlenerek başlangıç sürüsü oluşturulur
2. Sürüdeki her bir parçacığın uygunluk (fitness) değeri hesaplanır
3. Her parçacık için şimdiye kadarki en iyi bireysel konum (pbest) güncellenir
4. Tüm sürü içerisinde en iyi küresel değer (gbest) belirlenir
5. Hız ve pozisyon değerleri güncellenir
6. Durdurma kriteri sağlanıncaya kadar süreç ikinci adımdan itibaren tekrarlanır

Parçacık Sürü Optimizasyonu (PSO) algoritması, başlangıçta konum ve hızları rastgele atanmış parçacıklardan oluşan bir popülasyonla çalışmaya başlar. Algoritmanın temel işleyişinde, her parçacığın konumu ve hızı iteratif olarak güncellenir. Bu güncellemelerin yönü ve büyüklüğü, parçacığın bulunduğu çözümün performansına bağlıdır.

Her iterasyonda elde edilen çözümler, önceden tanımlanmış uygunluk fonksiyonu aracılığıyla değerlendirilir. Uygunluk fonksiyonu, çözümlerin kalitesini ölçmek için kullanılan ve probleme özgü olarak belirlenen bir ölçüttür. Bu değerlendirme sonucunda, her bir parçacık için o ana kadar elde edilen en iyi çözüm parçacık bazlı en iyi değer (pbest) olarak kaydedilir. Popülasyondaki tüm parçacıklar göz önünde bulundurulduğunda ise, pbest değerleri arasından elde edilen en iyi çözüm küresel en iyi değer (gbest) olarak adlandırılır.

Parçacıkların yeni konumları ve hızları, hem kendi pbest değerlerine hem de sürü genelinde belirlenen gbest değerine göre güncellenmektedir. Böylece her iterasyonda parçacıklar, hem kişisel deneyimlerinden hem de sürünün ortak bilgisinden yararlanarak arama uzayında daha iyi çözümlere yönelirler. Güncelleme süreci, hız ve konum denklemleri kullanılarak şekil 3.1 ve 3.2 deki gibi matematiksel olarak ifade edilir:

$$vi(t) = w.vi(t) + c1.r1.(Pbest - xi(t)) + c2.r2.(Gbest - xi(t)) \quad (3.1)$$

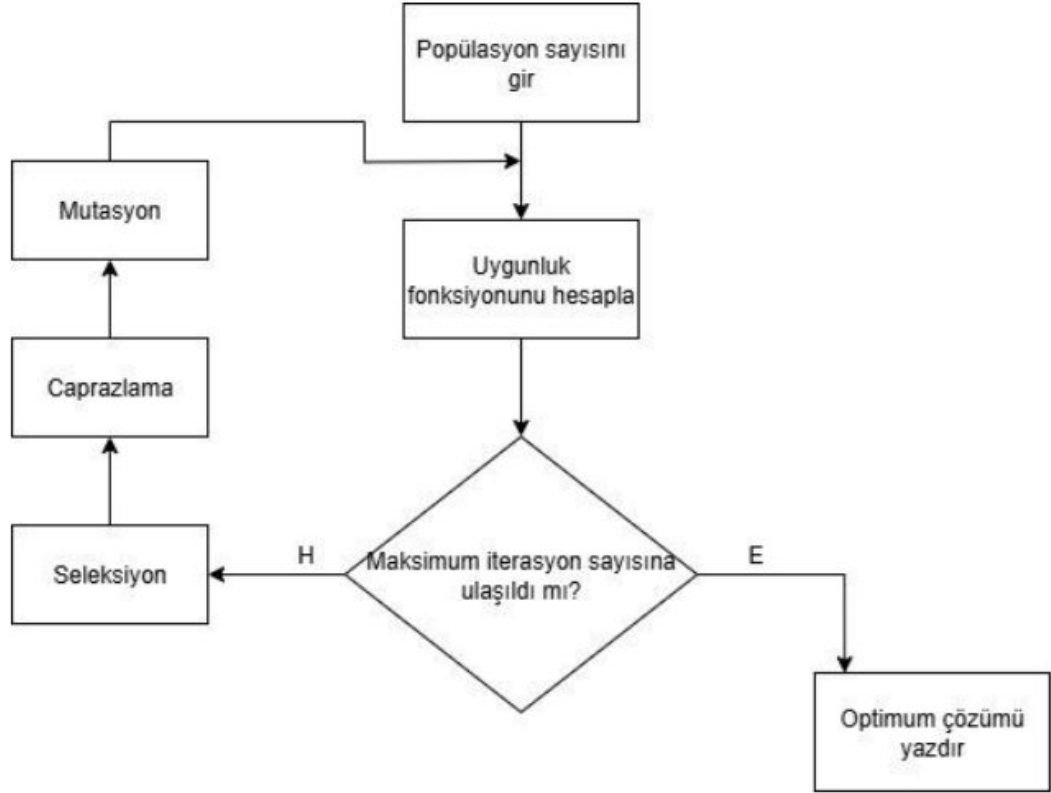
$$Xi(t + 1) = xi(t) + vi(t + 1) \quad (3.2)$$

Burada $vi(t+1)$ hız ve $xi(t+1)$ pozisyon değerlerini verirken, $r1$ ve $r2$ 0-1 arasında rastgele üretilen değerleri, w atalet ağırlığını, c_1 ve c_2 ise genellikle 2 olarak seçilen pozitif ivme katsayılarını ifade etmektedir. Ayrıca, $Pbest$ her bir parçacığın kişisel en iyi çözümünü, $Gbest$ ise tüm popülasyon içinde elde edilen genel en iyi çözümü göstermektedir.

3.1.3.2.2. Genetik Algoritma

Genetik algoritma 1975 yılında Michigan Üniversitesi'nden John Holland tarafından Darwin'in doğal seçim ve evrim teorisinden ilham alınarak geliştirilmiş bir metasezgisel optimizasyon algoritmasıdır (Keklik ve Özcan, 2023).

Doğal seçim, bir popülasyondaki bireylerin çevreye uyum sağlama yeteneklerine göre hayatta kalma kalma ve sonraki nesillere aktarılma durumudur. Başka bir deyişle, yüksek uyum yeteneğine sahip bireyler yaşam şanslarını artırarak gelecek nesillere genlerini aktarırken; düşük uyum yeteneğine sahip bireyler ise ya güçlü bir bireyle çiftleşme yoluyla daha avantajlı gen kombinasyonları elde etmeye çalışır veya mutasyon gibi işlemlerle çevresel koşullara uyum sağlama girişiminde bulunurlar. Ancak bu stratejilerden herhangi biri başarıya ulaşamazsa, ilgili bireyin soyunun zamanla tükenmesi kaçınılmazdır. Bu süreç, doğada gözlemlenen en iyinin hayatta kalması ilkesine dayanmaktadır. Genetik algoritmalar da benzer şekilde, başlangıçta rastgele üretilen çözüm adaylarından oluşan bir popülasyon ile başlar. Daha sonra seçim, çaprazlama ve mutasyon gibi genetik operatörler kullanılarak yeni bireylerden oluşan bir sonraki nesil üretilir. Her nesilde daha iyi çözümler seçilerek popülasyon sürekli olarak geliştirilir ve zamanla optimum veya optimuma yakın çözümler elde edilir (Keklik ve Özcan, 2023).



Şekil 3.4. Genetik Algoritma Optimizasyon Algoritması Akış Şeması

Genetik algoritma; parametre kodlama, başlangıç popülasyonunun belirlenmesi, seçim, çaprazlama, mutasyon, sonlandırma adımlarından oluşmaktadır. Her bir adımın uygulanması hakkında aşağıda bilgi verilmiştir.

3.1.3.2.2.1. Parametre kodlama

Genetik algoritmanın en önemli aşamalarından biri parametrelerin uygun bir şekilde kodlanmasıdır. Kodlama, problemin yapısına bağlı olarak çözüm uzayının bilgisayar ortamında temsil edilmesini sağlar. Kodlama aşamasında yapılan tercihler, algoritmanın hem doğruluğunu hem de hesaplama süresini doğrudan etkilemektedir. Literatürde farklı türde kodlama yöntemleri bulunmaktadır (Keklik ve Özcan,2023).

İkili (Binary) Kodlama: En yaygın kullanılan yöntemdir. Kromozomlar 0 ve 1'lerden oluşan bit dizileri şeklinde temsil edilir. Özellikle ayrık problemlerde etkilidir.

Permütasyon Kodlama: Kromozomlar, belirli bir sıradaki sayılardan oluşur. Bu yöntem özellikle gezgin satıcı problemi gibi sıralama gerektiren optimizasyonlarda tercih edilmektedir.

Değer (Value) Kodlama: Kromozomlar, gerçek sayılar, karakterler ya da nesnelerle temsil edilir. Daha karmaşık problemlerde kullanılabilir.

Ağaç (Tree) Kodlama: Çözümlerin hiyerarşik bir yapı veya ifade şeklinde tanımlandığı problemlerde kullanılır. Genetik programlama alanında yaygın olarak uygulanmaktadır.

Doğru kodlama yapılmadığı takdirde, algoritmanın çözüm uzayında etkili bir arama yapması zorlaşmaktadır.

3.1.3.2.2.2. Başlangıç Popülasyonu

Genetik algoritmanın başlangıç noktasını ilk popülasyonun oluşturulması oluşturmaktadır. Popülasyon, genellikle rastgele oluşturulmuş kromozomlardan meydana gelir. Bununla birlikte, bazı durumlarda probleme özgü bilgi kullanılarak daha iyi çözümler içeren popülasyonlar da oluşturulabilmektedir (Karaboğa, 2020).

Popülasyon büyüklüğü, algoritmanın başarısını etkileyen en önemli parametrelerden biridir. Küçük popülasyonlar çözüm uzayını yeterince temsil edemeyebilirken, çok büyük popülasyonlar ise hesaplama maliyetini artırmaktadır. Bu nedenle popülasyon büyüklüğü, çözüm kalitesi ile hesaplama maliyeti arasında denge gözetilerek belirlenmelidir.

3.1.3.2.2.3. Uygunluk Fonksiyonu

Her kromozomun problemin hedeflerine ne kadar uygun olduğunu değerlendirmek için uygunluk fonksiyonu kullanılmaktadır. Uygunluk fonksiyonu, genetik algoritmanın yönünü belirleyen en önemli bileşenidir (Karaboğa, 2020).

Uygunluk değeri yüksek bireyler, daha iyi çözümleri temsil eder ve bir sonraki nesilde daha yüksek oranda yer alma şansına sahiptir. Uygunluk fonksiyonunun doğru tanımlanmaması, algoritmanın optimal çözümlere ulaşmasını zorlaştırabilir.

3.1.3.2.2.4. Kromozom Seçimi

Uygunluk değerleri hesaplanan kromozomlar arasından, yeni nesillerde yer alacak bireyler belirlenir. Bu aşama doğal seçim mekanizmasını taklit etmektedir. Seçim sürecinde daha uygun bireylerin gelecek nesillere aktarılması amaçlanır. Literatürde sıklıkla kullanılan seçim yöntemleri şunlardır (Karaboğa,2020):

Rulet Tekerleği Seçimi (Roulette Wheel): Kromozomların seçilme olasılığı, uygunluk değerleri ile orantılıdır.

Sıralı Seçim (Rank Selection): Bireyler uygunluklarına göre sıralanır ve bu sıraya göre seçim yapılır.

Turnuva Seçimi (Tournament Selection): Rastgele seçilen birkaç birey arasından en iyisi bir sonraki nesle aktarılır.

Sabit Durum Seçimi (Steady-State): Her nesilde yalnızca birkaç birey değiştirilir, böylece popülasyonun büyük kısmı korunur.

3.1.3.2.2.5. Çaprazlama ve Mutasyon

Genetik algoritmaların yenilikçi bireyler üretmesini sağlayan temel aşama çaprazlama ve mutasyon işlemleridir.

Çaprazlama: İki ebeveyn kromozom arasındaki genlerin belirli yöntemlerle değiş tokuş edilmesi işlemidir. Bu sayede ebeveynlerin özelliklerini taşıyan fakat farklı yapıda yeni bireyler üretilir. Yaygın kullanılan çaprazlama yöntemleri arasında tek noktalı, çok noktalı, uniform ve aritmetik çaprazlama yer almaktadır (Karaboğa, 2020).

Mutasyon: Çözüm uzayında çeşitliliği artırmak amacıyla bireylerin bazı genlerinde rastgele değişiklikler yapılır. Mutasyon, algoritmanın yerel minimumlara sıkışmasını engelleyen en önemli mekanizmalardan biridir (Karaboğa, 2020).

Bu iki işlem sayesinde, popülasyon her nesilde hem mevcut iyi çözümleri korumakta hem de yeni çözüm alanlarını keşfetmektedir.

3.1.3.2.2.6. Sonlandırma

Genetik algoritmanın çalışma süresi sonlandırma kriteri ile belirlenir. Algoritma, genellikle şu durumlarda sonlandırılmaktadır (Karaboğa,2020):

- 1.Belirlenen maksimum nesil sayısına ulaşılması,
- 2.Uygunluk fonksiyonunun istenen eşik değerine erişmesi,
- 3.Popülasyonun çeşitliliğinin belirli bir seviyenin altına düşmesi.

Bu kriterlerden herhangi biri sağlandığında, algoritma durdurulur ve en iyi birey nihai çözüm olarak kabul edilir.

Genetik algoritmalar, 1970’lerden bu yana optimizasyon problemlerinde yaygın bir şekilde kullanılmakta olup, esnek yapıları sayesinde farklı problem türlerine kolayca uyarlanabilmektedir. En eski metasezgisel yöntemlerden biri olmasına rağmen günümüzde popülerliğini korumakta ve yeni algoritmaların geliştirilmesinde temel bir ilham kaynağı olmaktadır. Bununla birlikte, genetik algoritmaların bazı dezavantajları da bulunmaktadır. Özellikle parametre ayarlarının zorluğu, uygunluk fonksiyonunun doğru tanımlanması gerekliliği ve yüksek hesaplama maliyeti, bu durum yöntemin uygulanması sırasında dikkat edilmesi gereken unsurlardan biri olarak değerlendirilmektedir.

3.2. Önerilen SVM-DLOFS Kullanılarak Yapılan Çalışmalar

Bu bölümde Draco kertenkelelerin dış tehlikelere karşı korunma ve avlanma stratejisinden ilham alınarak ortaya çıkarılan yeni bir metasezgisel algoritmanın destek vektör makineleri algoritmaları ile birlikte kullanılarak öz nitelik seçimi problemi için ortaya çıkarılan SVM-DLOFS yönteminden bahsedilmiştir (Aydilek ve Açar ,2025).

3.2.1. Kullanılan Veri Kümeleri

Çizelge 3.2. Kullanılan veri kümeleri

Veri kümeleri	Öznitelik Sayısı	Gözlem Sayısı
Sonar	61	208
Turkish Music Emotion	51	400
Spectf Heart	45	80
Spect Heart	23	79
Libras Movement	91	359
Qsar Biodegradation	42	105

DeneySEL çalışmalar kapsamında kullanılan veri kümeleri Çizelge 3.2’de

sunulmuştur. Veri kümeleri UCI veri deposundan temin edilmiştir (Dua ve Graff, 2019). Sonar veri kümesi 208 gözlem ve 61 öznitelikten (Gorman ve Sejnowski, 1988), turkish music emotion veri kümesi 400 gözlem ve 51 öznitelikten (Türker vd., 2017), spectf heart veri kümesi 80 gözlem ve 45 öznitelikten (Detrano vd., 1989), spect heart veri kümesi 79 gözlem ve 23 öznitelikten (Detrano vd., 1989), libras movement veri kümesi 359 gözlem ve 91 öznitelikten (Freitas vd., 2009) ve qsar biodegradation veri kümesi 105 gözlem ve 42 öznitelikten oluşmaktadır (Mansouri vd., 2013).

3.2.2. Draco Lizard Optimizer

Draco Lizard Optimizer (DLO) yöntemi doğadaki draco kertenkelelerin dış tehlikelere karşı kendilerini koruma ve avlanma stratejilerinden ilham alınarak 2025 yılında Xiaowei Wang tarafından geliştirilen metasezgisel optimizasyon algoritmasıdır (Wang, 2025).

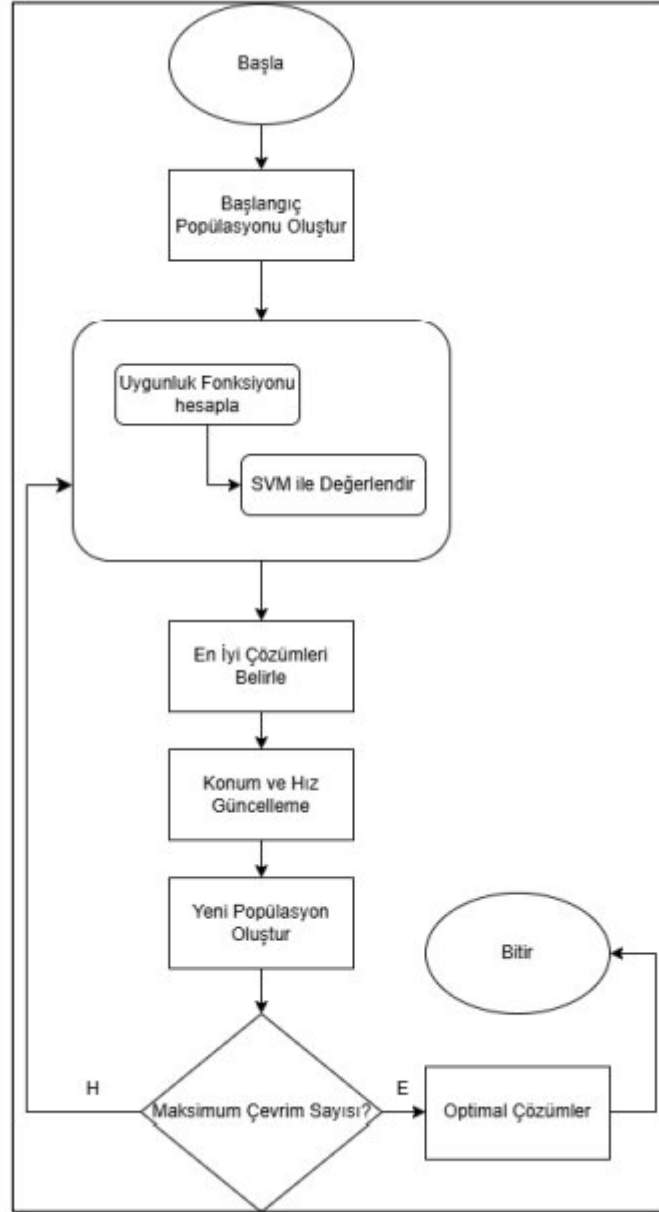
DLO algoritmasında, her bir bireyin bir Draco kertenkelesini temsil ettiği doğa esinli bir optimizasyon yaklaşımıdır. Algoritmanın başlangıç aşamasında, çözüm uzayı içerisinde bireylerin başlangıç konumları rastgele olarak oluşturulur. Daha sonra, bu bireyler tanımlanan uygunluk fonksiyonuna göre değerlendirilir. İkinci aşamada, bireylerin uygunluk değerleri hesaplanarak en yüksek performansa sahip bireyler belirlenir. Bu seçilen bireyler, hem bireysel en iyi çözüm hem de küresel en iyi çözüm olarak kaydedilmektedir (Wang, 2025).

Üçüncü aşamada, bireylerin konum ve hızları güncellenir. Bu güncellemeler, bireylerin kendi en iyi çözümlerine ve sürünün küresel en iyi çözümüne doğru yönelmelerini sağlayan matematiksel denklemler aracılığıyla gerçekleştirilir. Böylece algoritma, küresel arama ve yerel arama süreçleri arasında bir denge kurarak daha etkin bir optimizasyon süreci yürütür. DLO algoritmasında hareket ve konum güncellemeleri, Draco kertenkelesinin süzülme davranışını taklit eden matematiksel modellerle simüle edilir. Bununla birlikte, algoritmanın çeşitliliğini korumak ve yerel minimumlara takılma olasılığını azaltmak amacıyla rastgelelik unsurları da sürece dâhil edilir. Son aşamada, belirlenen durdurma kriteri karşılandığında algoritma sonlandırılarak en iyi çözüm elde edilir (Wang, 2025).

3.2.3. Öznitelik Seçimi için Önerilen SVM-DLOFS Sınıflandırıcısı

Önerilen SVM-DLOFS yönteminde, SVM sınıflandırıcı algoritması ile DLO algoritması birlikte kullanılarak yeni bir yöntem önerisi yapılmıştır. Önerilen yöntem, Draco Lizard Optimization (DLO) algoritmasının güçlü küresel ve yerel

arama yetenekleri, Destek Vektör Makineleri (SVM) sınıflandırıcısı ile entegre edilerek öznitelik seçimi sürecinde etkili bir optimizasyon çerçevesi sunmaktadır. Önerilen yöntemin akış şeması Şekil 3.5’de sunulmuştur (Aydilek ve Açar ,2025).



Şekil 3.5. SVM-DLOFS Akış Şeması

SVM-DLOFS algoritmasında ilk olarak başlangıç popülasyonu oluşturulmuştur. Bu çalışmada popülasyon büyüklüğü 50 olarak belirlenmiş ve her birey, rastgele seçilmiş bir öznitelik alt kümesini temsil edecek şekilde tanımlanmıştır. Bireylerin başlangıç değerleri 0 ile 1 aralığında rastgele üretilmiş olup, bu değerler ilgili öznitelik seçilip seçilmeyeceğini belirlemede kullanılmıştır (Aydilek ve Açar ,2025).

Öznitelik seçiminde DLO algoritması için kullanılan uygunluk fonksiyonu sınıflandırma doğruluğunun en üst düzeye çıkarılması ve seçilen öznitelik sayısının mümkün olduğunca en aza indirgeyecek şekilde tasarlanmıştır. Uygunluk fonksiyonu denklem 3.3’da verilmiştir (Aydilek ve Açar ,2025).

$$Uygunluk\ fonksiyonu = y1.err + y2.\frac{|R|}{|N|} \quad (3.3)$$

Bu çalışmada kullanılan uygunluk fonksiyonunda yer alan değişkenler şu şekilde tanımlanmaktadır: err, SVM sınıflandırıcısı kullanılarak 5 katlı çapraz doğrulama yöntemiyle elde edilen sınıflandırma hata oranını;|R| seçilen öznitelik sayısını, |N| toplam öznitelik sayısını , y1 ve y2 ise sınıflandırma doğruluğu ile seçilen öznitelik sayısı arasında bir denge sağlamak amacıyla belirlenen ağırlık katsayılarını göstermektedir. Bu çalışmada, sınıflandırma başarımının öncelikli olarak optimize edilmesi hedeflenmiş ve bundan dolayı y1 değeri 0,9, y2 ise 0,1 olarak seçilmiştir (Aydilek ve Açar, 2025).

Önerilen yöntemde, DLO algoritmasıyla seçilen öznitelik alt kümeleri, SVM sınıflandırıcısı kullanılarak değerlendirilmiştir. Popülasyondaki bireyler arasında en iyi uygunluk değerine sahip çözümler belirlenmiş, bireysel en iyi ve küresel en iyi değerler güncellenmiştir. Ardından, DLO’nun temel mekanizması olan konum ve hız güncellemesi, Levy uçuşu ve keşif-sömürü stratejileriyle gerçekleştirilmiştir. Bu sayede bireylerin çözüm uzayında daha etkin arama yapması sağlanmıştır. Güncellenen bireyler yeniden değerlendirilerek yeni popülasyon oluşturulmuş, uygunluk değerleri SVM sınıflandırıcısı ile hesaplanmıştır. Bu süreç 250 iterasyona ulaşana kadar tekrarlanmış ve sonunda en iyi öznitelik alt kümesi elde edilmiştir (Aydilek ve Açar, 2025).

3.3. Sınıflandırma Metrikleri

Bu bölümde sınıflandırıcıların performansını ölçmek için kullanılan metriklerden bahsedilmiştir. Şekil 3.6’da performans değerlendirmesi önemli bir yere sahip olan karmaşıklık matrisi sunulmuştur.

		GERÇEK	
		Pozitif	Negatif
TAHMİN	Pozitif	TP	FP
	Negatif	FN	TN

Şekil 3.6. Karmaşıklık matrisi

Gerçek Pozitif(TP): Pozitif sınıfa ait örneklerin model tarafından doğru biçimde pozitif olarak sınıflandırılması durumunu ifade eder. Başka bir ifadeyle, hem gerçek değer hem de tahmin pozitif olduğunda ortaya çıkar.

Gerçek Negatif(TN): Negatif sınıfa ait örneklerin model tarafından doğru biçimde negatif olarak sınıflandırılması durumunu ifade eder. Yani hem gerçek değer hem de tahmin negatif olduğunda gerçekleşir.

Yanlış Pozitif(FP): Gerçekte negatif olan örneklerin model tarafından yanlışlıkla pozitif olarak tahmin edilmesidir. Bu durum, modelin pozitif sınıfa ait olmayan verileri hatalı biçimde pozitif olarak sınıflandırdığını gösterir.

Yanlış Negatif(FN): Gerçekte pozitif olan örneklerin model tarafından yanlışlıkla negatif olarak tahmin edilmesi durumudur. Yani modelin, pozitif sınıfa ait verileri gözden kaçırarak negatif sınıfa dahil etmesi anlamına gelir.

Doğruluk(Accuracy): Sınıflandırma sonucunun hangi oranda doğru tahmin edildiğinin ölçütüdür. Hem doğru pozitif hem de doğru negatif tahminleri dikkate alır. Doğruluk metriği denklem 3.4’de gösterilmiştir.

$$Dogruluk = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.4)$$

Kesinlik(Precision): Modelin pozitif sınıfa yaptığı tahminlerin ne kadarının gerçekten pozitif olduğunu ölçer. Yanlış pozitiflerin etkisini azaltmada önemli bir metriktir. Kesinlik metriği denklem 3.5’te gösterilmiştir.

$$Kesinlik = \frac{TP}{TP + FP} \quad (3.5)$$

Duyarlılık(Recall): Modelin pozitif sınıfı ne kadar iyi yakalayabildiğini gösterir. Yanlış negatiflerin etkisini azaltmaya odaklanır. Duyarlılık metriği denklem 3.6’da gösterilmiştir.

$$Duyarlilik = \frac{TP}{TP + FN} \quad (3.6)$$

F1-Skor: Kesinlik ve duyarlılık ölçütlerinin harmonik ortalamasına dayanarak hesaplanmaktadır. Dengesiz veri kümelerinde (pozitif ve negatif örneklerin sayısı eşit olmadığında) daha güvenilir bir ölçüm sağlar. F1 skor metriği denklem 3.7’de gösterilmiştir.

$$F1 - Skor = 2 \times \frac{Kesinlik \times Duyarlilik}{Kesinlik + Duyarlilik} \quad (3.7)$$

4. BULGULAR

Bu tez kapsamında yürütülen iki ayrı deneysel çalışmadan elde edilen bulgular, birbirinden bağımsız başlıklar altında sunulmuştur.

4.1. Geleneksel Makine Öğrenmesi ve Metasezgisel Yöntemler Kullanılarak Elde Edilen Bulgular

Bu tez kapsamında 3 farklı veri kümesi kullanılarak geleneksel makine öğrenmesine dayalı ve metasezgisel yöntemlere dayalı 5 farklı öznitelik seçim yöntemi uygulanmış ve öznitelik seçimi uygulanmadan önceki hali ile karşılaştırılmıştır.

Deneysel çalışmalar kapsamında, seçilen öznitelik kümeleri üzerinde beş farklı makine öğrenmesi algoritması kullanılmıştır. Bu algoritmaların performanslarının güvenilir bir biçimde ölçülebilmesi için her bir deneyde 10 katlı çapraz doğrulama yöntemi tercih edilmiştir. Böylelikle elde edilen sonuçların rastlantısal etkilerden arındırılması ve daha genellenebilir bir yapıya kavuşturulması amaçlanmıştır.

Her bir veri kümesi için sınıflandırma performansı değerlendirilmesi, doğruluk, f1-skor, kesinlik ve duyarlılık metrikleri üzerinden yapılmış ve yöntemlerin işlem yükünü ortaya koyabilmek için mili saniye cinsinden çalışma süreleri de hesaplanmıştır. Tablo formatında sunulan sonuçlarda, ilgili veri kümesi için en yüksek başarımları elde eden yöntemler koyu yapılarak gösterilmiştir. Bu yaklaşım, öznitelik seçim yöntemlerinin hem doğruluk hem de verimlilik boyutunda nasıl bir katkı sağladığını açıkça ortaya koymaktadır.

Her bir veri kümesine öznitelik seçim yöntemleri uygulandıktan sonra elde edilen 10 katlı çapraz doğrulama sonuçları ve çalışma süreleri incelenmiş; üç farklı veri kümesinden elde edilen bulgular değerlendirilmiştir.

Deneysel çalışmalar kapsamında tüm veri kümelerinde 4/5 eğitim kümesi ve 1/5 oranında test kümesi olacak şekilde ayrılmıştır. Korelasyon analizinde threshold değeri 0.8 seçilmiştir. Genetik algoritma ve PSO algoritmalarında popülasyon sayısı tüm deneylerde 50 olarak seçilmiştir.

Her bir veri kümesi için deneysel çalışmalardan elde edilen sonuçlar doğruluk, kesinlik, f1-skor ve duyarlılık cinsinden farklı grafikler üzerinde değerlendirilmiştir.

4.1.1. Sonar Veri Kümesi

Bu bölümde, Sonar veri kümesi üzerinde gerçekleştirilen deneysel çalışmalardan elde edilen sonuçlar ayrıntılı olarak değerlendirilmiştir. Yapılan analizlerde, sınıflandırma performansı dört farklı değerlendirme metriği üzerinden incelenmiştir. Bu metrikler; doğruluk, kesinlik, F1-skor ve duyarlılık olup, her biri ayrı tablolar halinde sunulmuştur. Ayrıca her yöntemin çalışma süresi ms cinsinden alt başlıkta verilmiştir.

Çizelge 4.1. Sonar Veri kümesi Deney sonuçları (10 cv doğruluk sonuçları/(çalışma süresi (ms))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	61	0,7063 (52,71 ms)	0,7835 (1219,56 ms)	0,8070 (3234,79 ms)	0,8312 (472,07 ms)	0,8140 (1819,97 ms)
Korelasyon	38	0,6761 (43,82 ms)	0,7526 (1335,06 ms)	0,8125 (2435,82 ms)	0,7834 (324,22 ms)	0,7408 (441,61 ms)
Lasso	35	0,6816 (40,74 ms)	0,8316 (1350,40 ms)	0,7949 (2455,34 ms)	0,8319 (331,93 ms)	0,8257 (422,69 ms)
Hipotez testi	36	0,7349 (75,52 ms)	0,8011 (1316,84 ms)	0,7710 (2652,90 ms)	0,7709 (343,23 ms)	0,7772 (440,01 ms)
PSO	39	0,7118 (42,71 ms)	0,8434 (1458,02 ms)	0,8074 (2168,17 ms)	0,8136 (336,69 ms)	0,8132 (462,86 ms)
GA	31	0,7360 (60,58 ms)	0,7651 (1273,89 ms)	0,8250 (2090,06 ms)	0,8382 (334,80 ms)	0,8313 (424,39 ms)

Sonar veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçimi yöntemi ve beş farklı makine öğrenmesi algoritması kullanılarak elde edilen 10 katlı çapraz doğruluk sonuçları ile çalışma süreleri Çizelge 4.1’de sunulmuştur. Deney sonuçları öznitelik seçimi yöntemleri uygulandıktan sonraki sonuçlar ile veri kümesi üzerinde öznitelik seçimi yapılmamış hali ile karşılaştırılmıştır. Elde edilen sonuçlar incelendiğinde öznitelik seçim yöntemlerinin uygulanmasının ardından doğruluk başarımlarının arttığı ve çalışma sürelerinin genel olarak azaldığı görülmüştür.

Öznitelik seçimi yapılmadan algoritmaların performansı değerlendirildiğinde, LightGBM algoritmasının diğer yöntemlere kıyasla daha yüksek doğruluk oranı elde ettiği tespit edilmiştir. Öznitelik seçim yöntemlerinin uygulanmasının ardından ise, Parçacık Sürü Optimizasyonu (PSO) yöntemi 39 öznitelik ile %84,34 doğruluk oranına, Genetik Algoritma (GA) ise 31 öznitelik ile %83,82 doğruluk oranına ulaşmıştır. Bu bulgular, metasezgisel algoritmaların özellikle öznitelik seçiminde

yüksek düzeyde etkili olduğunu ve sınıflandırma performansını anlamlı ölçüde iyileştirdiğini göstermektedir.

Çizelge 4.2. Sonar Veri kümesi Deney sonuçları (10 cv kesinlik sonuçları/(çalışma süresi (ms))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	61	0,6974 (58,43 ms)	0,8330 (1412,96 ms)	0,8284 (2953,64 ms)	0,8484 (385,88 ms)	0,8016 (885,02 ms)
Korelasyon	38	0,6581 (45,44 ms)	0,7765 (1164,60 ms)	0,8109 (2345,72 ms)	0,7813 (389,50 ms)	0,7840 (813,58 ms)
Lasso	35	0,6642 (58,31 ms)	0,8911 (1238,98 ms)	0,7964 (2244,88 ms)	0,8396 (365,52 ms)	0,8506 (798,76 ms)
Hipotez testi	36	0,7150 (65,29 ms)	0,8216 (1180,13 ms)	0,7981 (1979,03 ms)	0,7741 (392,27 ms)	0,7823 (743,92 ms)
PSO	39	0,6894 (71,66 ms)	0,8919 (1354,60 ms)	0,8492 (2671,86 ms)	0,8448 (354,42 ms)	0,8308 (741,07 ms)
GA	31	0,7079 (48,41ms)	0,9014 (1525,29 ms)	0,8014 (2468,89 ms)	0,8377 (360,81 ms)	0,8292 (745,85 ms)

Sonar veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçimi yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama kesinlik (precision) değerleri ile çalışma süreleri Çizelge 4.2’de sunulmuştur. Deney sonuçları incelendiğinde, öznitelik seçimi için kullanılan Genetik Algoritma (GA) yönteminin Rastgele orman algoritması ile birlikte en yüksek kesinlik değerlerini sağladığı görülmektedir.

Öznitelik seçimi yapılmadan algoritmaların performansı değerlendirildiğinde, Hafif Gradyan Artırma (LightGBM) algoritmasının 0,8484 kesinlik oranı ile en başarılı sonucu verdiği belirlenmiştir. Öznitelik seçim yöntemleri uygulandığında ise, geleneksel makine öğrenmesi tabanlı yöntemlerde kesinlik oranlarının düşme eğiliminde olduğu; buna karşılık, GA ve Parçacık Sürü Optimizasyonu (PSO) yöntemlerinde kesinlik değerlerinin anlamlı ölçüde arttığı ve öznitelik sayısının azaldığı tespit edilmiştir. Bu bulgular, metasezgisel tabanlı öznitelik seçim yöntemlerinin Sonar veri kümesinde sınıflandırma performansını artırmada daha etkili olduğunu ortaya koymaktadır.

Çizelge 4.3. Sonar Veri kümesi Deney sonuçları (10 cv duyarlılık sonuçları/(çalışma süresi (ms))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	61	0, 6464 (86,95 ms)	0, 6589 (1134,65 ms)	0, 7482 (3319,10 ms)	0, 7749 (368,33 ms)	0, 7232 (868,54 ms)
Korelasyon	38	0, 6439 (51,33 ms)	0, 6554 (1222,52 ms)	0, 7696 (2158,67 ms)	0, 7357 (365,16 ms)	0, 6804 (797,22 ms)
Lasso	35	0, 6196 (52,36 ms)	0, 7232 (1129,89 ms)	0, 7500 (1915,31 ms)	0, 7892 (365,52 ms)	0, 7661 (781,23 ms)
Hipotez testi	36	0, 7107 (79,80 ms)	0, 7357 (1110,44 ms)	0, 6964 (2322,48 ms)	0, 7357 (354,24 ms)	0, 6857 (774,44 ms)
PSO	39	0, 6982 (51,91 ms)	0, 7500 (1378,30 ms)	0, 7250 (2409,46 ms)	0, 7250 (437,41 ms)	0, 7107 (757,90 ms)
GA	31	0, 6250 (77,80 ms)	0, 7089 (1158,71 ms)	0, 6607 (1804,63 ms)	0, 7642 (332,74 ms)	0, 7625 (766,80 ms)

Sonar veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçimi yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama duyarlılık (recall) değerleri ile çalışma süreleri Çizelge 4.3'te sunulmuştur. Deney sonuçları göz önüne alındığında, öznitelik seçimi için kullanılan Lasso yönteminin Hafif Gradyan Artırma algoritması ile birlikte en yüksek duyarlılık değerlerini sağladığı görülmektedir.

Öznitelik seçimi yapılmadan algoritmaların performansları göz önüne alındığında, Hafif Gradyan Artırma (LightGBM) algoritmasının 0,7749 duyarlılık oranı ile en başarılı sonucu verdiği görülmüştür. Öznitelik seçim yöntemleri uygulandığında ise, bazı geleneksel makine öğrenmesi tabanlı yöntemlerde duyarlılık oranlarının kısmen düştüğü, buna karşın Lasso yöntemi ile LightGBM algoritmasının 0,7892 duyarlılık değeri elde ederek en yüksek başarıyı sağladığı belirlenmiştir. Ayrıca, Rastgele Orman algoritması ile PSO yöntemi kullanıldığında da 0,7500 duyarlılık değeri elde edilerek rekabetçi bir performans sergilenmiştir.

Bu bulgular, Sonar veri kümesinde Lasso tabanlı öznitelik seçiminin özellikle LightGBM algoritmasıyla birlikte kullanıldığında duyarlılık performansını artırmada etkili olduğunu; bununla birlikte, metasezgisel tabanlı yöntemlerin de öznitelik boyutunu azaltarak tatmin edici sonuçlar verdiğini ortaya koymaktadır.

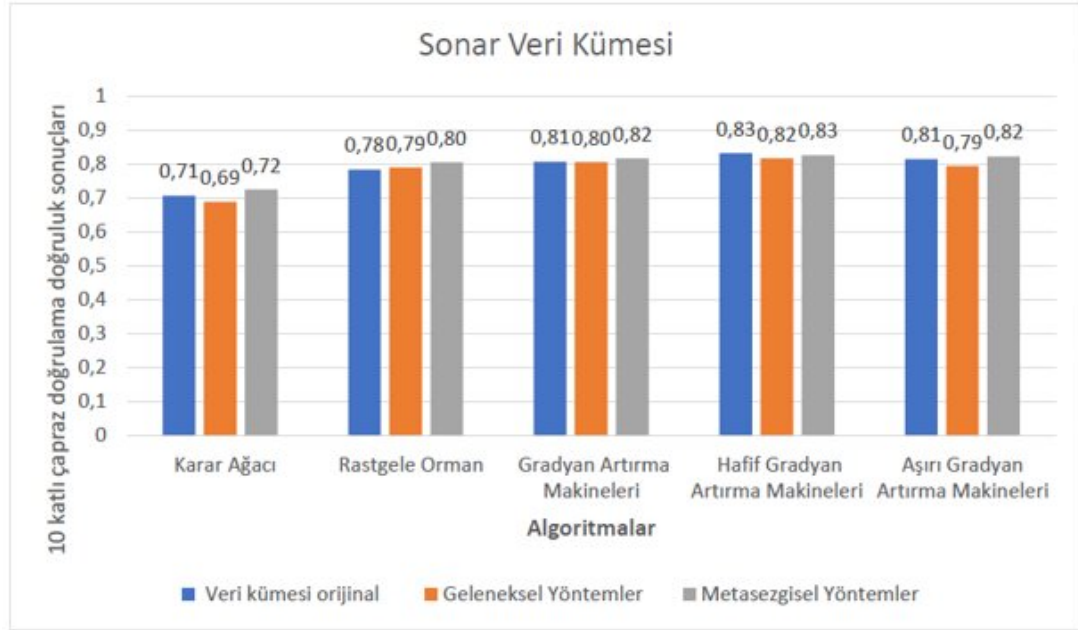
Çizelge 4.4. Sonar Veri kümesi Deney sonuçları (10 cv f1-skor sonuçları/(çalışma süresi (ms))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	61	0,6667 (60,96 ms)	0,7260 (1357,46 ms)	0,7803 (3514,93 ms)	0,8056 (473,86 ms)	0,7565 (2233,62 ms)
Korelasyon	38	0,6404 (73,64 ms)	0,7054 (2116,86 ms)	0,7823 (2216,38 ms)	0,7541 (374,85 ms)	0,7239 (809,68 ms)
Lasso	35	0,6349 (81,23 ms)	0,7927 (1095,80 ms)	0,7684 (2335,54 ms)	0,8095 (343,94 ms)	0,7990 (760,14 ms)
Hipotez testi	36	0,7090 (67,23 ms)	0,7659 (1068,36 ms)	0,7297 (1999,97 ms)	0,7436 (372,30 ms)	0,7257 (728,68 ms)
PSO	39	0,6982 (51,91 ms)	0,7500 (1378,30 ms)	0,7250 (2409,46 ms)	0,7250 (437,41 ms)	0,7107 (757,90 ms)
GA	31	0,6524 (77,75 ms)	0,7891 (1055,31 ms)	0,7036 (1903,27 ms)	0,7947 (373,43 ms)	0,7885 (743,74 ms)

Sonar veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçimi yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama F1-skor değerleri ile çalışma süreleri Çizelge 4.4’de sunulmuştur. Deneysel çalışmanın sonuçları incelendiğinde, öznitelik seçimi için kullanılan PSO algoritmasının Rastgele Orman algoritması ile birlikte en yüksek F1-skor değerini sağladığı görülmektedir.

Öznitelik seçimi yapılmadan algoritmaların performansları göz önüne alındığında, Hafif Gradyan Artırma (LightGBM) algoritmasının 0,8056 F1-skor değeri ile en başarılı sonucu verdiği görülmüştür. Öznitelik seçim yöntemleri uygulandığında ise, PSO yöntemi ile Rastgele Orman sınıflandırıcısının yanı sıra, Lasso yöntemi ile LightGBM algoritmasının 0,8095 F1-skoru elde ederek dikkat çekici sonuçlar verdiği belirlenmiştir.

Genel olarak sonuçlar, Sonar veri kümesinde metasezgisel tabanlı öznitelik seçim yöntemlerinin F1-skor performansını artırmada daha etkili olduğunu göstermektedir. Buna karşılık, korelasyon analizi ve hipotez testi gibi istatistiksel tabanlı yöntemlerde F1-skor değerlerinin diğer yöntemlere göre daha düşük kaldığı görülmüştür. Bu durum, karmaşık veri yapılarında metasezgisel yaklaşımların öznitelik seçiminde daha avantajlı olduğunu ortaya koymaktadır.



Şekil 4.1. Sonar Veri kümesi ortalama doğruluk sonuçları

Şekil 4.1'de Sonar veri kümesi kullanılarak, veri kümesinin öz nitelik seçimi yapılmamış hali ile geleneksel yöntemleri ve metasezgisel yöntemler ortalama doğruluk değerleri karşılaştırılmalı olarak gösterilmiştir.

Sonuçlar incelendiğinde, Karar Ağacı algoritması geleneksel öz nitelik seçim yöntemleri uygulandığında 0,69 doğruluk oranı ile en düşük performansı sergilemiştir. Buna karşın, Hafif Gradyan Artırma (LightGBM) algoritması, metasezgisel tabanlı yöntemlerle gerçekleştirilen öz nitelik seçimi sonrasında 0,83 doğruluk oranı ile en yüksek sınıflandırma başarımını elde etmiştir.

Genel eğilimler değerlendirildiğinde, metasezgisel yöntemlerin çoğu algoritma için doğruluk performansını artırdığı; özellikle LightGBM ve Gradyan Artırma Makinelerinde bu iyileşmenin belirgin olduğu görülmektedir. Buna karşılık, bazı geleneksel yöntemlerde doğruluk değerlerinde kısmi düşüşler gözlemlenmiştir. Bu bulgu, Sonar veri kümesinde öz nitelik seçiminin doğru yöntemlerle gerçekleştirildiğinde sınıflandırma performansına doğrudan katkı sağladığını göstermektedir.

4.1.2. Parkinson Veri Kümesi

Bu bölümde, Parkinson veri kümesi üzerinde gerçekleştirilen deneysel

çalışmalardan elde edilen sonuçlar ayrıntılı olarak değerlendirilmiştir. Yapılan analizlerde, sınıflandırma performansı dört farklı değerlendirme metriği üzerinden incelenmiştir. Bu metrikler; doğruluk, kesinlik ,f1-skor ve duyarlılık olup, her biri ayrı tablolar halinde sunulmuştur.

Çizelge 4.5. Parkinson Veri kümesi Deney sonuçları (10 cv doğruluk sonuçları/(çalışma süresi (ms)))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	61	0,7881 (66,65 ms)	0,8624 (2161,30 ms)	0,8762 (2512,22 ms)	0,8823 (469,69 ms)	0,8690 (417,57 ms)
Korelasyon	38	0,7933 (44,59 ms)	0,8638 (1883,91 ms)	0,8700 (1426,75 ms)	0,8771 (349,45 ms)	0,8848 (369,73 ms)
Lasso	35	0,8148 (69,20 ms)	0,8767 (2060,26 ms)	0,8690 (1586,41 ms)	0,8828 (399,42 ms)	0,8548 (381,95 ms)
Hipotez testi	36	0,8219 (70,05 ms)	0,8424 (1848,64 ms)	0,8690 (1680,00 ms)	0,8623 (402,22 ms)	0,8629 (394,16 ms)
PSO	39	0,8500 (81,49 ms)	0,8552 (1789,30 ms)	0,8767 (1851,08 ms)	0,8828 (399,97 ms)	0,8838 (399,68 ms)
GA	31	0,8576 (63,58 ms)	0,8762 (1879,56 ms)	0,8771 (1446,14 ms)	0,8971 (387,69 ms)	0,8938 (389,71 ms)

Parkinson veri kümesi üzerinde gerçekleştirilen deneylerde, 5 farklı öznitelik seçimi yöntemi ve 5 farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğruluk değerleri ve çalışma süreleri Çizelge 4.5'te verilmiştir. Deney sonuçları öznitelik seçimi yöntemleri uygulandıktan sonraki sonuçlar ile veri kümesi üzerinde öznitelik seçimi yapılmamış hali ile karşılaştırılmıştır.

Elde edilen bulgular incelendiğinde, öznitelik seçim yöntemlerinin uygulanmasının ardından doğruluk oranlarının genel olarak arttığı ve bazı yöntemlerde çalışma sürelerinin de kısaldığı gözlemlenmiştir. Öznitelik seçimi yapılmadan algoritmaların performansı değerlendirildiğinde, Hafif Gradyan Artırma (LightGBM) algoritmasının %88,23 doğruluk oranı ile en başarılı sonucu verdiği belirlenmiştir.

Öznitelik seçim yöntemleri uygulandığında ise, GA yöntemi 12 öznitelik seçerek %89,71 doğruluk oranı ile en yüksek başarıya ulaşmıştır. Ayrıca, PSO yöntemi ile 19 öznitelik kullanıldığında %88,38 doğruluk oranı elde edilmiştir. Bu bulgular, özellikle metasezgisel tabanlı yöntemlerin Parkinson veri kümesinde sınıflandırma performansını önemli ölçüde artırdığını ve daha az öznitelikle yüksek doğruluk oranları sağladığını göstermektedir.

Çizelge 4.6. Parkinson Veri kümesi Deney sonuçları (10 cv kesinlik sonuçları/(çalışma süresi (ms)))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	61	0,8562 (55,51 ms)	0,8899 (1503,97 ms)	0,9003 (1822,89 ms)	0,8999 (395,84 ms)	0,9079 (608,22 ms)
Korelasyon	38	0,8587 (59,67 ms)	0,8777 (1594,83 ms)	0,8906 (1251,20 ms)	0,8953 (458,55 ms)	0,8923 (647,00 ms)
Lasso	35	0,8605 (65,18 ms)	0,8963 (1482,39 ms)	0,9062 (1226,91 ms)	0,9059 (415,70 ms)	0,8876 (601,86 ms)
Hipotez testi	36	0,8758 (85,21 ms)	0,8684 (1801,50 ms)	0,8997 (980,26 ms)	0,9013 (388,62 ms)	0,9103 (623,28 ms)
PSO	39	0,9063 (55,60 ms)	0,8885 (1306,25 ms)	0,9066 (1524,15 ms)	0,8942 (369,09 ms)	0,8779 (628,15 ms)
GA	31	0,8932 (37,68 ms)	0,8850 (1562,91 ms)	0,9209 (910,58 ms)	0,9169 (349,24 ms)	0,8907 (548,09 ms)

Parkinson veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçme yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama kesinlik (precision) değerleri ile çalışma süreleri Çizelge 4.6'da sunulmuştur. Deneysel sonuçlar incelendiğinde, öznitelik seçimi için kullanılan GA yönteminin, özellikle Gradyan Artırma Makineleri ile birlikte %92,09 kesinlik oranı sağlayarak en yüksek başarıma ulaştığı görülmektedir.

Öznitelik seçimi yapılmadan algoritmaların performansları değerlendirildiğinde ise, XGBoost algoritmasının %90,79 kesinlik oranı ile en başarılı sonucu verdiği tespit edilmiştir. Bununla birlikte, öznitelik seçim yöntemlerinin uygulanmasıyla GA ve Lasso yöntemleri, farklı algoritmalarla birlikte dikkate değer iyileştirmeler sağlamıştır.

Genel olarak sonuçlar, Parkinson veri kümesinde metasezgisel yöntemlerin öznitelik seçiminde önemli avantajlar sunduğunu ve sınıflandırma doğruluğu kadar kesinlik performansını da artırdığını ortaya koymaktadır.

Çizelge 4.7. Parkinson Veri kümesi Deney sonuçları (10 cv duyarlılık sonuçları/(çalışma süresi (ms)))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	61	0,8627 (68,09 ms)	0,9364 (1531,19 ms)	0,9455 (1938,16 ms)	0,9545 (468,85 ms)	0,9273 (601,98 ms)
Korelasyon	38	0,8818 (89,85 ms)	0,9636 (1509,56 ms)	0,9545 (1167,72 ms)	0,9636 (379,08 ms)	0,9636 (650,90 ms)
Lasso	35	0,9000 (44,54 ms)	0,9545 (1120,46 ms)	0,9273 (1246,69 ms)	0,9454 (437,31 ms)	0,9455 (598,40 ms)
Hipotez testi	36	0,8891 (73,93 ms)	0,9364 (1469,43 ms)	0,9364 (1365,06 ms)	0,9181 (376,58 ms)	0,9273 (613,81 ms)
PSO	39	0,9000 (64,57 ms)	0,9273 (1315,02 ms)	0,9364 (1326,52 ms)	0,9636 (390,88 ms)	0,9091 (628,78 ms)
GA	31	0,9000 (56,67 ms)	0,9545 (1212,12 ms)	0,9364 (889,64 ms)	0,9545 (319,32 ms)	0,9364 (550,88 ms)

Parkinson veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçme yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama duyarlılık (recall) değerleri ile çalışma süreleri Çizelge 4.7’de sunulmuştur. Deney sonuçları göz önüne alındığında, öznitelik seçimi için kullanılan Korelasyon analizi yönteminin LightGBM algoritması ile birlikte %96,36 duyarlılık oranı sağlayarak en yüksek başarıma ulaştığı görülmektedir.

Öznitelik seçimi yapılmadan algoritmaların performansları değerlendirildiğinde, LightGBM algoritmasının %95,45 duyarlılık oranı ile öznitelik seçimi yapılmamış durumda dahi yüksek bir başarı gösterdiği tespit edilmiştir. Bununla birlikte, öznitelik seçimi yöntemlerinin uygulanması, özellikle Korelasyon Analizi ve PSO yöntemlerinde duyarlılık değerlerini artırmış ve daha az öznitelikle daha yüksek başarı elde edilmesini sağlamıştır.

Genel olarak sonuçlar, Parkinson veri kümesinde öznitelik seçimi yöntemlerinin sınıflandırma performansını olumlu yönde etkilediğini ve özellikle Korelasyon Analizi + LightGBM kombinasyonunun en başarılı sonuçları sunduğunu ortaya koymaktadır.

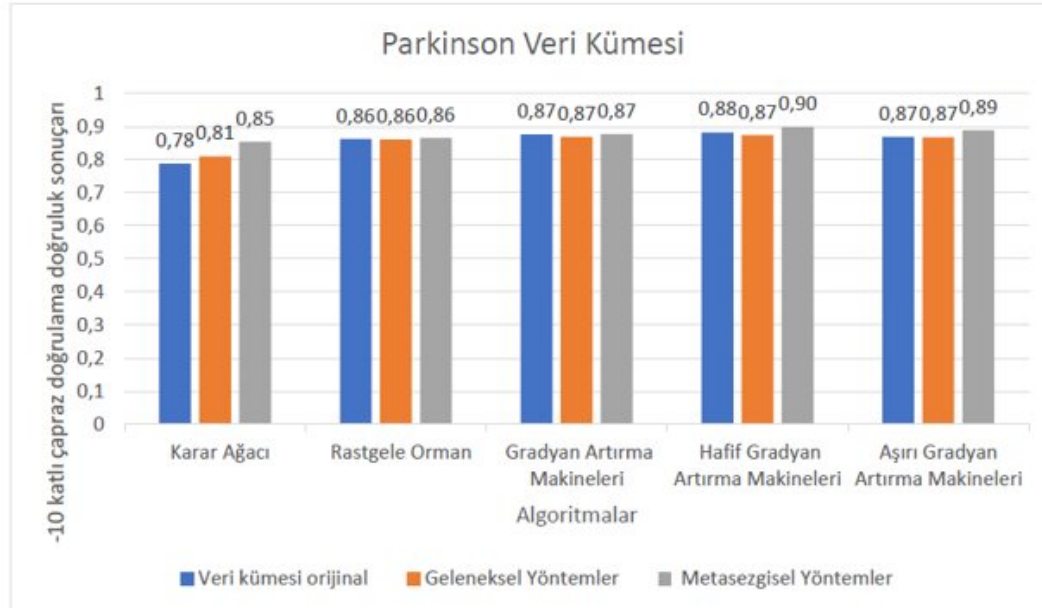
Çizelge 4.8. Parkinson Veri kümesi Deney sonuçları (10 cv f1-skor sonuçları/(çalışma süresi (ms)))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	23	0,8563 (82,45 ms)	0,9096 (1852,64 ms)	0,9186 (1840,95 ms)	0,9221 (442,76 ms)	0,9084 (618,21 ms)
Korelasyon	10	0,8614 (73,70 ms)	0,9127 (1555,38 ms)	0,9156 (885,64 ms)	0,9227 (365,03 ms)	0,9225 (664,63 ms)
Lasso	12	0,8760 (78,20 ms)	0,9200 (1411,82 ms)	0,9091 (1658,02 ms)	0,9214 (411,18 ms)	0,9097 (590,73 ms)
Hipotez testi	15	0,8811 (66,83 ms)	0,8980 (1774,74 ms)	0,9131 (1005,73 ms)	0,9045 (517,99 ms)	0,9121 (677,84 ms)
PSO	19	0,8982 (77,97 ms)	0,9032 (1427,25 ms)	0,9182 (1430,94 ms)	0,9244 (375,98 ms)	0,8874 (628,53 ms)
GA	12	0,8875 (67,79 ms)	0,9155 (1085,21 ms)	0,9222 (797,16 ms)	0,9315 (382,43 ms)	0,9082 (547,35 ms)

Parkinson veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçme yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama f1 skor değerleri ile çalışma süreleri Çizelge 4.8’de sunulmuştur. Deneysel çalışmanın sonuçları incelendiğinde, öznitelik seçimi için kullanılan GA yönteminin LightGBM algoritması ile birlikte %93,15 F1-skor oranı sağlayarak en yüksek başarıya ulaştığı görülmektedir.

Öznitelik seçimi yapılmadan algoritmaların performansları değerlendirildiğinde ise, LightGBM algoritmasının %92,21 F1-skor değeri ile en başarılı sonucu verdiği tespit edilmiştir. Öznitelik seçim yöntemlerinin uygulanması sonucunda, özellikle GA ve PSO tabanlı metasezgisel yöntemler F1-skor değerlerini artırmış, aynı zamanda daha az öznitelikle daha yüksek performans elde edilmesine katkı sağlamıştır.

Genel olarak, Parkinson veri kümesi üzerinde yapılan bu deneysel analizler, öznitelik seçimi yöntemlerinin sınıflandırma performansını artırmada kritik bir rol oynadığını ve GA + LightGBM kombinasyonunun en yüksek başarıyı sunduğunu ortaya koymaktadır.



Şekil 4.2. Parkinson Veri kümesi ortalama doğruluk sonuçları

Şekil 4.2’de Parkinson veri kümesi kullanılarak, öznelilik seçimi yapılmamış hali ile geleneksel yöntemler ve metasezgisel yöntemlerin ortalama doğruluk değerleri karşılaştırmalı olarak sunulmuştur.

Elde edilen sonuçlar incelendiğinde, Karar Ağacı algoritması öznelilik seçimi yapılmamış durumda %78 doğruluk oranı ile en düşük performansı göstermiştir. Buna karşılık, LightGBM algoritması metasezgisel tabanlı yöntemlerle yapılan öznelilik seçiminin ardından %90 doğruluk değerine ulaşarak tüm modeller arasında en yüksek sınıflandırma performansını elde etmiştir.

Genel eğilimler değerlendirildiğinde, metasezgisel yöntemlerin özellikle Karar Ağacı, XGBoost ve LightGBM algoritmalarında doğruluk değerlerini belirgin şekilde artırdığı görülmektedir. Öte yandan, bazı geleneksel yöntemlerde kısmi düşüşler gözlemlenmiştir. Bu bulgular, Parkinson veri kümesinde öznelilik seçiminin uygun yöntemlerle uygulanmasının sınıflandırma performansına doğrudan katkı sağladığını göstermektedir.

4.1.3. Breast Cancer Veri Kümesi

Bu bölümde, Breast Cancer veri kümesi üzerinde gerçekleştirilen deneysel çalışmalardan elde edilen sonuçlar ayrıntılı olarak değerlendirilmiştir. Yapılan analizlerde, sınıflandırma performansı dört farklı değerlendirme metriği üzerinden incelenmiştir. Bu metrikler; doğruluk, kesinlik , f1-skor ve duyarlılık olup, her biri

ayrı tablolar halinde sunulmuştur.

Çizelge 4.9. Breast Cancer Veri kümesi Deney sonuçları (10 cv doğruluk sonuçları/(çalışma süresi (ms)))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	32	0,9225 (169,16 ms)	0,9508 (4871,24 ms)	0,9602 (12285 ms)	0,9766 (3838,36 ms)	0,9648 (1334 ms)
Korelasyon	15	0,8969 (134,11 ms)	0,9461 (4088,72 ms)	0,9414 (6482,1 ms)	0,9484 (3035,88 ms)	0,9483 (1154 ms)
Lasso	15	0,9155 (120,44 ms)	0,9554 (3862,01 ms)	0,9530 (6312,27 ms)	0,9578 (3199,66 ms)	0,9601 (381,95 ms)
Hipotez testi	20	0,9342 (120,82 ms)	0,9555 (4497,90 ms)	0,9508 (8179,6 ms)	0,9648 (3215,57 ms)	0,9532 (1311 ms)
PSO	14	0,9131 (106,65 ms)	0,9438 (4285,74 ms)	0,9484 (6228,7 ms)	0,9650 (2865,00 ms)	0,9454 (1136,01 ms)
GA	16	0,9083 (108,91 ms)	0,9508 (4192,99 ms)	0,9508 (7122,6 ms)	0,9555 (387,69 ms)	0,9530 (1104 ms)

Breast Cancer veri kümesi üzerinde gerçekleştirilen deneylerde, 5 farklı öznitelik seçme yöntemi ve 5 farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğruluk değerleri ve çalışma süreleri Çizelge 4.9’da verilmiştir. Deney sonuçları öznitelik seçimi yöntemleri uygulandıktan sonraki sonuçlar ile veri kümesi üzerinde öznitelik seçimi yapılmamış hali ile karşılaştırılmıştır.

Elde edilen bulgular incelendiğinde, öznitelik seçimi yapılmadan deneyler yapıldığında LightGBM algoritmasının %97,66’lık oranı ile en yüksek doğruluk oranını verdiği gözlemlenmiştir. Öznitelik seçimi yöntemleri uygulandıktan sonra ise, Hipotez Testi yöntemi ile LightGBM algoritması %96,48 lik doğruluk oranı sağlamış ve öznitelik sayısını 20’ye indirmiştir.

Genel olarak elde edilen deneysel sonuçları incelediğimizde öznitelik seçimi uygulandıktan sonra doğruluk oranının %1-%2 bandında düşüşler gözlemlenmiştir. Bununla birlikte, öznitelik seçiminin uygulanması ile birlikte çalışma süresi ve öznitelik sayısı önemli ölçüde azalmıştır. Özellikle metasezgisel tabanlı yöntemlerin kullanılması ile hem daha az öznitelik hemde yüksek doğruluk oranlarına ulaşıldığı görülmüştür.

Çizelge 4.10. Breast Cancer Veri kümesi Deney sonuçları (10 cv kesinlik sonuçları/(çalışma süresi (ms)))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	32	0,9555 (86,38 ms)	0,9602 (1526,62 ms)	0,9672 (5404,35 ms)	0,9748 (1252,11 ms)	0,9715 (730,76 ms)
Korelasyon	15	0,9158 (47,80 ms)	0,9459 (1225,85 ms)	0,9402 (2313,14)	0,9536 (975,17 ms)	0,9391 (720,01 ms)
Lasso	15	0,9335 (54,25 ms)	0,9564 (1392,38 ms)	0,9566 (2431,46 ms)	0,9632 (1034,57 ms)	0,9668 (676,22 ms)
Hipotez testi	20	0,9490 (57,54 ms)	0,9602 (1166,96 ms)	0,9536 (2435,26 ms)	0,9715 (438,58 ms)	0,9606 (497,02 ms)
PSO	14	0,9331 (111,59 ms)	0,9596 (3649,98 ms)	0,9595 (5805,40 ms)	0,9707 (3222,44 ms)	0,9706 (1803,06 ms)
GA	16	0,9418 (140,57 ms)	0,9503 (3823,80 ms)	0,9629 (6480,09 ms)	0,9611 (3519,35 ms)	0,9599 (1808,25 ms)

Breast Cancer veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçme yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama kesinlik (precision) değerleri ve çalışma süreleri Çizelge 4.10'da sunulmuştur. Deney sonuçları, öznitelik seçimi yöntemleri uygulanmadan elde edilen değerler ile öznitelik seçimini uygulandıktan sonra elde edilen değerler karşılaştırılarak değerlendirilmiştir.

Sonuçlar değerlendirildiğinde, öznitelik seçimi yapılmadan gerçekleştirilen deneylerde en yüksek kesinlik oranı %97,48 ile LightGBM algoritması tarafından elde edilmiştir. Öznitelik seçimi yöntemleri uygulandıktan sonra ise, Hipotez Testi yöntemi ile LightGBM algoritması %97,15 ile kesinlik oranı yalnızca %0,3'lük küçük bir düşüş yaşanırken, çalışma süresinde %300 oranında iyileşme gözlemlenmiştir.

Sonuç olarak, Breast Cancer veri kümesi üzerinde yapılan bu deneyler, öznitelik seçimi yöntemlerinin özellikle LightGBM ve metasezgisel tabanlı yöntemler ile birlikte yüksek kesinlik performansını koruduğunu göstermektedir.

Çizelge 4.11. Breast Cancer Veri kümesi Deney sonuçları (10 cv duyarlılık sonuçları/(çalışma süresi (ms)))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	32	0,9212 (58,06 ms)	0,9627 (1890,22 ms)	0,9702 (5006,19 ms)	0,9888 (1163,77 ms)	0,9774 (758,70 ms)
Korelasyon	15	0,9252 (66,89 ms)	0,9701 (1660,59 ms)	0,9699 (3247,76 ms)	0,9662 (1008,27 ms)	0,9588 (710,68 ms)
Lasso	15	0,9326 (59,14 ms)	0,9738 (1476,64 ms)	0,9699 (2348,64 ms)	0,9699 (1093,98 ms)	0,9623 (691,65 ms)
Hipotez testi	20	0,9474 (78,72 ms)	0,9701 (1217,83 ms)	0,9701 (2540,04 ms)	0,9739 (371,36 ms)	0,9699 (517,81 ms)
PSO	14	0,9282 (113,13 ms)	0,9627 (3880,30 ms)	0,9588 (5477,36 ms)	0,9740 (2732,03 ms)	0,9625 (1831,85 ms)
GA	16	0,9289 (138,10 ms)	0,9701 (3720,05 ms)	0,9550 (6677,51 ms)	0,9813 (3638,11 ms)	0,9550 (2474,48 ms)

Breast Cancer veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçme yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama duyarlılık (recall) değerleri ve çalışma süreleri Çizelge 4.11’de sunulmuştur. Deney sonuçları , veri kümesinin orijinal hali ve öznitelik seçimi yöntemleri uygulandıktan sonra yapılan deneysel çalışmalar karşılaştırmalı olarak sunulmuştur.

Deneysel çalışmalardan elde edilen bulgular gözlemlendiğinde, öznitelik seçimi yapılmadan gerçekleştirilen deneylerde en yüksek duyarlılık oranı %98,88 ile LightGBM algoritması tarafından elde edilmiştir. Öznitelik seçimi yöntemleri uygulandığında ise, Genetik Algoritma (GA) LightGBM ile birlikte %98,13 duyarlılık oranı sağlayarak dikkat çekici bir başarı ortaya koymuştur. Benzer şekilde, Hipotez Testi yöntemi de LightGBM ile %97,39 kesinlik oranına ulaşmıştır.

Bununla birlikte, bazı öznitelik seçim yöntemlerinin uygulanması sonucunda duyarlılık değerlerinde küçük çaplı düşüşler gözlemlenmiştir. Öte yandan, Lasso yöntemi, özellikle Rastgele Orman ve Gradyan Artırma algoritmaları ile birlikte %97,38–%96,99 aralığında yüksek duyarlılık değerleri üretmiştir.

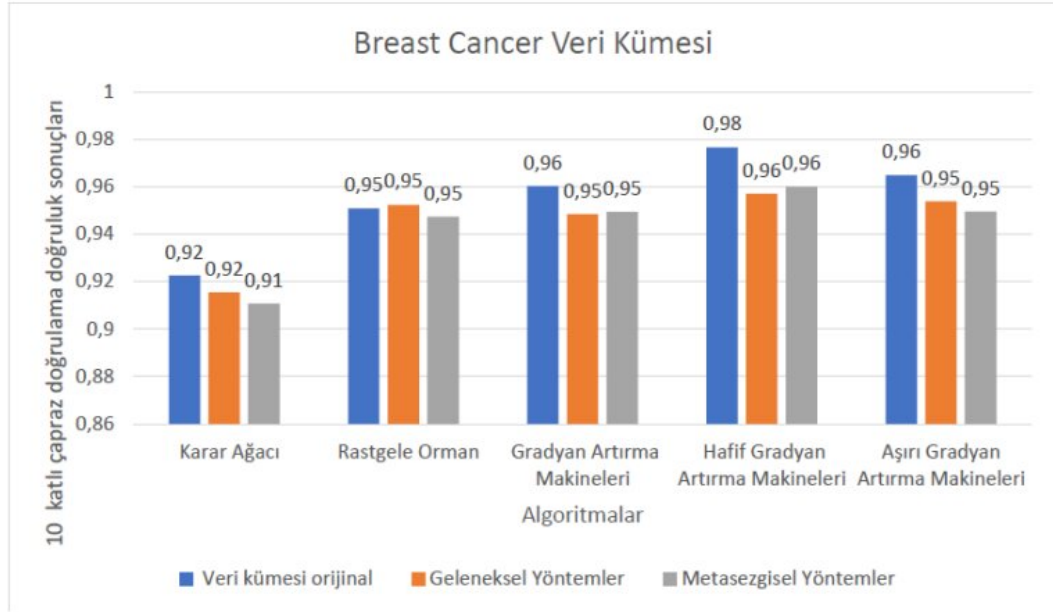
Çizelge 4.12. Breast Cancer Veri kümesi Deney sonuçları (10 cv f1-skör sonuçları/(çalışma süresi (ms)))

Kullanılan öznitelik seçim yöntemleri	Öznitelik sayısı	KA (Başarı oranı/süre)	RO (Başarı oranı/süre)	GBM (Başarı oranı/süre)	LightGBM (Başarı oranı/süre)	XGBoost (Başarı oranı/süre)
Orijinal	32	0,9362 (109,04 ms)	0,9608 (1929,69 ms)	0,9680 (5148,27 ms)	0,9816 (1285,28 ms)	0,9739 (733,20 ms)
Korelasyon	15	0,9185 (65,03 ms)	0,9576 (2215,01 ms)	0,9542 (2250,84 ms)	0,9592 (927,62 ms)	0,9482 (732,42 ms)
Lasso	15	0,9325 (51,94 ms)	0,9647 (1450,04 ms)	0,9629 (2975,14 ms)	0,9662 (1160,05 ms)	0,9640 (698,68 ms)
Hipotez testi	20	0,9474 (61,69 ms)	0,9648 (1226,85 ms)	0,9612 (2431,30 ms)	0,9722 (454,33 ms)	0,9647 (1824,36 ms)
PSO	14	0,9297 (140,34 ms)	0,9607 (3679,88 ms)	0,9585 (5869,39 ms)	0,9721 (3880,02 ms)	0,9661 (3504,59 ms)
GA	16	0,9339 (139,81 ms)	0,9594 (4179,31 ms)	0,9584 (6749,74 ms)	0,9708 (3448,28 ms)	0,9570 (1994,07 ms)

Breast Cancer veri kümesi üzerinde gerçekleştirilen deneylerde, beş farklı öznitelik seçim yöntemi ve beş farklı makine öğrenmesi yöntemi kullanılarak deneysel çalışmalar sonucunda elde edilen 10 katlı çapraz doğrulama f1-skör değerleri ve çalışma süreleri Çizelge 4.12’de sunulmuştur. Deney sonuçları, veri kümesinin orijinal hali ile öznitelik seçimi yöntemleri uygulandıktan sonra yapılan deneysel çalışmaların karşılaştırmalı bulgularını içermektedir.

Elde edilen sonuçlar incelendiğinde, öznitelik seçimi yapılmadan gerçekleştirilen deneylerde en yüksek f1-skör değeri %98,16 ile LightGBM algoritması tarafından elde edilmiştir. Öznitelik seçimi yöntemleri uygulandığında ise, Hipotez Testi yöntemi LightGBM ile birlikte %97,22 f1-skör oranına ulaşmış, böylece yalnızca sınırlı bir performans düşüşü yaşanırken, öznitelik sayısında ve çalışma sürelerinde önemli iyileşmeler sağlanmıştır. Benzer şekilde, PSO yöntemi LightGBM ile birlikte %97,21 oranında yakın düzeyde bir performans ortaya koymuştur.

Bunun yanında, Lasso yöntemi, özellikle Rastgele Orman (%96,47) ve Gradyan Artırma (%96,29) algoritmaları ile yüksek f1-skör değerleri üretmiş, bu yönüyle öznitelik seçimi sonrası güçlü bir performans göstermiştir. Öte yandan, Korelasyon Analizi yöntemi bazı algoritmalarla daha düşük sonuçlar elde etmiş ve genel olarak diğer yöntemlerin gerisinde kalmıştır.

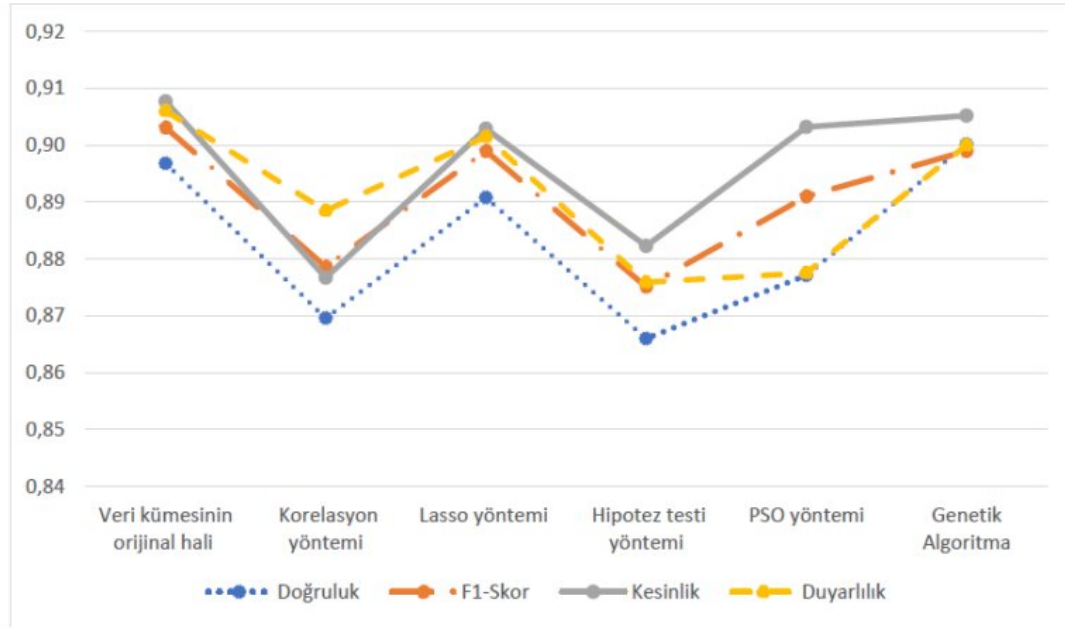


Şekil 4.3. Breast Cancer Veri kümesi ortalama doğruluk sonuçları

Şekil 4.3’de, Breast Cancer veri kümesi kullanılarak, öznelik seçimi yapılmamış hali ile geleneksel yöntemler ve metasezgisel yöntemlerin ortalama doğruluk değerleri karşılaştırmalı olarak sunulmuştur.

Elde edilen sonuçlar incelendiğinde, Karar Ağacı algoritması tüm yöntemlerde en düşük performansı göstermiş ve doğruluk oranı %91–%92 aralığında kalmıştır. Buna karşılık, en yüksek doğruluk oranı, öznelik seçimi yapılmamış durumda LightGBM algoritması tarafından %98 olarak elde edilmiştir. Metasezgisel yöntemler uygulandığında LightGBM doğruluk oranını %96 seviyesinde korumuş, daha az öznelikle yüksek başarımlar elde etmiştir.

Genel eğilimler değerlendirildiğinde, öznelik seçimi sonrasında doğruluk değerlerinde sınırlı düşüşler yaşanmasına karşın, öznelik sayısındaki azalma sayesinde işlem maliyetinde belirgin iyileşmeler sağlanmıştır.



Şekil 4.4. Hafif Gradyan Artırma Makineleri ile 3 veri kümesinin ortalama doğruluk, duyarlılık, kesinlik, F1- Skor sonuçları

Şekil 4.4’de, tüm deneysel sonuçlar dikkate alınarak doğruluk açısından en yüksek performansı gösteren LightGBM algoritması kullanıldığında, üç farklı veri kümesinden elde edilen doğruluk, F1- skoru, kesinlik ve duyarlılık metriklerinin ortalama değerleri sunulmuştur.

Sonuçlar incelendiğinde, öznelik seçim yöntemleri arasında Genetik Algoritma (GA) en yüksek performansı sergileyerek tüm değerlendirme metriklerinde başarılı sonuçlar üretmiştir. Buna karşılık, Korelasyon Analizi yöntemi genel olarak en düşük başarımlarını ortaya koymuş ve diğer yöntemlerin gerisinde kalmıştır.

Değerlendirme metrikleri karşılaştırıldığında, doğruluk değerlerinin diğer metriklerle kıyasla daha düşük seviyelerde seyrettiği; buna karşın, duyarlılık değerlerinin daha yüksek başarı gösterdiği tespit edilmiştir. Ayrıca, f1-skor ve kesinlik değerleri de metasezgisel yöntemlerin uygulanmasıyla genel olarak iyileşme göstermiştir.

Bu bulgular, öznelik seçiminde özellikle metasezgisel tabanlı yöntemlerin (GA ve PSO) sınıflandırma performansını artırmada daha etkili olduğunu ve LightGBM algoritması ile birlikte kullanıldığında güçlü sonuçlar ortaya koyduğunu göstermektedir.

4.2. Önerilen SVM-DLOFS Kullanılarak Elde Edilen Bulgular

Bu çalışmada, UCI veri deposundan temin edilen altı farklı veri kümesi kullanılarak deneysel çalışmalar gerçekleştirilmiştir. Her bir veri kümesi için, orijinal hali ve önerilen yöntem ayrı ayrı değerlendirilmiş; yöntem, 250 iterasyon boyunca 5 katlı çapraz doğrulama ile 10 kez çalıştırılarak elde edilen sonuçlar karşılaştırılmıştır. Deneylerde, öznelilik seçimi süreci için Destek Vektör Makineleri (SVM) sınıflandırıcısı, önerilen Draco Lizard Optimizer Feature Selection (DLOFS) yöntemiyle entegre edilmiştir. Performans değerlendirmesinde doğruluk (accuracy), geri çağırma (recall), kesinlik (precision) ve F1 skoru metrikleri kullanılmış; ayrıca seçilen ortalama öznelilik sayıları da analiz edilmiştir. Sonuçlar Çizelge 4.13'te gösterilmiştir.

Deneysel bulgulara göre:

Sonar veri kümesinde, önerilen yöntem doğruluk oranını %79,64'ten %83,19'a yükseltmiş; duyarlılık ve F1 skor değerlerinde de artış sağlamıştır. Ayrıca ortalama öznelilik sayısı 61'den 12,9'a düşürülerek modelin daha sade bir yapıya kavuşması sağlanmıştır.

Turkish Music Emotion veri kümesinde, dört farklı değerlendirme metriğinde de performans artışı gözlenmiş; ortalama öznelilik sayısı 51'den 13,2'ye inerek modelin boyutu önemli ölçüde azaltılmıştır.

Spectf Heart veri kümesinde, en iyi değerler açısından dört metrikte de iyileşme sağlanırken, ortalama sonuçlarda yalnızca F1 skorunda üstün performans elde edilmiştir. Ayrıca öznelilik sayısı 45'ten 8,1'e düşürülmüştür.

Spect Heart ve Qsar Biodegradation veri kümelerinde, dört değerlendirme metriğinin tamamında performans artışı elde edilmiş; ortalama öznelilik sayıları önemli ölçüde azalmıştır.

Libras Movement veri kümesinde, en iyi değerler açısından dört metriğin tamamında iyileşme sağlanırken, ortalama değerlerde F1 skoru ve kesinlik metriklerinde daha başarılı sonuçlar elde edilmiştir. Seçilen ortalama öznelilik sayısı ise 91'den 18,8'e düşürülerek modelin daha yalın bir yapıya kavuşturulması sağlanmıştır.

Tüm veri kümeleri için hesaplanan standart sapma değerleri, genel olarak önerilen yöntem ile SVM algoritmasının benzer düzeylerde sonuç verdiğini

göstermektedir. Bununla birlikte, bazı veri kümelerinde önerilen yöntemin standart sapma değerlerini daha da düşürdüğü ve bu durumun modelin daha istikrarlı bir performans sergilemesine katkıda bulunduğu görülmüştür.

Sonuç olarak, geliştirilen DLOFS yöntemi, hem sınıflandırma doğruluğunu artırmakta hem de gereksiz öznitelikleri elimine ederek modelin boyutunu küçültmektedir. Böylece, yüksek boyutlu veri kümelerinde aşırı öğrenme riskini azaltmakta ve daha genellenebilir modeller üretilmesini mümkün kılmaktadır.

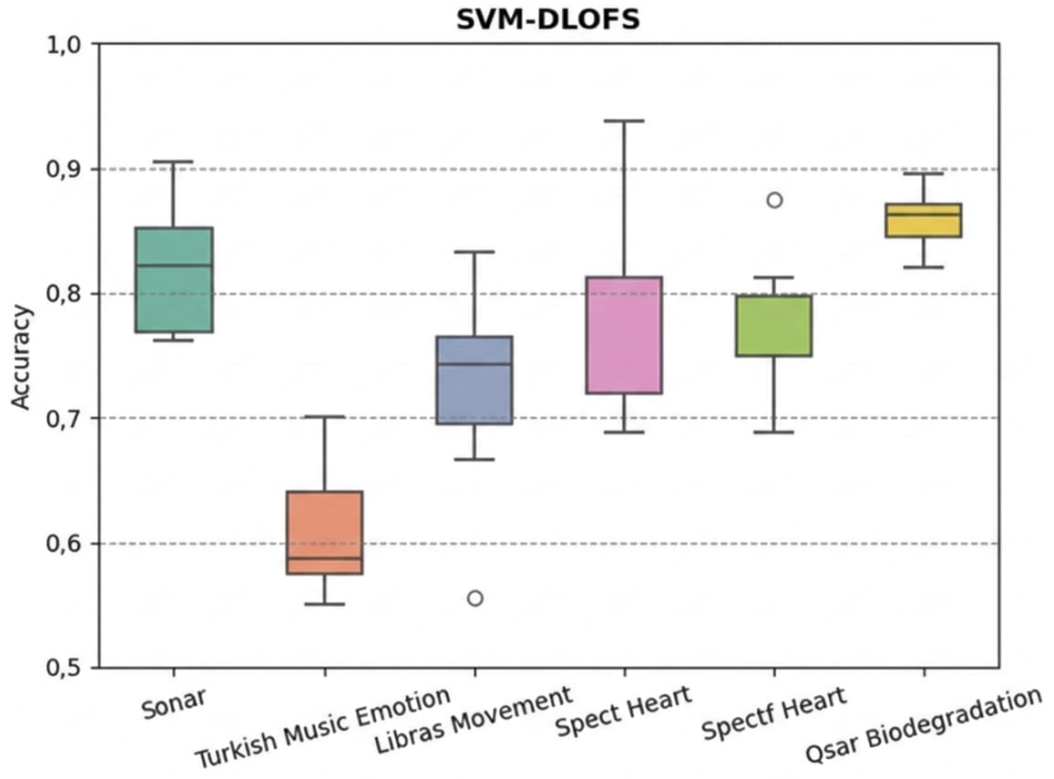
Çizelge 4.13. SVM-DLOFS ve SVM deney sonuçları

		Doğruluk		Duyarlılık		Kesinlik		F1-Score	
		SVM	SVM-DLOFS	SVM	SVM-DLOFS	SVM	SVM-DLOFS	SVM	SVM-DLOFS
Sonar	En İyi	0,8364	0,9048	0,7801	0,9048	0,9230	0,9213	0,8076	0,9062
	Ortalama	0,7964	0,8319	0,7457	0,8319	0,8724	0,8403	0,7964	0,8280
	En Kötü	0,7758	0,7619	0,6879	0,7619	0,8512	0,7619	0,6879	0,7418
	Std,	0,0585	0,0531	0,0347	0,0531	0,0208	0,0510	0,0190	0,0531
	Ort, Öznitelik	61	12,9						
Turkis Music Emotion	En İyi	0,5281	0,7000	0,5281	0,7000	0,6096	0,7022	0,5042	0,6991
	Ortalama	0,4731	0,6098	0,4731	0,6098	0,5301	0,6101	0,4466	0,6067
	En Kötü	0,4219	0,5500	0,4219	0,5500	0,4491	0,5463	0,3865	0,5453
	Std,	0,0318	0,0538	0,0318	0,0538	0,0391	0,0486	0,0391	0,0538
	Ort, Öznitelik	51	13,2						
Spectf Heart	En İyi	0,7987	0,8750	0,7987	0,8750	0,8447	0,8750	0,7925	0,8750
	Ortalama	0,7655	0,7625	0,7655	0,7625	0,8088	0,7858	0,7535	0,7553
	En Kötü	0,7199	0,6875	0,7199	0,6875	0,7215	0,7312	0,6895	0,6875
	Std,	0,0267	0,0515	0,0515	0,0515	0,0337	0,0534	0,0305	0,0533
	Ort, Öznitelik	45	8,1						
Spect Heart	En İyi	0,7744	0,9375	0,8095	0,9375	0,8171	0,9453	0,7781	0,9377
	Ortalama	0,7297	0,8063	0,7125	0,8063	0,7423	0,8165	0,7163	0,7957
	En Kötü	0,6821	0,6875	0,6000	0,6875	0,6878	0,7004	0,6526	0,6838
	Std,	0,0292	0,0744	0,0784	0,0744	0,0396	0,0674	0,0480	0,0730
	Ort, Öznitelik	23	5,6						
Libras Movement	En İyi	0,7667	0,8333	0,7667	0,8333	0,7900	0,8770	0,7459	0,8295
	Ortalama	0,7415	0,7256	0,7415	0,7256	0,7630	0,7854	0,7243	0,7245
	En Kötü	0,7108	0,5556	0,7108	0,5556	0,7207	0,6090	0,6895	0,5411
	Std,	0,0153	0,0844	0,0153	0,0844	0,0233	0,0801	0,0172	0,0852
	Ort, Öznitelik	91	18,8						
Qsar Biodegradation	En İyi	0,8151	0,8957	0,8151	0,8957	0,8139	0,8959	0,8091	0,8921
	Ortalama	0,8039	0,8600	0,8039	0,8600	0,8019	0,8590	0,7951	0,8577
	En Kötü	0,7867	0,8199	0,7867	0,8199	0,7845	0,8197	0,7743	0,8158
	Std,	0,0087	0,0207	0,0087	0,0207	0,0088	0,0210	0,0105	0,0210
	Ort, Öznitelik	42	8,1						

Önerilen yöntemin altı farklı veri kümesi üzerindeki sınıflandırma doğruluk performansına ilişkin sonuçlar, on çalıştırmadan elde edilen değerlerin dağılımı Şekil 4.5'te gösterilmektedir. Boxplot grafiği, her bir veri kümesi için doğruluk değerlerinin merkezi eğilimlerini, dağılım genişliklerini ve kararlılık düzeylerini değerlendirmede kullanılmıştır. Grafik incelendiğinde, özellikle Sonar ve Qsar Biodegradation veri kümelerinde yüksek doğruluk ve istikrarlı bir performans elde edildiği görülmektedir. Bu bulgu, önerilen yöntemin söz konusu veri kümelerinde

öznitelik seçimi ve sınıflandırma açısından etkin bir şekilde çalıştığını ortaya koymaktadır.

Buna karşılık, Turkish Music Emotion veri kümesi, diğer veri kümelerine kıyasla daha düşük doğruluk değerleri sergilemektedir. Medyan doğruluk değerinin yaklaşık %60 seviyelerinde olması, bu veri kümesinin sınıflandırma açısından daha zorlu ve karmaşık bir problem sunduğunu göstermektedir.



Şekil 4.5. SVM-DLOFS doğruluk kutu grafiği

5. TARTIŞMA

Bu tez çalışması kapsamında, makine öğrenmesi ve veri madenciliği problemlerinde karşılaşılan yüksek boyutluluk, hesaplama maliyeti ve aşırı öğrenme sorunlarına çözüm üretmek amacıyla öznitelik seçimi kullanılmıştır. İlk aşamada geleneksel istatistiksel yöntemler ile metasezgisel öznitelik seçim yöntemlerinin performansı karşılaştırılmış; ikinci aşamada ise yeni bir metasezgisel yaklaşım olan Draco Kertenkele Optimizasyon (DLO) algoritması temel alınarak özgün bir öznitelik seçim yöntemi (SVM-DLOFS) önerilmiştir.

Çalışmanın ilk aşamasında Sonar, Parkinson ve Breast Cancer veri kümeleri üzerinde gerçekleştirilen deneylerde, öznitelik seçiminin sınıflandırma başarımı üzerindeki etkisi incelenmiştir. Çizelge 4.1, 4.5 ve 4.9’da sunulan sonuçlar incelendiğinde, öznitelik seçim yöntemlerinin uygulanmasının ardından doğruluk oranlarının arttığı ve çalışma sürelerinin belirgin şekilde düştüğü gözlemlenmiştir. Özellikle Genetik Algoritma (GA) ve Parçacık Sürü Optimizasyonu (PSO) gibi metasezgisel yöntemlerin, Korelasyon Analizi ve Hipotez Testi gibi geleneksel yöntemlere kıyasla daha yüksek sınıflandırma doğruluğuna ulaştığı tespit edilmiştir. Örneğin Parkinson veri kümesinde GA ile LightGBM kombinasyonu %89,71 doğruluk ve %93,15 F1-skor değerine ulaşarak en başarılı sonucu vermiştir. Bu durum, metasezgisel algoritmaların arama uzayında küresel optimuma yaklaşma yeteneğinin, istatistiksel tabanlı filtre yöntemlerine göre daha üstün olmasıyla açıklanabilir. Sınıflandırıcı performansı açısından değerlendirildiğinde ise Hafif Gradyan Artırma (LightGBM) algoritmasının, hem orijinal veri setlerinde hem de öznitelik seçimi sonrasında diğer algoritmalara (KA, RO, GBM, XGBoost) kıyasla daha istikrarlı ve üstün bir performans sergilediği görülmüştür. Bu bulgu, Alp ve Soygazi (2023) tarafından yapılan ve LightGBM ile GA kombinasyonunun diyabet tahmininde en yüksek doğruluğu verdiğini belirten çalışma ile paralellik göstermektedir.

Tezin ikinci ve özgün katkısı olan SVM-DLOFS yöntemi, 6 farklı veri kümesi üzerinde test edilmiştir. DLO algoritmasının avlanma ve korunma stratejilerinden ilham alan yapısı (Wang, 2025), öznitelik seçim probleminde arama uzayını etkin bir şekilde taramayı sağlamıştır. Çizelge 4.12’deki bulgular incelendiğinde, SVM-DLOFS yönteminin standart SVM algoritmasına göre tüm veri kümelerinde doğruluk, kesinlik, duyarlılık ve F1-skor değerlerini artırdığı görülmüştür. Özellikle Sonar veri kümesinde doğruluk oranı %79,64’ten %83,19’a çıkarken, öznitelik sayısı 61’den ortalama 12,9’a düşürülmüştür. Benzer şekilde Libras Movement veri kümesinde 91 öznitelikten 18,8’e inilmesine rağmen performans artışı sağlanmıştır.

Bu sonuçlar, önerilen yöntemin "boyutlanma laneti" problemini aşmada ve gereksiz özniteliklerin yarattığı gürültüyü elimine etmede son derece başarılı olduğunu göstermektedir. Ayrıca, elde edilen düşük standart sapma değerleri (örneğin Qsar Biodegradation veri kümesinde 0,0207), yöntemin kararlı ve güvenilir sonuçlar ürettiğini kanıtlamaktadır.

6. SONUÇLAR

Bu tez çalışmasında, öznitelik seçimi problemini farklı açılardan ele alan iki bağımsız deneysel çalışma sunulmuştur. İlk çalışmada, UCI veri deposundan alınan üç farklı veri kümesi (Sonar, Parkinson ve Breast Cancer) üzerinde geleneksel makine öğrenmesi algoritmalarıyla çeşitli öznitelik seçimi yöntemlerinin performansları karşılaştırılmıştır. Öznitelik seçimi sürecinde Korelasyon Analizi, Lasso, Hipotez Testi, Parçacık Sürü Optimizasyonu (PSO) ve Genetik Algoritma (GA) yöntemleri kullanılmış; sınıflandırma aşamasında ise Karar Ağacı, Rastgele Orman, Gradyan Artırma Makineleri, LightGBM ve XGBoost algoritmaları değerlendirmeye alınmıştır. Deneysel bulgular, doğruluk, kesinlik, duyarlılık ve F1-skor metrikleri ile çalışma süreleri dikkate alınarak analiz edilmiştir. Elde edilen sonuçlar, öznitelik seçimi uygulamasının hem sınıflandırma performansını iyileştirdiğini hem de hesaplama maliyetini önemli ölçüde azalttığını göstermektedir. Öznitelik seçimi uygulandığında, öznitelik sayısında dikkate değer azalmalar sağlanırken doğruluk değerlerinde yalnızca sınırlı düşüşler gözlenmiştir. Bazı durumlarda, örneğin Sonar veri kümesinde PSO ve GA yöntemleri kullanıldığında doğruluk oranlarının arttığı görülmüştür. Algoritmalar arasında LightGBM, öznitelik seçimi uygulanıp uygulanmamasına bakılmaksızın en yüksek doğruluk ve duyarlılık oranlarını sağlayarak en başarılı yöntem olmuştur. GA ve PSO tabanlı metasezgisel yaklaşımlar, özellikle yüksek doğruluk, kesinlik ve F1-skor değerleri ile öne çıkarken; Korelasyon Analizi en düşük başarıyı göstermiştir. Lasso ve Hipotez Testi yöntemleri ise daha kısa çalışma süreleri sağlamasına karşın, performans açısından metasezgisel yöntemlerin gerisinde kalmıştır. Genel olarak, bu çalışmanın bulguları metasezgisel tabanlı yöntemlerin yüksek boyutlu veri kümelerinde öznitelik seçimi için güçlü bir alternatif olduğunu göstermektedir. İkinci çalışmada ise, 2025 yılında literatüre kazandırılan güncel bir metasezgisel algoritma olan Draco Kertenkele Optimizasyonu (DLO), Destek Vektör Makineleri (SVM) ile bütünleştirilerek SVM-DLOFS yöntemi geliştirilmiştir. Bu yöntem, UCI veri deposundan elde edilen altı farklı veri kümesi üzerinde uygulanmış ve performansı 5 katlı çapraz doğrulama ile 250 iterasyon boyunca 10 tekrarlı deneyler sonucunda değerlendirilmiştir. Deneysel süreçte doğruluk, kesinlik, duyarlılık ve F1-skor metrikleri kullanılmış; ayrıca seçilen ortalama öznitelik sayıları analiz edilmiştir. Bulgular, önerilen SVM-DLOFS yönteminin genel olarak sınıflandırma doğruluğunu artırdığını ve modelin daha az öznitelikle daha yalın bir yapıya kavuşmasını sağladığını göstermiştir. Örneğin, sonar veri kümesinde doğruluk oranı %79,64'ten %83,19'a yükselmiş ve öznitelik sayısı 61'den 12,9'a düşürülmüştür. Turkish Music Emotion veri kümesinde tüm değerlendirme metriklerinde artış sağlanırken, öznitelik sayısı 51'den

13,2'ye düşmüştür. Spectf Heart veri kümesinde en iyi değerler açısından dört metrikte de iyileşme kaydedilirken, ortalama değerlerde yalnızca F1 skorunda üstünlük sağlanmıştır. Qsar Biodegradation ve Spect Heart veri kümelerinde dört metrikte de artış elde edilmiş, Libras Movement veri kümesinde ise özellikle F1 skoru ve kesinlik metriklerinde ortalama değerler bakımından gelişim gözlemlenmiştir. Tüm veri kümeleri için standart sapma değerleri incelendiğinde, önerilen yöntemin genellikle SVM ile benzer düzeyde sonuçlar ürettiği; bazı durumlarda ise daha düşük standart sapmalar ile daha kararlı bir performans sergilediği belirlenmiştir. Genel olarak, geliştirilen DLOFS yöntemi yalnızca sınıflandırma doğruluğunu artırmakla kalmamış, aynı zamanda gereksiz öznitelikleri elimine ederek modelin boyutunu küçültmüş ve yüksek boyutlu veri kümelerinde aşırı öğrenme riskini azaltarak daha genellenebilir sonuçlar üretmiştir. Sonuç olarak, tez kapsamında gerçekleştirilen her iki çalışma da öznitelik seçiminin sınıflandırma performansını iyileştirme ve modellerin verimliliğini artırma noktasında kritik bir role sahip olduğunu göstermektedir. İlk çalışmada geleneksel ve metasezgisel yöntemlerin kıyaslaması yapılırken, ikinci çalışmada güncel bir metasezgisel algoritmanın öznitelik seçimine entegre edilmesiyle daha yüksek performanslı ve kararlı modeller elde edilmiştir. Bu bulgular, öznitelik seçimi alanında metasezgisel algoritmaların önemini ve potansiyelini güçlü bir biçimde ortaya koymaktadır.

7. ÖNERİLER

Bu tez çalışması kapsamında geliştirilen ve Destek Vektör Makineleri (SVM) ile entegre edilen Draco Kertenkele Optimizasyon (DLO) tabanlı öznitelik seçim yaklaşımı (SVM-DLOFS), literatüre özgün bir katkı sunmakla birlikte, yöntemin etkinliğinin artırılması ve uygulama alanının genişletilmesi adına gelecek çalışmalarda dikkate alınabilecek çeşitli hususlar bulunmaktadır.

Öncelikle, mevcut çalışmada sarmalayıcı (wrapper) yöntem olarak yalnızca SVM algoritması tercih edilmiştir. Yöntemin farklı sınıflandırma dinamiklerindeki başarısını daha kapsamlı değerlendirebilmek adına, gelecek araştırmalarda DLO algoritmasının K-En Yakın Komşu (KNN), Rastgele Orman (Random Forest) veya Hafif Gradyan Artırma (LightGBM) gibi farklı yapıdaki sınıflandırıcılarla entegre edilerek hibrit yapıların performans analizlerinin yapılması önerilmektedir. Buna paralel olarak, çalışmanın kıyaslama aşamasında kullanılan Genetik Algoritma (GA) ve Parçacık Sürü Optimizasyonu (PSO) gibi klasik yöntemlerin ötesine geçilerek; Balina Optimizasyon Algoritması (WOA), Gri Kurt Optimizasyonu (GWO) veya Harris Şahinleri Optimizasyonu (HHO) gibi literatürdeki daha güncel metasezgisel algoritmalarla kapsamlı karşılaştırmaların yapılması, yöntemin rekabet gücünü daha net ortaya koyacaktır.

Yöntemin genelleştirilebilirliği ve ölçeklenebilirliği açısından veri çeşitliliği de kritik bir önem taşımaktadır. Bu çalışmada kullanılan UCI tabanlı sayısal veri kümelerinin ötesine geçilerek; görüntü işleme, metin madenciliği veya mikro-dizilim genetik verileri gibi çok daha yüksek boyutlu ve karmaşık veri tipleri üzerinde çalışılması literatüre önemli katkılar sağlayacaktır. Ayrıca, önerilen yöntemin uygulama alanı sadece sınıflandırma problemleriyle sınırlandırılmamalı; regresyon problemleri veya kümeleme analizleri gibi farklı makine öğrenmesi görevlerinde de öznitelik seçimi amacıyla kullanılarak performansının test edilmesi faydalı olacaktır.

Son olarak, algoritmanın yapısal iyileştirmelerine odaklanılmalıdır. Tez kapsamında sınıflandırma hatası ve öznitelik sayısı üzerine kurulan uygunluk fonksiyonu (fitness function); modelin eğitim süresi, hesaplama karmaşıklığı ve model kararlılığı gibi parametrelerin de dahil edildiği çok amaçlı (multi-objective) bir yapıya dönüştürülebilir. Ayrıca, DLO algoritmasının yerel optimuma takılma riskini minimize etmek ve arama uzayını daha etkin taramak amacıyla, kaotik haritalar veya farklı metasezgisel operatörlerle zenginleştirilmiş hibrit modellerin tasarlanması, öznitelik seçimi başarımını daha üst seviyelere taşıyacaktır.

KAYNAKLAR

- Açar, İ., & Aydılek, İ. B. (2024). Geleneksel makine öğrenmesi yöntemleri ve metasezgisel yöntemlerle öznelilik seçim yöntemlerinin karşılaştırılması. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 15(2), 397–409.
- Alicković, E., & Subasi, A. (2017). Breast cancer diagnosis using GA feature selection and rotation forest. *Neural Computing and Applications*, 28(4), 753–763.
- Alp, G., & Soygazi, F. (2024). Diyabet tahmini için sınıflandırma algoritmalarında meta-sezgisel destekli öznelilik seçimi. *Meta-Heuristic Supported Feature Selection in Classification Algorithms for Diabetes Diagnosis*.
- Atalay, T., Ayata, D., & Yaslan, Y. (2018). Comparison of feature selection methods in voice based emotion recognition systems. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1–4). IEEE.
- Aydilek, İ., & Açar, İ. (2025). Improving SVM classification performance with Draco Lizard feature selection optimizer. *Proceedings of the Bulgarian Academy of Sciences*, 78(7), 1020–1029.
- Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227.
- Bilen, A., & Özer, A. B. (2022). Siber saldırılar için rastgele orman algoritması kullanılarak öznelilik seçimi. *Fırat Üniversitesi Fen Bilimleri Dergisi*, 34(1), 31–37.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Bulut, C., Ballı, T., & Yetkin, E. (2023). Comparative classification performances of filter model feature selection algorithms in EEG based brain computer interface system. *Journal of the Faculty of Engineering and Architecture of Gazi University*, 38(4).
- Candan, H., Durmuş, A., & Harman, G. (2019). Genetik algoritma ve sınıflandırıcı yöntemler ile kanser tahmini. *Veri Bilimi*, 2(1), 30–34.
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(1), 20–28.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Chen, T., & Guestrin, C. (2017). XGBoost: Reliable large-scale tree boosting system. In *Proceedings of the 22nd SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 13–17).
- Chen, X. W., & Wasikowski, M. (2008). Fast: A ROC-based feature selection metric for small samples and imbalanced data classification problems. In *Proceedings of the 14th ACM SIGKDD International Conference on*

Knowledge Discovery and Data Mining (pp. 124–132).

- Ciran, A., & Özbay, E. (2024). Derin öğrenme ve özellik seçimi yaklaşımları kullanılarak göz hastalıkları tespiti. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 15(2), 421–433.
- Çetin Yağmur, E., & Yağmur, S. (2022). *Rüzgar gücü tahmininde genetik algoritma ile öznitelik seçimi*.
- Deperlioğlu, Ö. (2023). Tıbbi veri kümelerinde genetik algoritmalarla özellik seçimi ve sınıflandırma başarımına etkisi. *Mühendislik Bilimleri ve Tasarım Dergisi*, 11(1), 68–80.
- Detrano, R., Janosi, A., Steinbrunner, C., Griggs, R., Froelicher, V., Young, J., Sandhu, S., Detrano, R., Slobodnick, J., Kloos, T., Bryan, J., Goodman, R., Seager, J., Mitchell, H., Sherwin, R., & Ragosta, M. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology*, 64(5), 304–310.
- Dua, D., & Graff, C. (2019). *UCI Machine Learning Repository*. University of California, School of Information and Computer Science, Irvine, CA.
- Ekrem, Ö., Salman, O. K. M., Aksoy, B., & İnan, S. A. (2020). Yapay zekâ yöntemleri kullanılarak kalp hastalığının tespiti. *Mühendislik Bilimleri ve Tasarım Dergisi*, 8(5), 241–254.
- Emanet, S., Baydoğan, G. K., & Demir, Ö. (2021). Öznitelik seçme yöntemlerinin makine öğrenmesi tabanlı saldırı tespit sistemi performansına etkileri. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 12(5), 743–755.
- Emhan, Ö., & Akın, M. (2019). Filtreleme tabanlı öznitelik seçme yöntemlerinin anomali tabanlı ağ saldırısı tespit sistemlerine etkisi. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 10(2), 549–559.
- Emmanuel, I., Sun, Y., & Wang, Z. (2024). A machine learning-based credit risk prediction engine system using a stacked classifier and a filter-based feature selection method. *Journal of Big Data*, 11(1), 23.
- Farsi, M. (2021). Filter-based feature selection and machine-learning classification of cancer data. *Intelligent Automation & Soft Computing*, 28(1).
- Freitas, C. O. A., Rezende, S. O., & de Carvalho, A. C. P. L. F. (2009). Evaluating classifier ensembles on hand movement recognition from electromyographic signals. In *Proceedings of the International Conference on Machine Learning and Applications (ICMLA)* (pp. 41–46). IEEE.
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 1189–1232.
- Gorman, R. P., & Sejnowski, T. J. (1988). Analysis of hidden units in a layered network trained to classify sonar targets. *Neural Networks*, 1(1), 75–89.
- Gümüşçü, A., Aydılek, İ. B., & Taştaltın, R. (2016). Mikro-dizilim veri sınıflandırmasında öznitelik seçme algoritmalarının karşılaştırılması. *Harran*

- Gümüştü, A., Karadağ, K., Tenekeci, M. E., & Aydılek, İ. B. (2017). Genetic algorithm based feature selection on diagnosis of Parkinson disease via vocal analysis. In *2017 25th Signal Processing and Communications Applications Conference (SIU)* (pp. 1–4). IEEE.
- Harb, H. M., & Desuky, A. S. (2014). Feature selection on classification of medical datasets based on particle swarm optimization. *International Journal of Computer Applications, 104*(5).
- Hasan, R., Biswas, B., Samiun, M., Saleh, M. A., Prabha, M., Akter, J., Joya, F. H., & Abdullah, M. (2025). Enhancing malware detection with feature selection and scaling techniques using machine learning models. *Scientific Reports, 15*(1), 9122. <https://doi.org/10.1038/s41598-025-93447-x>
- Karaboğa, D. (2020). *Yapay zeka optimizasyon algoritmaları*. Nobel Akademik Yayıncılık.
- Kartal, M. (1998). *Bilimsel araştırmalarda hipotez testleri-Parametrik ve nonparametrik teknikler*. Nobel Akademik Yayıncılık.
- Kayaalp, F., Basarslan, M. S., & Polat, K. (2018). A hybrid classification example in describing chronic kidney disease. In *2018 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)* (pp. 1–4). IEEE.
- Kaynar, O., Arslan, H., Görmez, Y., & Işık, Y. E. (2018). Makine öğrenmesi ve öznetelik seçim yöntemleriyle saldırı tespiti. *Bilişim Teknolojileri Dergisi, 11*(2), 175–185.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems, 30*.
- Keklik, G., & Özcan, B. D. (2023). Genetik algoritmaların işleyişi ve genetik algoritma uygulamalarında kullanılan operatörler. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 6*(1), 1052–1066.
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of ICNN'95-International Conference on Neural Networks* (Vol. 4, pp. 1942–1948). IEEE.
- Kılıçarslan, S., Adem, K., & Cömert, O. (2019). Parçacık sürü optimizasyonu kullanılarak boyutu azaltılmış mikrodizi verileri üzerinde makine öğrenmesi yöntemleri ile prostat kanseri tespiti. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 7*(1), 769–777.
- Kumar, S. S., & Shaikh, T. (2017). Empirical evaluation of the performance of feature selection approaches on random forest. In *2017 International Conference on Computer and Applications (ICCA)* (pp. 227–231). IEEE.
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*,

50(6), 1–45.

- Little, M. A., McSharry, P. E., Hunter, E. J., Ramig, L. O., & Clifford, G. D. (2007). Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Transactions on Biomedical Engineering*, 56(4), 1015–1022.
- Ma, D., Shu, M., & Zhang, H. (2025). *Feature selection optimization for employee retention prediction: A machine learning approach for human resource management*.
- Mansouri, K., Carrasco, N., Lewis, K. C., & Jaworska, J. (2013). Quantitative structure–activity relationship models for ready biodegradability of chemicals. *Journal of Chemical Information and Modeling*, 53(4), 867–878.
- Nemati, Z., Mohammadi, A., Bayat, A., & Mirzaei, A. (2024). Metaheuristic and data mining algorithms-based feature selection approach for anomaly detection. *IETE Journal of Research*, 70(7), 6040–6054.
- Osanaiye, O., Ogundile, O., Aina, F., & Periola, A. (2019). Feature selection for intrusion detection system in a cluster-based heterogeneous wireless sensor network. *Facta Universitatis, Series: Electronics and Energetics*, 32(2), 315–330.
- Özsağlam, M. Y., & Çunkaş, M. (2008). Optimizasyon problemlerinin çözümü için parçacık sürü optimizasyonu algoritması. *Politeknik Dergisi*, 11(4), 299–305.
- Ranstam, J., & Cook, J. A. (2018). LASSO regression. *Journal of British Surgery*, 105(10), 1348.
- Sağbaş, E. A. (2023). Filtre tabanlı öznelik seçim yöntemleri kullanılarak metinlerde duygu sınıflandırması üzerine karşılaştırmalı bir çalışma. *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, 35(1), 239–250.
- Sakri, S. B., Rashid, N. B. A., & Zain, Z. M. (2018). Particle swarm optimization feature selection for breast cancer recurrence prediction. *IEEE Access*, 6, 29637–29647.
- Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. *Babylonian Journal of Machine Learning*, 69–79. <https://doi.org/10.58496/bjml/2024/007>
- Shabbir, A., Ali, R. H., Shabbir, M. Z., Abideen, Z. U., Khan, T. A., Ijaz, A. Z., & Bakar, M. A. (2023). Genetic algorithm-based feature selection for accurate breast cancer classification. In *2023 International Conference on IT and Industrial Technologies (ICIT)* (pp. 1–6). IEEE.
- Sharawi, M., Zawbaa, H. M., & Emary, E. (2017). Feature selection approach based on whale optimization algorithm. In *2017 Ninth International Conference on Advanced Computational Intelligence (ICACI)* (pp. 163–168). IEEE.
- Sulaiman, M. H., Mustaffa, Z., Saari, M. M., & Abas, M. F. (2024). Wind power forecasting with metaheuristic-based feature selection and neural networks. *Cleaner Energy Systems*, 9, 100149.

- Şengür, D. (2021). Öğretmenlerde etkili psikososyal risk düzeylerinin belirlenmesi için parçacık sürü optimizasyonu tabanlı öznitelik seçme. *Türkiye Eğitim Dergisi*, 6(2), 488–496.
- Tokmak, M. (2023). XGBoost algoritması ile ikili parçacık sürü optimizasyonu öznitelik seçme tabanlı jar kötü amaçlı yazılımlarının tespiti. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, 10(1), 140–152.
- Türker, B., Yıldırım, S., & Akın, M. (2017). *Turkish music emotion dataset*. [UCI Repository].
- UCI Machine Learning Repository. (2025). *UCI Machine Learning Repository*. <https://archive.ics.uci.edu/>
- Vaishali, R., Sasikala, R., Ramasubbareddy, S., Remya, S., & Nalluri, S. (2017). Genetic algorithm based feature selection and MOE Fuzzy classification algorithm on Pima Indians diabetes dataset. In *2017 International Conference on Computing Networking and Informatics (ICCNi)* (pp. 1–5). IEEE.
- Vatansever, B., Aydın, H., & Çetinkaya, A. (2021). Genetik algoritma yaklaşımıyla öznitelik seçimi kullanılarak makine öğrenmesi algoritmaları ile kalp hastalığı tahmini. *Journal of Science, Technology and Engineering Research*, 2(2), 67–80.
- Venkatesh, B., & Anuradha, J. (2019). A review of feature selection and its methods. *Cybernetics and Information Technologies*, 19(1), 3–26.
- Wang, X. (2025). Draco lizard optimizer: A novel metaheuristic algorithm for global optimization problems. *Evolutionary Intelligence*, 18, Article 10. <https://doi.org/10.1007/s12065-024-00998-5>
- Wolberg, W. H., & Mangasarian, O. L. (1990). Multisurface method of pattern separation for medical diagnosis applied to breast cytology. *Proceedings of the National Academy of Sciences*, 87(23), 9193–9196.
- Yağmur, E., & Yağmur, S. (2022). Rüzgar gücü tahmininde genetik algoritma ile öznitelik seçimi. *AKÜ FEMÜBİD*, 22, 1028–1040. <https://doi.org/10.35414/akufemubid.1117779>
- Yılmaz, A., & Sümer, E. (2021). Relief özellik seçim yöntem tabanlı önerilen hibrit model ile kalp hastalığı teşhisi. *Avrupa Bilim ve Teknoloji Dergisi*, (31), 609–615.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

İsim Soyisim : İSMAİL AÇAR
Doğum Tarihi : 0000-00-00
Doğum Yeri : BİRECİK
Telefon : 0500 000 00 00
E-Posta : ismailacr63@gmail.com

T.C
YÜKSEKÖĞRETİM KURULU
ULUSAL TEZ MERKEZİ

TEZ VERİ GİRİŞ FORMU

[Ljava.lang.String;@55186241

Referans No	10788879
Yazar Adı / Soyadı	İSMAİL AÇAR
Orcid	0009-0007-9888-7429
T.C.Kimlik No	10398065956
Telefon	5374872632
E-Posta	ismailacr63@gmail.com
Tezin Dili	Türkçe
Tezin Özgün Adı	METASEZGİSEL ALGORİTMALARLA ÖZNİTELİK SEÇİMİ YÖNTEMLERİNİN UYGULAMASI VE GELİŞTİRİLMESİ
Tezin Tercümesi	APPLICATION AND DEVELOPMENT OF FEATURE SELECTION METHODS USING METAHEURISTIC ALGORITHMS
Konu	Bilgisayar Mühendisliği Bilimleri-Bilgisayar ve Kontrol = Computer Engineering and Computer Science and Control
Üniversite	Harran Üniversitesi
Enstitü / Hastane	Fen Bilimleri Enstitüsü
Ana Bilim Dalı	Bilgisayar Mühendisliği Ana Bilim Dalı
Bilim Dalı	Bilgisayar Mühendisliği Bilim Dalı
Tez Türü	Yüksek Lisans
Yılı	2026
Sayfa	65
Tez Danışmanları	PROF. DR. İBRAHİM BERKAN AYDİLEK
Dizin Terimleri	
Önerilen Dizin Terimleri	Metasezgisel algoritmalar, Draco kertenkele optimizasyon, Genetik algoritma, Parçacık sürü optimizasyon algoritması, Öznitelik seçimi

10.03.2026

İmza:.....

