

“

BİLGİSAYAR MÜHENDİSLİĞİ

ALANINDA ULUSLARARASI ARAŞTIRMA VE DEĞERLENDİRMELER

Aralık 2024

EDİTÖR

PROF. DR. SELAHATTİN BARDAK

”

Genel Yayın Yönetmeni / Editor in Chief • C. Cansın Selin Temana

Kapak & İç Tasarım / Cover & Interior Design • Serüven Yayınevi

Birinci Basım / First Edition • © Aralık 2024

ISBN • 978-625-5955-84-5

© copyright

Bu kitabın yayın hakkı Serüven Yayınevi'ne aittir.

Kaynak gösterilmeden alıntı yapılamaz, izin almadan hiçbir yolla çoğaltılamaz.

The right to publish this book belongs to Serüven Publishing. Citation can not be shown without the source, reproduced in any way without permission.

Serüven Yayınevi / Serüven Publishing

Türkiye Adres / Turkey Address: Kızılay Mah. Fevzi Çakmak 1. Sokak

Ümit Apt No: 22/A Çankaya/ANKARA

Telefon / Phone: 05437675765

web: www.serüvenyayınevi.com

e-mail: serüvenyayınevi@gmail.com

Baskı & Cilt / Printing & Volume

Sertifika / Certificate No: 47083

BİLGİSAYAR MÜHENDİSLİĞİ

Alanında Uluslararası Araştırma ve Değerlendirmeler

ARALIK 2024

EDİTÖR

PROF. DR. SELAHATTİN BARDAK

İÇİNDEKİLER

BÖLÜM 1

GRİ KURT ALGORİTMASI VE VARYANTLARININ SINIFLANDIRMA PERFORMANSININ DEĞERLENDİRİLMESİ

Rabia UYAR 1

BÖLÜM 2

TESPİT SİSTEMLERİ: BİR KUANTUM UYGULAMASI

Gamzepelin AKSOY 13

Murat KARABATAK² 13

BÖLÜM 3

GELİŞMİŞ BOYUT İNDİRGEME TEKNİKLERİ: DOĞRUSAL OLMAYAN ALGORİTMALARIN ROLÜ

Zeynep Özpolat 27

BÖLÜM 1

GRİ KURT ALGORİTMASI VE VARYANTLARININ SINIFLANDIRMA PERFORMANSININ DEĞERLENDİRİLMESİ

Rabia UYAR¹

¹ Arş. Gör. Rabia UYAR, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü,
Bilgisayar Bilimleri Anabilim dalı

GİRİŞ

Teknolojideki gelişmeler, veri hacminde hızlı bir artışa yol açmıştır. Bu verilere hızlı ve kolay erişim sağlamak analistler ve veri bilimciler için önemli bir gereksinim haline gelmiştir. Bu ihtiyacı karşılamak için veri madenciliği adı verilen yöntemler geliştirilmiştir. Veri madenciliği, büyük veri kümelerindeki gizli ilişkileri, örüntüleri ve bilgileri ortaya çıkarmayı amaçlayan çok adımlı bir süreçtir (Emre & Selçukcan Erol, 2017). Bu süreç çeşitli araçlar ve teknolojiler kullanılarak gerçekleştirilir ve veri setinde saklı olan anlamlı bilgileri ortaya çıkarır. Veri madenciliği sürecindeki en kritik adımlardan biri özellik seçimidir (Onuralp Gökalp et al., 2018). Özellik seçimi, performanstan ödün vermeden orijinal veri setini en iyi temsil eden alt kümenin seçilmesi işlemidir. Bu süreç, büyük veri setlerinin analizinde önemli bir rol oynar (Çelik & Akdamar, 2018). Özellik seçimi, veri setindeki n özellik arasından en iyi k özelliğin seçilmesi işlemi olarak tanımlanır. Özellik seçiminin temel amacı, veri kümesindeki özellik sayısını azaltmak ve en anlamlı ve önemli özellikleri seçmektir (Uladi, A. İ. & Arı, E. S., 2023).

Özellik sayısının azaltılması, analiz ve modelleme sürecinde önemli faydalar sağlar. Bu faydalar arasında daha hızlı model yürütme, daha düşük hesaplama maliyetleri ve daha iyi model netliği yer alır. Ayrıca, veri boyutunun azaltılması analizi daha yönetilebilir ve sonuçları daha yorumlanabilir hale getirir (Tekerek, 2011). Bu nedenle, özellik sayısının azaltılması işlemi veri madenciliği sürecinde kritik bir adımdır (Savaş et al., 2012).

Literatürde birçok özellik seçme yöntemi bulunmaktadır (Tekerek, 2011). Bu yöntemlerden bazıları karar ağaçlarıdır. Karar ağaçları özellikle sınıflandırma problemlerinde yaygın olarak kullanılmaktadır ve özellik seçimi için etkili bir araçtır. Karar ağaçları veri setini belirli kurallara göre dallara ayırarak en iyi özellikleri bulmaya çalışır. Karar ağaçlarının yapısında farklı algoritmalar vardır ve bu algoritmalarından en yaygın olanı ID3 algoritmasıdır. ID3 algoritması, bir özellik seçme ve örnek kümesini o özelliğin değerlerine göre bölme işlemi tekrarlar. Ayrıca, geriye doğru özellik seçimi için destek vektör makineleri (SVM) ve tekrarlanan özellik eleme gibi yöntemler kullanılmaktadır (Altan & Karasu, 2020; Emeksiz & Tan, 2021). Bu yöntem, en düşük puan alan özelliği veri kümesinden çıkararak sınıflandırma başarısını optimize etmek amacıyla bir amaç fonksiyonuna dayalı olarak tüm özelliklere bir derece atar.

Veri madenciliği süreçlerinde, son yıllarda popülerlik kazanan algoritmalarından biri de sezgisel algoritmalar (Han & Kim, 2002). Sezgisel algoritmalar, problem çözümüne sezgisel yaklaşımlar sağlar ve matematiksel modelin kesin olarak bilinmediği durumlarda kullanılır (Rajabioun, 2011). Sezgisel algoritmaların doğruluğu kanıtlanmaz; bunun yerine problem çözümünü basitleştirmeleri veya tatmin edici bir sonuç üretmeleri beklenir (Xue et al., 2013).

Popülasyon temelli Gri Kurt Optimizasyon algoritması, sezgisel algorit-

malara bir örnektir (Mirjalili et al., 2014). Bu algoritma, gri kurtların liderlik hiyerarşisi ve avlanma davranışlarına dayanır. Gri kurtların doğada avlanma için sergiledikleri içgüdüsel davranışlar, bu algoritmanın temelini oluşturur. İyileştirilmiş Gri Kurt Optimizasyon algoritması (IGWO)'nun amacı, klasik GWO'nun hızını ve doğruluğunu artırarak daha etkili ve verimli çözümler sunmaktır (Rajwar et al., 2023). Bu iyileştirmeler, özellikle karmaşık ve büyük veri kümelerinde önemli performans kazanımları sağlayabilir.

Essemble Gri Kurt Optimizasyon algoritması (EGWO) ise, klasik GWO algoritmasının çözüm doğruluğunu ve optimizasyon süreçlerindeki hızını artırmak amacıyla geliştirilmiştir (Mosavi et al., n.d.). EGWO, GWO'nun temel ilkelerine bağlı kalarak, özellikle arama uzayının daha etkili taranmasını ve küresel optimuma daha hızlı ulaşılmasını sağlar. Bu iyileştirmeler, özellikle yüksek boyutlu, karmaşık ve gürültülü problemler üzerinde belirgin performans artışları sunarak daha güvenilir ve verimli sonuçlar elde edilmesine olanak tanır (Storn & Price, 1997). Bu algoritmalar genel olarak, klasik Gri Kurt Optimizasyon algoritmasından daha etkili sonuçlar elde etmeyi amaçlar.

Bu çalışma, veri madenciliği ve özellik seçimi konularında yapılan mevcut çalışmaları ele alarak, GWO algoritmasının ve varyantlarının test edilmesini amaçlamaktadır. IGWO ve EGWO algoritmalarının, mevcut kaynaklardan elde edilen sınıflandırma test verileri için orijinal GWO algoritması ile karşılaştırılarak, performansı ve etkinliği analiz edilmiştir. Elde edilen sonuçlar, varyant algoritmaların veri madenciliği ve özellik seçimi süreçlerindeki potansiyel avantajlarını göstermektedir. Bu çalışmada mevcut kaynaklardan elde edilen Zoo sınıflandırma veri seti için GWO, IGWO ve EGWO algoritmaları eş zamanlı olarak çalıştırılarak, veri seti için sınıflandırma hatası, zaman, doğruluk ve özellik sayısı ölçümleri istatistiksel olarak elde edilip incelenmiştir.

MATERYAL VE METHOD

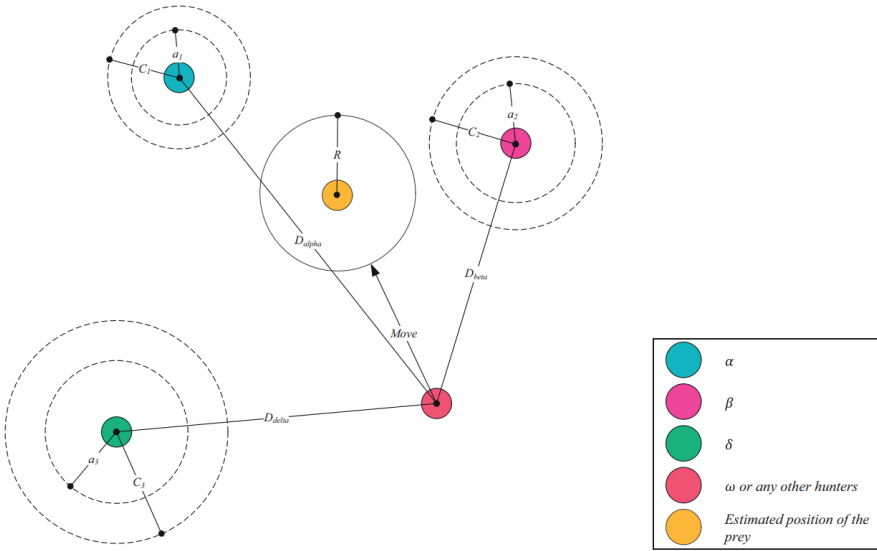
Gri Kurt Optimizasyon Algoritması (GWO)

Gri Kurt Optimizasyonu (GWO), doğal bir meta-sezgisel algoritmadır ve gri kurtların avlanma davranışlarını taklit ederek çözüm arama işlemini gerçekleştirir. Algoritma, doğadaki kurt sürülerinin avlanma sürecini model alarak tasarlanmıştır. GWO, sürüdeki en iyi üç çözümü alfa, beta ve delta olarak belirler ve diğer bireyler (omega) bu üç liderin etrafında avı (optimum çözüm) yakalamak için hareket eder. GWO'nun temel prensibi, keşif ve sömürü aşamaları arasında dinamik bir denge kurmaktır. Keşif aşamasında, kurtlar geniş bir arama alanında avı bulmaya çalışırken, sömürü aşamasında avı yakalamaya odaklanırlar. Bu iki aşama arasındaki geçiş, algoritmanın farklı optimizasyon problemlerinde etkin bir şekilde kullanılmasını sağlar. Alfa, beta ve delta kurtlarının konumları, diğer kurtların yeni pozisyonlarını belirlemede anahtar rol oynar. GWO algoritması, çözümlerin çeşitli pozisyonlarını güncelleyerek küresel optimuma ulaşmayı amaçlar. Kurtların hareketleri, avın pozisyonuna

olan uzaklıklarına bağlı olarak ayarlanır. Bu süreçte, kurtlar avın etrafında spiral veya rastgele hareketlerle dolaşarak daha iyi çözümler ararlar. Bu mekanizma, algoritmanın yerel optimumlara sıkışmasını önler ve global optimumu bulma şansını artırır. GWO, parametrik yapısı ve esnekliği ile birçok farklı optimizasyon problemine uygulanabilir. Algoritmanın basit yapısı ve az sayıda parametre gereksinimi, onu diğer karmaşık optimizasyon yöntemlerine göre daha çekici kılar. GWO, sürekli ve kesikli optimizasyon problemlerinde etkili bir şekilde kullanılabilir ve çeşitli alanlarda başarılı sonuçlar elde etmiştir (Zhao et al., 2019).

Sonuç olarak, GWO algoritması, doğadan ilham alan ve avlanma davranışlarını taklit eden bir optimizasyon yöntemidir. Keşif ve sömürü aşamaları arasında dinamik bir denge kurarak, çözümlerin yerel optimumlarda sıkışmasını önler ve küresel optimuma ulaşmayı hedefler. Basitliği ve etkinliği ile birçok farklı optimizasyon problemine kullanılabilir, bu da GWO'yu geniş bir uygulama yelpazesinde popüler bir seçim haline getirir.

GWO algoritmasının pozisyon güncellemeleri Şekil 1'de sunulmaktadır.



Şekil 1: GWO Pozisyon Güncellemesi (Mirjalili vd., 2014)

Şekil 1, GWO algoritmasında kullanılan avlanma stratejisini ve kurtların avın yerini tespit etme sürecini görselleştirmektedir. Burada;

- Mavi (α), grubun lideri ve en uygun çözümü temsil eder. α , avın yerini bulmada en etkili bireydir.

- Pembe (β), ikinci en uygun çözümü temsil eder ve liderden sonra grubu yönlendiren ikinci en güçlü bireydir.

- Yeşil (δ), üçüncü en uygun çözümü ifade eder ve α ve β tarafından yönlendirilir.

- Kırmızı (ω veya diğer avcılar), geri kalan kurtları temsil eder ve avın yerini bulmak için α , β ve δ 'ya bağlı hareket eder.

- Sarı ise optimizasyon sürecinde, bulunan en iyi tahmini çözüm avın konumunu ifade etmektedir.

Çizimde, D_{α} , D_{β} ve D_{δ} , avcılarının avın konumuna olan mesafelerini temsil eder. Avcılar (kurtlar), bu mesafeleri küçülterek avın (optimum çözümün) yerini bulmaya çalışır. Her bir kurt, rastgele bir bileşen ve yönlendirilmiş bir hareket ile sürekli olarak avın konumuna yaklaşmaktadır.

İyileştirilmiş Gri Kurt Optimizasyon Algoritması (IGWO)

İyileştirilmiş Gri Kurt Optimizasyonu (IGWO), kurt sürüsü liderlerinin yeniden konumlandırılması için yeni metodolojiler kullanarak algoritmanın daha uygun ve optimal çözümler belirlemesini sağlar (Long vd., 2018)

IGWO, kurt sürüsü liderlerinin konumlarını yeniden kalibre etmek için yeni metodolojiler kullanır. Bu stratejiler, çözüm uzayının daha geniş bir bölümünün taranmasına izin verir ve algoritmanın daha iyi performans göstermesine yardımcı olur. IGWO, klasik Gri Kurt Optimizasyonu'nun (GWO) bazı eksikliklerini gidermek için geliştirilmiştir. Yeni uygulanan güncelleme mekanizmaları sayesinde, algoritma yerel minimumlarda sıkışıp kalmayı önler. Bu mekanizmalar, kurtların dinamik ve uyarlanabilir hareketlerini geliştirerek daha geniş bir arama alanı sağlar ve böylece daha iyi çözümler elde edilmesine yardımcı olur. IGWO algoritması, başlangıç pozisyonlarının daha hassas bir şekilde belirlenmesiyle öne çıkar. Bu da başlangıç aşamasında daha üstün çözümler elde edilmesini sağlar. Algoritma, lider kurtların konumlarını dinamik olarak güncelleyerek arama kapasitesini artırır ve çözüm uzayının daha geniş bir bölümünü tarar. Bu, IGWO'nun karmaşık ve büyük ölçekli problemlerde daha başarılı olmasına katkıda bulunur. Algoritma, daha hızlı bir yakınsamayı kolaylaştırmak için optimize edilmiştir (Dada et al., 2022). Dinamik lider güncellemeleri ve adaptif parametre ayarlamaları, algoritmanın performansını artırarak daha hızlı ve daha etkili çözümler üretmesine olanak tanır. IGWO'nun bu özelliği, özellikle zaman kısıtlaması olan projelerde büyük bir avantaj sağlar. IGWO'nun hesaplama verimliliği artırılmış ve böylece daha büyük ölçekli problemlere uygulanabilirliği genişletilmiştir. Genel olarak, IGWO hem daha hızlı hem de daha etkili çözümler üretmektedir. Karmaşık problemlerle karşılaşıldığında, IGWO'nun daha yüksek bir başarı oranına sahip olduğu gösterilmiştir. Bu, IGWO'nun geniş uygulama yelpazesinde kullanılabilirliğini artırır ve onu çeşitli optimizasyon problemlerinde tercih edilen bir yöntem haline getirir.

Ensemble Gri Kurt Optimizasyon Algoritması (EGWO)

Ensemble GWO Gri Kurt Optimizasyonu (EGWO), Gri Kurt Optimizasyonu algoritmasının geliştirilmiş bir türevi olarak global optimizasyon problemlerinin çözümünde daha etkin sonuçlar elde etmeyi amaçlayan bir meta-sezgisel yaklaşımdır (Yu & Wu, 2022)

EGWO, klasik GWO'nun doğadan esinlenen avlanma stratejisini temel almakta ve keşif (exploration) ile sömürü (exploitation) süreçleri arasındaki dengeyi iyileştirmek için birden fazla optimizasyon mekanizmasını bir araya getirmektedir. Bu yöntem, GWO'nun temel prensiplerine bağlı kalarak, özellikle karmaşık ve çok modlu arama uzaylarında daha istikrarlı ve güvenilir çözümler sunmaktadır. EGWO'nun geliştirilmesi sürecinde, klasik GWO algoritmasının sınırlılıkları değerlendirilmiş ve çeşitli iyileştirmeler hayata geçirilmiştir. Bu bağlamda, adaptif ağırlıklar, dinamik parametre güncellemeleri ve popülasyon çeşitliliğini artırmayı hedefleyen stratejiler entegre edilmiştir. EGWO, farklı alt algoritmaların avantajlarını bir araya getirerek, her birinin güçlü yönlerinden faydalanmakta ve böylece lokal minimumlara sıkışma olasılığını azaltmaktadır. Bu birleşik yapı, keşif ve sömürü süreçleri arasında dengeli bir geçiş sağlayarak, algoritmanın küresel optimuma ulaşma başarısını artırmaktadır. Ayrıca, EGWO'nun düşük hesaplama maliyetine sahip olması ve hızlı yakınsama yeteneği, büyük boyutlu ve zorlu problemler için uygulanabilirliğini önemli ölçüde artırmaktadır. Ensemble Gri Kurt Optimizasyonu, mühendislik tasarımı, makine öğrenmesi, veri analitiği ve enerji sistemleri gibi geniş bir yelpazede uygulama potansiyeline sahiptir. Özellikle çok modlu fonksiyonların optimizasyonu ve karmaşık kısıtlar içeren problemler üzerinde üstün performans göstermektedir. Diğer meta-sezgisel algoritmalara kıyasla, EGWO'nun esnek yapısı, çeşitli problem gereksinimlerine uyulanabilirliğini kolaylaştırmaktadır. Bu yenilikçi yöntem, yüksek doğruluk oranları ve güvenilir çözümler sunarak optimizasyon süreçlerinde önemli bir katkı sağlamaktadır. EGWO, alan yazınına sunduğu yenilikler ve uygulama potansiyeli ile meta-sezgisel optimizasyon yaklaşımlarında öncü bir yere sahiptir.

Veri Seti Özellikleri

Bu veri seti bir hayvanat bahçesinden 101 hayvandan oluşmaktadır. Hayvanları tanımlamak için çeşitli özelliklere sahip 16 değişken vardır. 7 Sınıf Türü şunlardır: Memeli, Kuş, Sürüngen, Balık, Amfibi, Böcek ve Omurgasız. Bu veri setinin amacı, değişkenlere dayanarak hayvanların sınıflandırmasını tahmin edebilmektir.

Deneysel Çalışma

Gri Kurt Optimizasyonu (GWO), İyileştirilmiş Gri Kurt Optimizasyonu (IGWO) ve Ensemble Gri Kurt Optimizasyonu (EGWO) algoritmalarının doğruluk performanslarının karşılaştırılması tablo 1'de sunulmaktadır.

Performans değerlendirmesi en iyi, en kötü, ortalama ve standart sapma değerleri üzerinden yapılmıştır. Bu sonuçlar, algoritmaların optimizasyon problemlerinde sergiledikleri doğruluk düzeylerini ve istikrarını ölçmek için önemli bir kıyaslama sağlamaktadır.

Tablo 1 Algoritmaların Doğruluk Değeri Karşılaştırması

Doğruluk Değeri				
Algoritma Adı	En İyi	En Kötü	Ortalama	Standart Sapma
GWO	0.880095	0.93809	0.97164	0.005910
IGWO	0.950000	0.87000	0.96000	0.004000
EGWO	0.976000	0.83400	0.95120	0.003890

EGWO, tabloda en yüksek doğruluk performansını sergileyen algoritma olarak dikkat çekmektedir. En iyi doğruluk değeri açısından EGWO (0.9760), IGWO (0.9500) ve GWO'dan (0.9381) üstün bir performans göstermiştir. Bu durum, EGWO'nun global optimuma ulaşma kapasitesinin diğer algoritmalara göre daha yüksek olduğunu göstermektedir. Ortalama doğruluk değerlerinde ise IGWO (0.9600), EGWO (0.9512) ve GWO'dan (0.9716) daha yüksek bir sonuç elde etmiş olsada, EGWO'nun standart sapma değeri (0.00389) en düşük olarak hesaplanmıştır. Bu, EGWO'nun sonuçlarındaki kararlılığı ve istikrarını ortaya koymaktadır. Sonuç olarak, EGWO'nun standart sapma değerinin düşük olması, bu algoritmanın doğruluk değerlerinde daha az varyans sergilediğini ve daha tutarlı bir performans sunduğunu ifade etmektedir. En kötü doğruluk değerlerinde EGWO'nun IGWO ve GWO'dan daha düşük bir performans göstermesi, algoritmanın potansiyel olarak lokal minimumlara sıkışma durumunu tamamen ortadan kaldırmadığını göstermektedir. Ancak genel doğruluk ve istikrar açısından EGWO, diğer iki algoritmaya kıyasla daha üstün bir performans sergilemektedir. Bu sonuçlar, EGWO'nun optimizasyon problemlerinde etkin bir yöntem olarak kullanılabileceğini kanıtlamaktadır.

Gri Kurt Optimizasyonu (GWO), İyileştirilmiş Gri Kurt Optimizasyonu (IGWO) ve Ensemble Gri Kurt Optimizasyonu (EGWO) algoritmalarının sınıflandırma hatası performanslarının karşılaştırılması tablo 2'de sunulmaktadır. Değerlendirme metrikleri arasında en iyi, en kötü, ortalama ve standart sapma değerleri yer almakta olup, algoritmaların hata oranlarını ve performans kararlılıklarını ölçmek için önemli bir karşılaştırma sağlamaktadır.

Tablo 2 Algoritmaların Sınıflandırma Hatası Karşılaştırması

Sınıflandırma Hatası				
Algoritma Adı	En İyi	En Kötü	Ortalama	Standart Sapma
GWO	3.78e-04	2.39e-03	1.67e-02	5.23e-02
IGWO	3.50e-04	2.10e-03	1.50e-02	5.17e-02
EGWO	3.63e-04	2.25e-03	1.53e-02	5.00e-02

En iyi sınıflandırma hatası açısından IGWO, $3.50e-04$ ile en düşük hata oranını sergilemiş ve GWO ($3.78e-04$) ile EGWO'yu ($3.63e-04$) geride bırakmıştır. Bu sonuç, IGWO'nun sınıflandırma hatasını minimize etme potansiyelinin daha yüksek olduğunu göstermektedir. En kötü sınıflandırma hatası değerlerinde ise IGWO, $2.10e-03$ ile en düşük değere ulaşırken, EGWO ($2.25e-03$) ve GWO'dan ($2.39e-03$) daha başarılı bir performans göstermiştir. Bu durum, IGWO'nun hata oranlarındaki güvenilirliği açısından öne çıktığını ortaya koymaktadır. Ortalama sınıflandırma hatası değerleri açısından, IGWO ($1.50e-02$) en düşük hata oranına sahip olurken, EGWO ($1.53e-02$) ve GWO ($1.67e-02$) sırasıyla daha yüksek değerlere sahiptir. Bununla birlikte, standart sapma değerlerinde EGWO'nun $5.00e-02$ ile en düşük değeri sergilemesi, bu algoritmanın hata oranlarında daha kararlı bir performans sunduğunu göstermektedir. Standart sapma sonuçları, EGWO'nun sınıflandırma hatasının sonuçlar arasında daha az varyans sergilediğini ve istikrar açısından avantajlı bir yöntem olduğunu ortaya koymaktadır. Sonuç olarak, IGWO'nun genel olarak sınıflandırma hatasını minimize etme kapasitesinin diğer algoritmalara kıyasla daha iyi olduğu görülmekte; ancak EGWO'nun daha düşük standart sapma değerine sahip olması, sınıflandırma hatalarında daha tutarlı bir sonuç sunduğunu göstermektedir. Bu durum, EGWO'nun özellikle hata oranındaki istikrarın kritik olduğu uygulamalar için tercih edilebilir bir algoritma olduğunu işaret etmektedir.

Gri Kurt Optimizasyonu (GWO), İyileştirilmiş Gri Kurt Optimizasyonu (IGWO) ve Ensemble Gri Kurt Optimizasyonu (EGWO) algoritmalarının işlem sürelerini değerlendirmesi tablo 3'te sunulmaktadır. Zaman parametresi açısından algoritmalar, en iyi, en kötü, ortalama ve standart sapma değerleri ile karşılaştırılmıştır. Bu değerlendirme, algoritmaların işlem süresi performansı ve kararlılığı hakkında önemli bilgiler sağlamaktadır.

Tablo 3 Algoritmaların Zaman Karşılaştırması

Zaman				
Algoritma Adı	En İyi	En Kötü	Ortalama	Standart Sapma
GWO	248.33	226.90	276.70	3.90
IGWO	245.70	230.66	270.00	3.50
EGWO	237.85	229.93	262.55	3.33

En iyi işlem süresi açısından değerlendirildiğinde, EGWO 237.85 birim ile en düşük işlem süresine sahiptir ve IGWO (245.70) ile GWO'dan (248.33) daha hızlı bir performans sergilemiştir. Bu sonuç, EGWO'nun işlem süresini minimize etme kapasitesinin daha yüksek olduğunu göstermektedir. En kötü işlem süresi metriklerinde de benzer bir durum gözlenmiş; EGWO 229.93 ile en düşük değeri elde ederek, IGWO (230.66) ve GWO'dan (226.90) daha tutarlı bir sonuç sunmuştur. Ortalama işlem süresi bakımından EGWO (262.55), IGWO (270.00) ve GWO'dan (276.70) daha düşük bir süre ile üstünlük sağlamıştır.

Ayrıca, standart sapma değerlerinde de EGWO'nun 3.33 birim ile en düşük değere sahip olması, bu algoritmanın işlem sürelerinde daha kararlı ve tutarlı bir performans sunduğunu ortaya koymaktadır. GWO (3.90) ve IGWO'nun (3.50) standart sapma değerlerinin daha yüksek olması, bu algoritmaların işlem sürelerinde daha fazla değişkenlik sergilediğini göstermektedir. Genel olarak, EGWO, işlem süresi açısından hem ortalama hem de standart sapma bakımından en iyi sonuçları sağlayarak, diğer algoritmalara kıyasla daha hızlı ve kararlı bir optimizasyon süreci sunmaktadır. Bu durum, EGWO'nun zaman açısından kritik uygulamalarda daha uygun bir yöntem olduğunu göstermekte ve bu algoritmanın kullanım potansiyelini artırmaktadır.

Gri Kurt Optimizasyonu (GWO), İyileştirilmiş Gri Kurt Optimizasyonu (IGWO) ve Ensemble Gri Kurt Optimizasyonu (EGWO) algoritmalarının seçtiği özellik sayısına ilişkin performans değerlendirmesi tablo 4'te sunulmaktadır. Özellik sayısı, bir optimizasyon algoritmasının problemi çözmek için gerekli minimum bilgi miktarını temsil ettiğinden, bu değerlendirme algoritmaların etkinliğini ve gereksiz veriyi elemeyi ne kadar iyi başardığını göstermektedir.

Tablo 4 Algoritmaların Özellik Sayısı Karşılaştırması

Özellik Sayısı				
Algoritma Adı	En İyi	En Kötü	Ortalama	Standart Sapma
GWO	8	12	6.37	1.53
IGWO	7	11	6.50	1.40
EGWO	5	9	7.20	1.27

En düşük özellik sayısı açısından EGWO, 5 ile en iyi sonucu elde etmiş ve bu, diğer algoritmalara kıyasla daha az özellik seçerek problemi çözebildiğini göstermektedir. IGWO (7) ve GWO (8) ile karşılaştırıldığında, EGWO'nun daha az özellik ile yeterli bilgiyi sağlayabilme yeteneği, algoritmanın daha etkili bir özellik seçim süreci sunduğunu ifade etmektedir. En yüksek özellik sayısı bakımından EGWO, 9 ile yine daha az özellik seçerek GWO (12) ve IGWO'dan (11) daha üstün bir performans sergilemiştir. Ortalama özellik sayısı açısından değerlendirildiğinde, EGWO 7.20 ile IGWO (6.50) ve GWO'dan (6.37) biraz daha yüksek bir değere sahiptir. Bu durum, EGWO'nun belirli durumlarda daha fazla özelliği seçebildiğini göstermektedir; ancak bu seçim, genellikle gereksiz özellikleri ortadan kaldırmayı başaran algoritmanın genel etkinliğini etkilememektedir. Ayrıca, standart sapma değerlerinde EGWO'nun 1.27 ile en düşük değeri göstermesi, algoritmanın özellik seçimi sürecinde daha kararlı olduğunu ve seçilen özellik sayısında daha az değişkenlik sergilediğini ortaya koymaktadır.

Sonuç olarak, EGWO, özellik seçiminde daha az özellik ile aynı veya daha iyi performans sağlayarak, özellikle yüksek boyutlu veri kümelerinde etkili bir

yöntem olduğunu göstermektedir. Daha düşük standart sapma değeri, algoritmanın istikrarlı bir performans sunduğunu ve farklı problemlerde seçilen özellik sayısında tutarlılık sağladığını kanıtlamaktadır. Bu durum, EGWO'nun veri boyutunu azaltırken bilgi kaybını minimumda tutmayı başaran bir algoritma olarak öne çıktığını göstermektedir.

BULGULAR VE SONUÇ

Bu çalışma, Gri Kurt Optimizasyonu (GWO), İyileştirilmiş Gri Kurt Optimizasyonu (IGWO) ve Ensemble Gri Kurt Optimizasyonu (EGWO) algoritmalarının veri madenciliği ve özellik seçimi süreçlerindeki performanslarını kapsamlı bir şekilde karşılaştırmıştır. Veri madenciliği, büyük veri kümelerinde anlamlı bilgileri ortaya çıkarmayı amaçlayan bir süreçtir ve özellik seçimi, bu süreçte analizin doğruluğunu ve verimliliğini artırmada kritik bir rol oynamaktadır. Çalışmada, sınıflandırma hatası, doğruluk, işlem süresi ve seçilen özellik sayısı gibi metrikler kullanılarak bu üç algoritmanın performansları değerlendirilmiştir.

Elde edilen sonuçlar, EGWO algoritmasının genel olarak diğer iki algoritmaya kıyasla daha üstün performans sergilediğini göstermiştir. Sınıflandırma hatası açısından, IGWO en düşük ortalama hata oranını sergilese de, EGWO'nun düşük standart sapma değeri, hata oranlarının daha kararlı bir dağılım gösterdiğini ortaya koymaktadır. Doğruluk performansında EGWO, en iyi değerleri elde ederek, özellikle çok modlu ve karmaşık problemler üzerinde daha başarılı sonuçlar sunmuştur. Ayrıca, EGWO'nun standart sapma değerinin düşük olması, sonuçlarındaki istikrarı desteklemektedir. İşlem süresi bakımından EGWO, ortalama ve en iyi sürelerde en düşük değerleri sergileyerek diğer algoritmalara göre daha hızlı bir optimizasyon süreci sunmuştur. Bu durum, EGWO'nun zaman açısından verimliliğini ve büyük veri setlerinde uygulanabilirliğini artırmaktadır. Ayrıca, özellik seçimi sürecinde EGWO, hem en iyi hem de en kötü durumda daha az özellik seçerek diğer algoritmalara kıyasla daha etkili bir süreç gerçekleştirmiştir. EGWO'nun düşük standart sapma değeri, seçilen özellik sayısındaki tutarlılığı işaret ederken, bu algoritmanın gereksiz özellikleri elemekteki başarısını vurgulamaktadır.

Sonuç olarak, EGWO algoritması, doğruluk, işlem süresi ve özellik seçimi metriklerinde üstün performans sergilemiştir. Özellikle zaman verimliliği ve kararlılığı ile veri madenciliği ve özellik seçimi süreçlerinde önemli bir araç olarak öne çıkmaktadır. Bununla birlikte, IGWO'nun düşük hata oranı ve daha az karmaşıklıkla belirli senaryolarda avantaj sağladığı görülmüştür. Bu bulgular, EGWO'nun farklı veri kümelerinde ve karmaşık problemlerde tercih edilebilecek etkili bir yöntem olduğunu göstermektedir. Gelecekte, EGWO'nun farklı veri setleri ve optimizasyon problemlerinde test edilmesi, algoritmanın daha geniş bir uygulama yelpazesindeki etkinliğini daha iyi anlamayı sağlayacaktır.

REFERANSLAR

- Altan, A., & Karasu, S. (2020). Ayırıştırma Yöntemlerinin Derin Öğrenme Algoritması ile Tanımlanan Rüzgâr Hızı Tahmin Modeli Başarımına Etkisinin İncelenmesi. *European Journal of Science and Technology*. <https://doi.org/10.31590/ejosat.785699>
- Uladi, A. İ., & Arı, E. S. (2023). Büyük Veri, Büyük Veri Analizi ve Uygulama Alanları. *Yönetim Bilişim Sistemleri Dergisi*, 9(1), 1-14.
- Çelik, S., & Akdamar, E. (2018). Büyük Veri Ve Veri Görşelleştirme. *Akademik Bakış Uluslararası Hakemli Sosyal Bilimler Dergisi*, 65, 253-264.
- Dada, E. G., Joseph, S. B., Oyewola, D. O., Fadele, A. A., Chiroma, H., & Abdulhamid, S. M. (2022). Application of Grey Wolf Optimization Algorithm: Recent Trends, Issues, and Possible Horizons. *Gazi University Journal of Science*, 35(2), 485-504. <https://doi.org/10.35378/gujs.820885>
- Emeksiz, C., & Tan, M. (2021). Geliştirilmiş EEMD-EWT Tabanlı Yapay Sinir Ağı Modeli Kullanarak Çok Adımlı Rüzgar Hızı Tahmini. *Avrupa Bilim Ve Teknoloji Dergisi*, 26, 165-173. <https://doi.org/10.31590/ejosat.948661>
- Emre, İ. E., & Selçukcan Erol, Ç. (2017). Veri Analizinde İstatistik mi Veri Madenciliği mi? *Bilişim Teknolojileri Dergisi*, 161-161. <https://doi.org/10.17671/gazibtd.309297>
- Han, K. H., & Kim, J. H. (2002). Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. *IEEE Transactions on Evolutionary Computation*, 6(6), 580-593. <https://doi.org/10.1109/TEVC.2002.804320>
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46-61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- Mosavi, S. K., Jalalian, E., & Gharahchopog, F. S. (n.d.). A Comprehensive Survey Of Grey Wolf Optimizer Algorithm And Its Application. In *International Journal of Advance Robotics & Expert Systems (JARES)*, 1(6), 23-45.
- Onuralp Gökalp, M., Kayabay, K., Çoban, S., Yandık, Y. B., & Erhan Eren, P. (2018). *Büyük Veri Çağında İşletmelerde Veri Bilimi*, Ankara.
- Rajabioun, R. (2011). Cuckoo optimization algorithm. *Applied Soft Computing Journal*, 11(8), 5508-5518. <https://doi.org/10.1016/j.asoc.2011.05.008>
- Rajwar, K., Deep, K., & Das, S. (2023). An exhaustive review of the metaheuristic algorithms for search and optimization: taxonomy, applications, and open challenges. *Artificial Intelligence Review*, 56(11), 13187-13257. <https://doi.org/10.1007/s10462-023-10470-y>
- Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. *İstanbul Commerce University Journal of Science*, 11(21), 1-23.
- Storn, R. & Price, K. (1997). Differential Evolution-A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. In *Journal of Global Optimization*, 11, 341-359.

- Tekerek, A. (2011). *Data Mining Processes and Open Source Data Mining Tools*. <https://www.researchgate.net/publication/325441295>
- Xue, B., Zhang, M., & Browne, W. N. (2013). Particle swarm optimization for feature selection in classification: A multi-objective approach. *IEEE Transactions on Cybernetics*, 43(6), 1656–1671. <https://doi.org/10.1109/TSMCB.2012.2227469>
- Zhao, X., Zhang, X., Cai, Z., Tian, X., Wang, X., Huang, Y., Chen, H., & Hu, L. (2019). Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients. *Computational Biology and Chemistry*, 78, 481–490. <https://doi.org/10.1016/j.compbiolchem.2018.11.017>
- Arora, S. & Joshi, H. (2017). Enhanced grey wolf optimisation algorithm for constrained optimisation problems. *International Journal of Swarm Intelligence*, 3,126. <https://doi.org/10.1504/IJSI.2017.10008734>
- Long, W., Jiao, J., Liang, X., & Tang, M. (2018). An exploration-enhanced grey wolf optimizer to solve high-dimensional numerical optimization. *Engineering Applications of Artificial Intelligence*, 68, 63-80. <https://doi.org/10.1016/j.engappai.2017.10.024>
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46-61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- Yu, X., & Wu, X. (2022). Ensemble grey wolf Optimizer and its application for image segmentation. *Expert Systems with Applications*, 209, 118267. <https://doi.org/10.1016/j.eswa.2022.118267>

BÖLÜM 2

YALAN TESPİT SİSTEMLERİ: BİR KUANTUM UYGULAMASI

Gamzepelin AKSOY¹

Murat KARABATAK²

¹ Dr., Isparta Uygulamalı Bilimler Üniversitesi Teknoloji Fakültesi
Bilgisayar Mühendisliği Bölümü. gamzepelina@gmail.com ORCID No:
0000-0002-5328-2983

² Prof. Dr., Fırat Üniversitesi Teknoloji Fakültesi Yazılım Mühendisliği
Bölümü. mkarabatak@firat.edu.tr ORCID No: 0000-0002-6719-7421
Bu çalışma Gamzepelin AKSOY'un EEG Sinyallerinin Sınıflandırılmasında
Kuantum Tabanlı Karar Destek Sisteminin Gerçekleştirilmesi isimli
doktora tezinden üretilmiştir.

1. GİRİŞ

Yalan, insanlık tarihinin başlangıcından itibaren bireyler arası ilişkilerde güveni zedeleyen en temel sorunlardan biri olmuştur. Manipülasyon, çıkar sağlama veya zor durumlardan kurtulma amacıyla yalan söyleme davranışı, toplumsal düzenin bozulmasına ve adaletin zedelenmesine neden olmaktadır. Özellikle hukuk, güvenlik ve finans gibi kritik alanlarda yalanın doğru bir şekilde tespit edilmesi büyük önem taşımaktadır.

Modern teknoloji ve bilimsel ilerlemeler, yalan tespiti yöntemlerini fizyolojik ölçümlerden bilişsel ve dil temelli analizlere kadar geniş bir yelpazede geliştirmiştir. Başlangıçta suçlu sorgulamalarında kullanılan yalan makineleri, zamanla daha karmaşık ve çok boyutlu sistemlere dönüşmüştür. Beyin sinyallerinin, dil yapı analizlerinin ve makine öğrenimi algoritmalarının bu süreçlere dahil edilmesiyle yalan tespitindeki başarı oranları önemli ölçüde artmıştır. Bununla birlikte, kuantum makine öğrenme yöntemlerinin bu alandaki potansiyeli henüz araştırılmamıştır.

Bu çalışma, yalan tespiti yöntemlerinin tarihsel bağlamdan günümüzün ileri teknolojilerine kadar olan evrimini ele almakta ve Varyasyonel Kuantum Sınıflandırıcı (VQC) kullanılarak yalan tespiti üzerine bir araştırma sunmaktadır. Halka açık bir veri seti üzerinde gerçekleştirilen analizler ile VQC algoritmasının bu alandaki performansı değerlendirilmiştir.

2. TARİHSEL GELİŞİM SÜRECİ

1920’li yılların başında Larson tarafından modern yalan makinesi icat edilmiştir (Grubin & Madsen, 2005). Yalan makinesinin ortaya çıkışının ardından özellikle adli vakalar üzerinde incelemeler yapılmıştır. Bu yöntem ile suçlulardan kalp atış hızı, solunum ve cilt iletkenliği verileri fizyolojik bir kayıt cihazı kullanılarak elde edilmiştir. Yalan tespit cihazı kullanılan araştırmalarda aynı zamanda sorgulama teknikleri ve yalan testlerinden de faydalanılmıştır (Saxe & Ben-Shakhar, 1999). 1930’larda fizyolojik yalan tespiti ile başlayan süreç 1950’li yıllarda sözsüz yalan tespiti ile devam etmiştir. 1970’lerde yalan ile belirli kelimeler arasındaki ilişkinin incelenmesiyle yapılan çalışmalar ile sözlü yalan tespiti araştırmaları başlamıştır. 1980’lerin sonlarına doğru “kriter tabanlı içerik analizi” yöntemi oluşturulmuştur (Vrij vd., 2022). Bu yöntem yazılı metinlerin doğruluğunun değerlendirilmesini sağlayan bir araçtır. Dünyanın birçok bölgesinde ceza mahkemelerinde kullanılmıştır. Yalan söylemenin, doğruyu söylemekten daha zor olduğu ve yalan söyleyenlerin, başkaları üzerinde bıraktıkları itibayı daha fazla umursadıkları varsayımı üzerine bu görüş geliştirilmiştir (Vrij & Mann, 2006). Bu yöntemi ‘gerçeklik izleme kriterleri’ ve ‘ifade analizi’ takip etmiştir. Gerçeklik izleme, algıyla oluşturulan gerçek anılar ile hayal gücü tarafından üretilen anılar arasındaki ayrımı yapma sürecini ifade eden bir terim olarak kullanılmaktadır (Subramaniam vd., 2018). Psikolojik rahatsızlığı olan kişilerde ve ilerleyen yaşlarda gerçeklik al-

gısında bozulmalar meydana gelmektedir. Bu sebeple genelde klinik psikoloji alanında ve adli arařtırmalarda sıklıkla kullanılarak gerçeğliğin test edilmesi sağlanmaktadır (Gancedo vd., 2021). İfade analizi ise kelimelerin anlamına ve dil bilgisi kurallarına odaklanmaktadır. Ancak insanların yaşadıkları çevre ve kültürel etkiler düşünöldüğünde yanlış kararlara sebep olabilmektedir (Ma & Lin, 2015). 1980'lerde yalan tespiti için beyin aktivitesi sonucu elde edilen veriler kullanılmaya başlanmıştır. Yapay zekâ ve makine öğrenme alanında yaşanan gelişmeler ile birlikte bu alanda yapılan analizler bilgisayar ortamında geliştirilmeye başlanmıştır.

Yalan makinesi ile elde edilen verilerin incelenmesi ile başlayan süreç, yerini 2000'li yıllarda doğal dil işleme algoritmalarında yaşanan gelişmeler ve makine öğrenme algoritmalarının yaygın kullanımı sonucunda bilgisayar destekli sistemlere bırakmıştır. Bu sistemlerde kalp atış hızı, solunum gibi verilerin yanı sıra yüz ifadeleri gibi biyometrik veriler, ses düzeyi ve tonlama, dil yapılarının kullanımı ve EEG sinyalleri gibi bilgilerin incelenmesi ile bu çalışmalarda daha başarılı sonuçlar elde edilmiştir.

Bu alanda beyin sinyalleri kullanılarak yapılan ilk çalışmalardan biri Rosenfeld vd. tarafından gerçekleştirilmiştir. Bir deney sistemi kurularak, bireylerin uyarılara karşı olay ilişkili potansiyellerinin kaydedilmesi sağlanmıştır. Bu kayıtlar incelenerek, bireylerin yalan söyleyip söylemediklerini belirlemek amaçlanmıştır (Rosenfeld vd., 1987). Farwell, EEG sinyalleri ile gizli bilgilerin ortaya çıkarılması ve doğruluğın tespiti üzerine geliştirdiğı sisteme 'beyin parmak izi' adını vermiştir. Bu sistemle, mevcut bilgiyi tespit ederek istatistiksel olarak doğruluğunu incelemiştir (Farwell, 2012). Makine öğrenme ve derin öğrenme tekniklerinin bu alanda uygulanmasıyla birlikte, yalan tespiti için bilgisayar tabanlı otomatik sistemler geliştirilmiştir.

Karnati vd., yanıltıcı davranışları otomatik olarak tespit etmek amacıyla LieNet adını verdikleri bir sistem geliştirmiştir. Bu sistem, video, ses ve EEG sinyallerinden özellikler çıkarmaya olanak tanımakta ve yeni bir derin evrişimli sinir ağı ortaya koymaktadır. LieNet'in etkinliğı, 'Bag-of-Lies (BoL)', 'real-life (RL) trail' ve 'Miami University Deception Detection (MU3D)' veritabanlarında doğrulanmıştır. Bag-of-Lies veritabanında %95,91 başarı elde edilmiştir (Karnati vd., 2022). Anwar vd. ise gizli bilgilerin tespiti için yalan makinesine alternatif olarak beyin aktivitesinin kullanılmasını incelemiştir. Makine öğrenmesi tabanlı bir sınıflandırma algoritması geliştirmek için giyilebilir bir EEG cihazı kullanmıştır. Çalışmada, verilerin boyutunu azaltmak için 3. seviye ayrık dalgacık dönüşümü ve temel bileşen analizi kullanılmıştır. Sınıflandırma işlemi, SVM kullanılarak gerçekleştirilmiş ve %83'lük bir doğruluk elde edilmiştir (Anwar vd., 2019). Javaid vd., yalan tespiti için görsel, ses ve EEG bilgilerinin önemini arařtırdıkları bir çalışma gerçekleştirmiştir. Çalışmalarında, sistem yüz hareketlerini görsel olarak incelerken evrişimli sinir ağı kullanılmıştır. Ses sinyallerinin frekans dağılımlı spektrogramları çıkarılarak

CNN modelinde analiz edilmiştir. EEG sinyallerine ise çift yönlü uzun kısa süreli bellek uygulanarak, geçmiş ve gelecek sinyaller arasındaki ilişkiler incelenmiştir. Deneyler ‘Bag-Of-Lies’ veri setinde gerçekleştirilmiş ve tüm bileşenlerin dâhil edildiği analizde %83.5 doğrulukla yalan tespiti yapılmıştır (Javaid vd., 2022).

Bir başka yalan tespiti çalışmasında, EEG verileri BoL veri setinden elde edilmiştir. EEG verilerine ayrık dalgacık dönüşümü uygulanmış ve sinyaller alt sinyallere ayrılmıştır. NCP derin öğrenme modeli ve LSTM yönteminin kullanıldığı hibrit bir sistem geliştirilmiş ve yalan tespiti deneylerinde kullanılmıştır. Bu deneyler sonucunda %97.88’lik bir performans elde edilmiştir (Aslan, Baykara, & Alakuş, 2024).

Gao vd., suçlu ve masumlardan oluşan 28 kişilik denek grubunda yalan söylemeyi sınıflandırmak için yeni bir yöntem önermişlerdir. EEG sinyallerinden elde edilen her bir özelliğin ayırt edici etkisini belirlemek için F-skorunu kullanmıştır. ‘F-skor_ELM’ olarak adlandırdıkları bu yöntemde, ortalama test doğruluğunun %98.72 olduğunu belirtmişlerdir (Gao vd., 2013). BrainVision kayıt cihazı ve analizörü ile elde edilen EEG sinyallerinin kullanıldığı bir yalan tespit sistemi çalışmasında, bant geçiren filtre kullanılmıştır. Dalgacık Paket Dönüşümü uygulanan sinyaller, 10 gizli katmanlı ve iki otomatik kodlayıcıya sahip bir Derin Sinir Ağına girdi olarak verilmiştir. Yapılan deneyler sonucunda %95 doğruluk elde edilmiştir (Edla vd., 2021). Gao vd., yalan tespiti sırasında doğruluğu artırabilmek için gürültüden arındırılmış P3 bileşeni ve makine öğrenimine dayalı yeni bir gizli bilgi testi yöntemi önermişlerdir. Çalışmada, suçlu ve masum katılımcılar olarak 30 kişi seçilmiş ve üç farklı uyaran paradigması uygulanmıştır. Bağımsız bileşen analizi ile artefaktları ayırmak için kullanılmıştır. Oluşturulan yeni dalga formlarından zaman, frekans ve dalgacıklara dayalı 3 grup özellik çıkarılmıştır. Bu özellikler, iki sınıf örneği kullanılarak SVM sınıflandırıcısında değerlendirilmiş ve %84.29’luk bir başarı elde edilmiştir (Gao vd., 2012). Farahani ve Morani, gizli bilgilerin çoklu duyusal tespitini sağlamak amacıyla bir çalışma gerçekleştirmişlerdir. Sınıflandırıcı olarak Genetik Destek Vektör Makinesi kullandıkları bu çalışmada, otonom özellikler ve bunların kombinasyonları kullanıldığında %95.45’lik bir doğruluk oranına ulaşıldığını belirtmişlerdir (Farahani & Moradi, 2017). Alarfaj ve Mahmoud, HD-72 Cognionics kulaklık kullanarak elde ettikleri EEG sinyallerinde Fourier serisi dönüşümünü uygulamıştır. EEG sinyallerini hem zamansal hem de mekânsal boyutlarda analiz etmiştir. Veri setini %80 eğitim ve %20 test olarak ayırdıkları çalışmada, model performansını doğrulamak için çapraz doğrulama yöntemini kullanmıştır. V-TAM dikkat modeline dayalı bir analiz platformunu yalan tespiti için önermişlerdir ve önerilen yöntemin %98.5 hassasiyetle yalan tespiti analizi gerçekleştirdiğini belirtmiştir (Alarfaj & Ahmed, 2022).

3. MATERYAL METOT

3.1. LieWaves Veri Seti

LieWaves veri seti, AF3, T7, Pz, T8 ve AF4 kanallarını içeren 5 kanallı Emotiv Insight cihazı kullanılarak, sağlıklı 27 denekten toplanan verileri içermektedir (Aslan vd., 2023). Denekler, doğruyu söylemeyi ve yalan söylemeyi içeren iki farklı deneye katılmıştır. Deneyler sırasında görsel uyaran olarak 10 farklı tespih kullanılmıştır. Bu tespihlerin 5'i birinci deneyde, diğer 5'i ise ikinci deneyde kullanılmıştır.

Uyaranlar, video klipler olarak hazırlanmış ve katılımcılara projeksiyon cihazıyla 2-3 metre mesafeden duvara yansıtılarak izlettirilmiştir. Deneyler, dış gürültü ve parazitlerden arındırılmış sessiz bir ortamda gerçekleştirilmiş, elektrot yerleşimi için uluslararası 10-20 sistemi kullanılmıştır. Tablo 1'de veri setine ilişkin genel bilgiler sunulmaktadır.

Tablo 1. *LieWaves veri seti hakkında genel bilgiler*

Sinyal Tipi	EEG sinyalleri (Zaman Serisi)
Kullanım Amacı	Yalan Tespiti
Kullanılan Cihaz	Emotiv Insight
Kanal Sayısı	5
Katılımcı Sayısı	27
Katılımcı Yaş Ortalaması	23,1
Deney Sayısı	2
Örnek Sayısı	9600
Uyaran Türü	Resim

EEG kayıtları sırasında, çalışmada kullanılan tüm tespihler, her biri 2 saniye boyunca gösterilecek şekilde hazırlanmış, 25 kareden oluşan video kayıtları ile sunulmuştur. Her bir sunumda, önce 1 saniyelik siyah ekran gösterilmiş, ardından 2 saniyelik tespih görüntüsü sunulmuş ve toplamda 75 saniyelik EEG verisi kaydedilmiştir. Bu süreç, her katılımcı için 5 kez tekrarlanmıştır.

Deney öncesinde katılımcılardan, bir kutu içerisindeki tespihlerden ikisini seçmeleri istenmiştir. Daha sonra, hangi deneyde doğruyu, hangi deneyde yalan söyleyeceklerine karar vermeleri talep edilmiştir. Katılımcılara bu kararlarını vermeleri ve deney için konsantre olmaları amacıyla 20 saniyelik süre tanınmıştır. İlk deneyde yalan söylemeye karar veren katılımcılar, seçtikleri tespih resmini gördüklerinde “HAYIR” yazan sol tuşa basmıştır. Doğruyu söylemeye karar verdiklerinde ise “EVET” yazan sağ tuşa basmıştır. Bu süreç boyunca EEG sinyalleri sürekli olarak kaydedilmiş ve bir veri seti oluşturulmuştur (Aslan, Baykara, & Alakus, 2024).

3.2. Ön İşlem Süreci

Deneyisel çalışmada kullanılan veri seti, analizlerde kullanılmak üzere çeşitli ön işlem adımlarından geçirilerek hazırlanmıştır. Bu süreçte, EEG sinyalleri 5 alt banda ayrılmış ve her bir alt band üzerinde belirli istatistiksel analizler gerçekleştirilmiştir.

EEG sinyallerinin analizi ve ham verilerden özellik çıkarımı için, Dalgacık Dönüşümü (Wavelet Transform, WT) kullanılmıştır. WT, özellikle zamanla değişen ve durağan olmayan biyomedikal sinyallerin analizi için etkili bir yöntemdir. Bu yöntemde, orijinal sinyal, dalgacıklar olarak adlandırılan yapı taşlarıyla temsil edilerek bir dizi dalgacık katsayısına ayrıştırılır. Böylece sinyal, daha az parametre ile sıkıştırılarak farklı frekans bantlarında detaylı şekilde analiz edilebilir.

İlk adımda, sinyal düşük frekanslı bileşenleri temsil eden yaklaşık katsayılarına ve yüksek frekanslı bileşenleri temsil eden detay katsayılarına ayrılmıştır. Daha sonra, ardışık yüksek geçiş ve alçak geçiş filtrelemeleri ile bu katsayılar daha ayrıntılı bileşenlere ayrıştırılmıştır. Bu işlem, sinyalin daha iyi anlaşılmasını ve işlenmesini sağlayarak WT'yi güçlü bir analiz yöntemi haline getirmektedir (Al-Fahoum & Al-Fraihat, 2014).

Son olarak, ayrık dalgacık dönüşümü tamamlandıktan sonra sinyaller üzerinde çeşitli istatistiksel teknikler uygulanmıştır. Bu tekniklere ilişkin kullanılan formüller, Tablo 2'de gösterilmektedir.

Tablo 2. *Özellik çıkarımı aşamasında kullanılan istatistiksel yöntemler*

İstatistiksel Yöntemler	Formüller
Ortalama	$\mu_{ti} = 1/N \sum_{j=i}^N M_{ij}$
Standart sapma	$SS = \sqrt{\frac{1}{N} \sum_{n=1}^N (x(n) - \mu_n)^2}$
Varyans	$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2$
Logaritmik Entropi	$H_{LogEn}(x) = - \sum_{i=0}^{N-1} (log_2(p_i(x)))^2$
Shanon Entropi	$H_{SEn}(x) = - \sum_{i=0}^{N-1} p_i(x)^2 (log_2(p_i(x)))^2$

$$E = \text{Varyans}(x(t))$$

$$H = \sqrt{\frac{\text{Varyans}\left(\frac{dx(t)}{dt}\right)}{E}}$$

Hjort Parametreleri

$$K = \sqrt{\frac{H\left(\frac{dx(t)}{dt}\right)}{H(x(t))}}$$

3.3. VQC Algoritması

Varyasyonel kuantum sınıflandırıcı, gürültülü orta ölçekli kuantum cihazları (NISQ) üzerinde sınıflandırma problemleri için yaygın kullanılan, gözetimli bir kuantum makine öğrenimi algoritmasıdır. Havlicek vd., bu cihazlarda ekstra hata düzeltme yöntemlerine ihtiyaç duymadan keşifsel sonuçlar elde etmeyi sağlayan VQC modelini önermiştir. VQC, gürültülü verileri yinelemeli cihaz ölçümleriyle optimizasyon hesaplamalarına entegre ederek maliyet fonksiyonunu hesaplar ve hataları azaltır (Maheshwari vd., 2022).

VQC, kuantum bilgisayarlar üzerinde $f(x, \theta) = y$ fonksiyonunu temel alır; burada giriş verisi x , bir kuantum devresi aracılığıyla kuantum durumuna kodlanır (Schuld vd., 2020). VQC'nin çalışma aşamaları ikiye ayrılır: Durum hazırlama ve sınıflandırıcı oluşturma. İlk aşamada, klasik veriler, kuantum durumları olarak kodlanır ve kuantum özellik haritası ile Hilbert uzayında temsil edilir. İkinci aşamada, parametrelili bir kuantum devresi (U_θ) kullanılarak giriş verisinin durumu değiştirilir. Bu operatör, modelin tahmin sürecinde temel rol oynar. Giriş verisinin kuantum durumunu, bir başka kuantum durumuna dönüştürmek için üniter kapılar kullanılır. Kodlanan veriler üzerinde işlem yapıldıktan sonra ölçüm gerçekleştirilir. Bu ölçüm, çıkışın 0 veya 1 olasılıklarıyla tahmin edilmesini sağlar ve ikili sınıflandırma için temel oluşturur. Varyasyonel devrenin parametreleri klasik yöntemlerle optimize edilerek sınıflandırma doğruluğu artırılır (Zeguendry vd., 2023). Böylece, VQC algoritması, NISQ cihazlarında sınıflandırma problemlerine etkili bir çözüm sunar.

4. DENEYSEL ÇALIŞMA

Bu çalışmada, kuantum tabanlı varyasyonel kuantum sınıflandırıcı algoritmasının yalın tespitindeki performansını değerlendirmek amacıyla, P300 verilerinin bulunduğu *Lie Waves* veri seti kullanılmıştır. Çalışmada, yalın tespiti için örüntü tanıma süreci beş ana aşamada gerçekleştirilmiştir:

Ön İşleme: EEG sinyalleri üzerinde ayrık dalgacık dönüşümü uygulanmıştır.

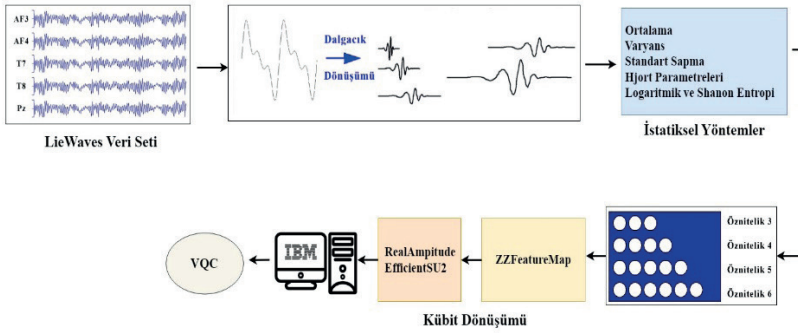
Özellik Çıkarma: Dalgacık dönüşümü ile elde edilen sinyallerden belirli özellikler (ör. ortalama, varyans, entropi) çıkarılmıştır.

Boyut İndirgeme: PCA (Temel Bileşen Analizi) yöntemi kullanılarak, veriler farklı boyutlara indirgenmiş ve analiz için optimize edilmiştir.

Özellik Haritası: Boyutu indirgenmiş veriler, Hilbert uzayında tanımlanabilmeleri için uygun kuantum özellik haritalarıyla işlenmiştir.

Sınıflandırma: Parametrik kuantum devreleriyle oluşturulan kuantum tabanlı modellerin doğrulukları değerlendirilmiştir.

Şekil 1'de EEG sinyallerinde VQC algoritması ile gerçekleştirilen deneysel çalışmanın şematik görünümüne yer verilmiştir.



Şekil 1. Önerilen modelinin şematik görünümü

EEG sinyallerini içeren LieWaves veri setinde, sağ frontal kutup bölgesinden AF4, sol frontal kutup bölgesinden AF3, sağ temporal bölgeden T8, sol temporal bölgeden T7 ve parietal bölgeden Pz olmak üzere 5 kanaldan elde edilen veriler bulunmaktadır. Çalışmanın ilk aşamasında, ayrık dalgacık dönüşümü ile sinyaller farklı frekans bantlarına ayrılmış ve bu bantlardaki özellikler istatistiksel yöntemlerle işlenmiştir.

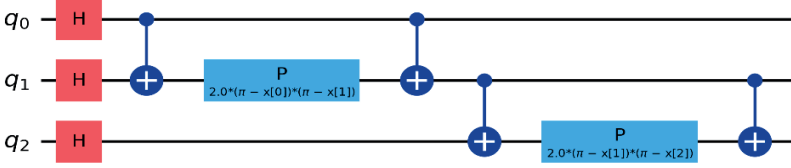
Çalışmada uygulanan istatistiksel yöntemler; ortalama, varyans, standart sapma, Hjort parametreleri, logaritmik entropi ve Shannon entropisi olarak belirlenmiştir. Bu yöntemlerle, toplamda biri çıkış değeri olmak üzere 36 özneliğe sahip yeni bir veri seti oluşturulmuştur.

Analizlerin etkin bir şekilde gerçekleştirilmesi için, boyut indirgeme adımlarında Temel bileşen analizi (PCA) yöntemi kullanılmıştır. PCA, veri setinin öznelik sayısını optimize ederek 3 ile 6 arasında değişen dört farklı boyuta indirgenmiştir.

Veri seti, eğitim ve test verilerine ayrılmadan önce rastgele bölüme

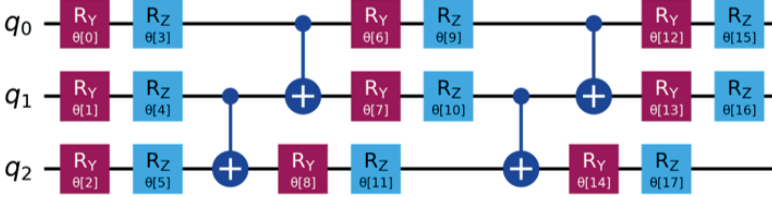
yöntemi uygulanmış ve veri setinin %70'i eğitim, %30'u test verisi olarak belirlenmiştir. Analizler, Jupyter Notebook ortamında Qiskit kütüphaneleri ve Statevector simülatörü kullanılarak gerçekleştirilmiştir.

Varyasyonel Kuantum Sınıflandırıcı (VQC) algoritmasının veri setindeki başarımını değerlendirmek için ZZFeatureMap kullanılarak veriler kübit formuna dönüştürülmüştür. Bu özellik haritası, 3 kübit ve tek tekrar için oluşturulmuş olup, ilgili devre yapısı Şekil 2'de gösterilmektedir. Bu süreç, kuantum tabanlı sınıflandırma modeli için gerekli olan kuantum verilerinin hazırlanmasını sağlamıştır.

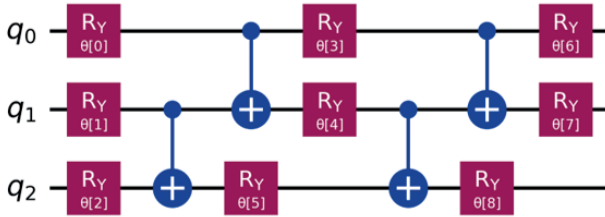


Şekil 2. ZZFeaturemap

VQC algoritmasının önemli bir bileşeni, parametrik kuantum devreleridir. Bu çalışmada, kuantum hesaplamada yaygın olarak kullanılan RealAmplitude ve EfficientSU2 devreleri tercih edilmiştir. Her iki devre, 2 tekrar içerecek şekilde yapılandırılmıştır. Bu devreler, kuantum algoritmasının esnekliğini ve performansını artırmak amacıyla farklı parametrelerle optimize edilmiştir. Kullanılan devrelerin şematik görünümleri, Şekil 3'te sunulmaktadır.



EfficientSU2



RealAmplitudes

Şekil 3. VQC algoritmasında kullanılan parametrik kuantum devreleri

VQC algoritmasının eğitim sürecinde, kullanılan RealAmplitude ve EfficientSU2 devrelerinden elde edilen kapı açılarını optimize etmek için COBYLA (Constrained Optimization By Linear Approximations) algoritması tercih edilmiştir. Bu optimizasyon süreci, modelin performansını artırmak amacıyla 100 iterasyon boyunca sürdürülmüştür.

5. SONUÇLAR

Bu çalışmada, varyasyonel kuantum sınıflandırıcı algoritmasının yalan tespiti üzerindeki performansı, farklı EEG kanalları ve kübit sayıları dikkate alınarak analiz edilmiştir. Çalışmada kullanılan parametrik kuantum devreleri (RealAmplitude ve EfficientSU2), doğruluk oranlarını artırma potansiyelleri açısından değerlendirilmiştir.

Farklı kanallar ve kübit sayıları için iki farklı devre yapısı ile elde edilen doğruluk oranları, analiz edilerek Tablo 3'te sunulmuştur. Bu veriler, algoritmanın performansının hem devre yapısına hem de kullanılan kanal ve kübit sayısına bağlı olarak değiştiğini göstermektedir.

Tablo 3. VQC algoritmasının sınıflandırma performansı sonuçları

Kanal	Kübit Sayısı	RealAmplitude Doğruluk Oranları (%)		EfficientSU2 Doğruluk Oranları (%)	
		Ortalama	Maksimum	Ortalama	Maksimum
AF3	3	68.65	88.24	75.29	94.12
AF3	4	63.35	76.47	66.88	82.35
AF3	5	59.06	76.47	69.29	88.24
AF3	6	58.71	82.35	68.18	88.24
AF4	3	78.24	82.35	74.65	94.12
AF4	4	61.24	82.35	73.24	88.24
AF4	5	73.59	94.12	66.82	88.24
AF4	6	58.24	82.35	53.71	76.47
Pz	3	43.18	58.82	40.65	70.59
Pz	4	74.65	82.35	75.12	88.24
Pz	5	71.24	94.12	71.29	88.24
Pz	6	57.18	76.47	62.24	88.24
T7	3	73.53	76.47	70.65	70.65
T7	4	71.65	71.65	73.18	73.18
T7	5	67.94	67.94	67.59	67.59
T7	6	61.82	61.82	67.94	67.94
T8	3	70.18	76.47	56.94	70.59
T8	4	71.00	88.24	72.47	82.35
T8	5	57.71	94.12	69.47	69.47
T8	6	60.53	60.53	60.71	60.71

Tablo 3'te sunulan verilere göre:

- En yüksek doğruluk oranı, AF3 ve AF4 kanalları için 3 kübitli yapı ve EfficientSU2 parametrik devresi kullanıldığında %94.12 olarak ölçülmüştür.

- Benzer şekilde AF4, Pz ve T8 kanalları, 5 kübitli yapı ve RealAmplitude devresi ile %94.12 doğruluk oranına ulaşmıştır. Bu sonuçlar, bu kanalların yalan tespiti açısından diğerlerine kıyasla daha etkili olduğunu göstermektedir.

- T7 kanalı, diğer kanallara göre daha düşük bir performans sergilemiş ve maksimum doğruluk oranları genellikle %61-%76 aralığında kalmıştır.

EfficientSU2 devresi, çoğu durumda RealAmplitude devresine göre daha tutarlı sonuçlar sağlamış ve özellikle 3 kübitli yapıda optimum performansı göstermiştir. Ayrıca, AF3 ve AF4 kanallarının farklı kübit sayılarında daha yüksek performans sergilediği gözlenmiştir.

Sonuçlar, VQC algoritmasının farklı devre yapılandırmaları ve EEG kanallarına göre özelleştirilebilir olduğunu ve kuantum tabanlı yaklaşımların yalan tespiti gibi karmaşık problemlerde etkili bir çözüm sağlayabileceğini ortaya koymuştur. Çalışma, bu alandaki araştırmalar için bir referans niteliği taşımaktadır.

KAYNAKÇA

- AlArfaj, A., & Ahmed, H. (2022). A Deep Learning Model for EEG-Based Lie Detection Test Using Spatial and Temporal Aspects. *Computers, Materials & Continua*, 73(3), 5655-5669. <https://doi.org/10.32604/cmc.2022.031135>
- Al-Fahoum, A. S., & Al-Fraihat, A. A. (2014). Methods of EEG Signal Features Extraction Using Linear Analysis in Frequency and Time-Frequency Domains. *ISRN Neuroscience*, 2014, 730218. <https://doi.org/10.1155/2014/730218>
- Anwar, S., Batool, T., & Majid, M. (2019). Event Related Potential (ERP) based Lie Detection using a Wearable EEG headset. 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 543-547. <https://doi.org/10.1109/IBCAST.2019.8667131>
- Aslan, M., Baykara, M., & Alakus, T. (2024). LieWaves: Dataset for lie detection based on EEG signals and wavelets. *Medical and Biological Engineering and Computing*, 62(5). <https://doi.org/10.1007/s11517-024-03021-2>
- Aslan, M., Baykara, M., & Alakuş, T. B. (2023). LieWaves. 1. <https://doi.org/10.17632/5gzxb2bzs2.1>
- Aslan, M., Baykara, M., & Alakuş, T. B. (2024). LSTMNCP: Lie detection from EEG signals with novel hybrid deep learning method. *Multimedia Tools and Applications*, 83(11), 31655-31671. <https://doi.org/10.1007/s11042-023-16847-z>
- Edla, D. R., Dodia, S., Bablani, A., & Kuppili, V. (2021). An Efficient Deep Learning Paradigm for Deceit Identification Test on EEG Signals. *ACM Transactions on Management Information Systems*, 12(3), 25:1-25:20. <https://doi.org/10.1145/3458791>
- Farahani, E. D., & Moradi, M. H. (2017). Multimodal detection of concealed information using Genetic-SVM classifier with strict validation structure. *Informatics in Medicine Unlocked*, 9, 58-67. <https://doi.org/10.1016/j.imu.2017.05.004>
- Farwell, L. A. (2012). Brain fingerprinting: A comprehensive tutorial review of detection of concealed information with event-related brain potentials. *Cognitive Neurodynamics*, 6(2), 115-154. <https://doi.org/10.1007/s11571-012-9192-2>
- Gancedo, Y., Fariña, F., Seijo, D., Vilariño, M., & Arce, R. (2021). Reality Monitoring: A Meta-analytical Review for Forensic Practice. *European Journal of Psychology Applied to Legal Context*, 13(2), 99-110. <https://doi.org/10.5093/ejpalc2021a10>
- Gao, J., Lu, L., Yang, Y., Yu, G., Na, L., & Rao, N. (2012). A novel concealed information test method based on independent component analysis and support vector machine. *Clinical EEG and Neuroscience*, 43(1), 54-63. <https://doi.org/10.1177/1550059411428715>
- Gao, J., Wang, Z., Yang, Y., Zhang, W., Tao, C., Guan, J., & Rao, N. (2013). A Novel Approach for Lie Detection Based on F-Score and Extreme Learning Machine. *PLOS ONE*, 8(6), e64704. <https://doi.org/10.1371/journal.pone.0064704>

- Grubin, D., & Madsen, L. (2005). Lie detection and the polygraph: A historical review. *The Journal of Forensic Psychiatry & Psychology*, 16(2), 357-369. <https://doi.org/10.1080/14789940412331337353>
- Javaid, H., Dilawari, A., Khan, U. G., & Wajid, B. (2022). EEG Guided Multimodal Lie Detection with Audio-Visual Cues. 2022 2nd International Conference on Artificial Intelligence (ICAI), 71-78. <https://doi.org/10.1109/ICAI5435.2022.9773469>
- Karnati, M., Seal, A., Yazidi, A., & Krejcar, O. (2022). LieNet: A deep convolution neural Network framework for detecting deception. 971-984. <https://doi.org/10.1109/TCDS.2021.3086011>
- Ma, D., & Lin, D. (2015). Statement Analysis of Deception Detection. *Open Access Library Journal*, 02(10), Article 10. <https://doi.org/10.4236/oalib.1101975>
- Maheshwari, D., Sierra-Sosa, D., & Garcia-Zapirain, B. (2022). Variational Quantum Classifier for Binary Classification: Real vs Synthetic Dataset. *IEEE Access*, 10, 3705-3715. <https://doi.org/10.1109/ACCESS.2021.3139323>
- Rosenfeld, J. P., Nasman, V. T., Whalen, R., Cantwell, B., & Mazzeri, L. (1987). Late Vertex Positivity in Event-Related Potentials as a Guilty Knowledge Indicator: A New Method of Lie Detection. *International Journal of Neuroscience*, 34(1-2), 125-129. <https://doi.org/10.3109/00207458708985947>
- Saxe, L., & Ben-Shakhar, G. (1999). Admissibility of polygraph tests: The application of scientific standards post- Daubert. *Psychology, Public Policy, and Law*, 5(1), 203-223. <https://doi.org/10.1037/1076-8971.5.1.203>
- Schuld, M., Bocharov, A., Svore, K. M., & Wiebe, N. (2020). Circuit-centric quantum classifiers. *Physical Review A*, 101(3), 032308. <https://doi.org/10.1103/PhysRevA.101.032308>
- Subramaniam, K., Kothare, H., Mizuiri, D., Nagarajan, S. S., & Houde, J. F. (2018). Reality Monitoring and Feedback Control of Speech Production Are Related Through Self-Agency. *Frontiers in Human Neuroscience*, 12. <https://doi.org/10.3389/fnhum.2018.00082>
- Vrij, A., Granhag, P. A., Ashkenazi, T., Ganis, G., Leal, S., & Fisher, R. P. (2022). Verbal Lie Detection: Its Past, Present and Future. *Brain Sciences*, 12(12), Article 12. <https://doi.org/10.3390/brainsci12121644>
- Vrij, A., & Mann, S. (2006). Criteria-Based Content Analysis: An empirical test of its underlying processes. *Psychology, Crime & Law*. <https://doi.org/10.1080/10683160500129007>
- Zeguendry, A., Jarir, Z., & Quafafou, M. (2023). Quantum Machine Learning: A Review and Case Studies. *Entropy*, 25(2), Article 2. <https://doi.org/10.3390/e25020287>

BÖLÜM 3

GELİŞMİŞ BOYUT İNDİRGEME TEKNİKLERİ: DOĞRUSAL OLMAYAN ALGORİTMALARIN ROLÜ

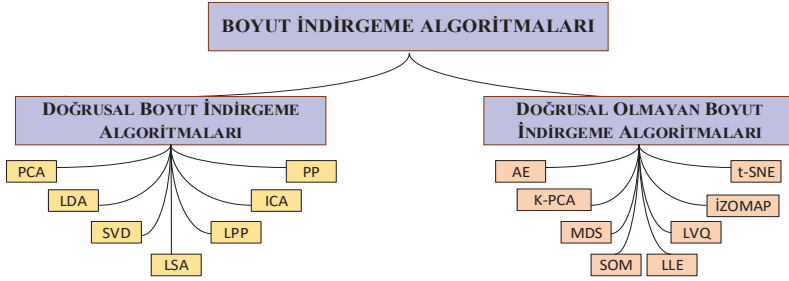
Zeynep Özpolat¹

¹ Muş Alparslan Üniversitesi, Mühendislik-Mimarlık Fakültesi, Yazılım Mühendisliği,
Merkez, Muş, z.ozpolat@alparslan.edu.tr, ORCID: 0000-0003-1549-1220

1. Giriş

Büyük veri kavramının bilgisayar bilimlerine girmesiyle birlikte, bu verilerin saklanması, ön işlem süreçleri ve algoritmalar tarafından işlenmesi konularında çeşitli zorluklar ortaya çıkmıştır. Teknolojik araçlar yardımıyla elde edilen büyük veri içerisinde yer alan alakasız bilgiler ya da tahmin sistemini olumsuz yönde etkileyebilecek girdi unsurları, veri setinin dikkatle şekillendirilmesini gerekli kılmaktadır (Reddy et al., 2020). Bu sorunların üstesinden gelmek için, verilerin boyutlarını küçültmek en etkili yöntemlerden biri olarak kabul edilmektedir. Boyut indirgeme, gereksiz boyutları elimine ederken temel bilgiyi koruyarak algoritmaların daha hızlı ve etkili çalışmasını sağlamaktadır.

Boyut indirgeme algoritmaları iki ana kategoriye ayrılmaktadır: Doğrusal Boyut İndirgeme Algoritmaları ve Doğrusal Olmayan Boyut İndirgeme Algoritmaları (Mendez, 2023). Doğrusal yöntemler, verilerin lineer yapısını korumaya odaklanırken, doğrusal olmayan yöntemler daha karmaşık veri yapılarında etkili sonuçlar sunar. Bu iki kategoriye ait önemli algoritmalar Şekil 1’de özetlenmektedir.



Şekil 1. Boyut indirgeme algoritmaları

PCA: Temel Bileşen Analizi (Principal Component Analysis),
LDA: Lineer Diskriminant Analizi (Linear Discriminant Analysis),
SVD: Tekil Değer Ayrışımı (Singular Value Decomposition), **LSA:**

Gizli Anlamsal Analiz (Latent Semantic Analysis), **PP**: Projeksiyon Takibi (Projection Pursuit), **ICA**: Bağımsız Bileşen Analizi (Independent Component Analysis), **LPP**: Yerel Koruma Çıkarılması (Locality Preserving Projection), **AE**: Otoenkoder (Autoencoder), **K-PCA**: Çekirdek Temelli Temel Bileşen Analizi (Kernel Principal Component Analysis), **MDS**: Çok Boyutlu Ölçekleme (Multidimensional Scaling), **SOM**: Öz Düzenlemeli Haritalar (Self-Organizing Maps), **LVQ**: Öğrenen Vektör Nicemleme (Learning Vector Quantization), **LLE**: Yerel Doğrusal Gömme (Local Linear Embedding), **t-SNE**: t-dağılımlı Stokastik Komşu Yerleştirme (t-distributed Stochastic Neighbor Embedding)

Veri setinin yapısına bağlı olarak doğru yöntemin seçimi, hem işlem etkinliği hem de elde edilen sonuçların doğruluğu açısından büyük önem taşımaktadır. Doğrusal boyut indirgeme yöntemleri, verilerin doğrusal ilişkiler taşıdığı varsayımına dayanmakta ve bu nedenle karmaşık yapılar içeren veri setleri üzerinde sınırlı bir performans göstermektedir. Günümüzde veri setleri giderek daha karmaşık, çok boyutlu ve doğrusal olmayan yapılar sergilediğinden, doğrusal yöntemlerin bu tür veri setlerinde yetersiz kaldığı görülmektedir.

Bu çalışmada, doğrusal boyut indirgeme yöntemlerinden ziyade doğrusal olmayan boyut indirgeme algoritmalarına odaklanılacaktır. Doğrusal olmayan yaklaşımlar, veri yapısını daha iyi temsil edebilmekte ve özellikle karmaşık veri setlerinde daha esnek ve etkili çözümler sunmaktadır. Bu algoritmalar, verilerin özünü yakalamada ve bilgiyi daha düşük boyutlu bir uzaya taşımada doğrusal yöntemlere kıyasla üstün avantajlara sahiptir. Çalışmanın amacı, doğrusal olmayan boyut indirgeme algoritmalarının temellerini, işleyiş mekanizmalarını ve bu yöntemlerin geniş uygulama alanlarını detaylı bir şekilde ele almaktır. Doğrusal olmayan yöntemlerin hem teorik

altyapısı hem de veri analizi süreçlerindeki pratik faydaları, bu algoritmaları veri bilimi, yapay zekâ ve diğer alanlarda kritik bir araç haline getirmektedir. Şimdi, bu yaklaşımların sunduğu avantajları ve kullanım senaryolarını keşfetmek için detaylı bir inceleme yapılacaktır.

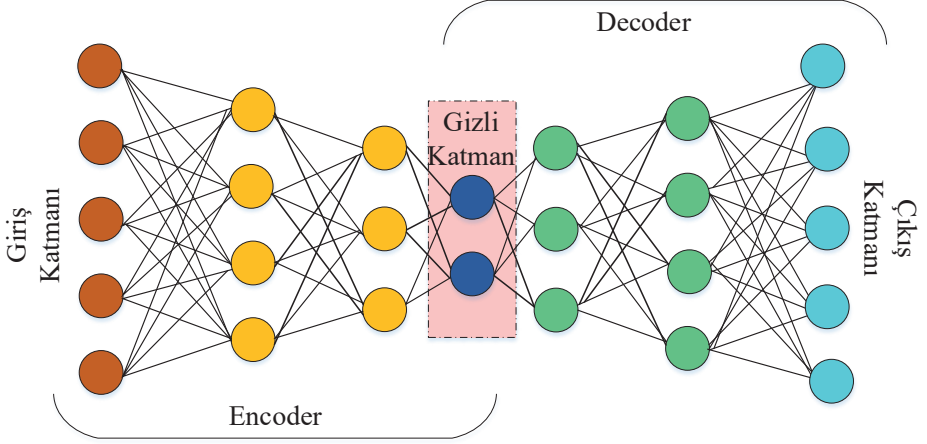
2. Doğrusal Olmayan Boyut İndirgeme Yöntemleri

Doğrusal olmayan boyut indirgeme yöntemleri, özellikle karmaşık ve doğrusal olmayan yapıya sahip veri setlerini işlemek için etkili araçlar sunar. Doğrusal boyut indirgeme yöntemlerine benzer şekilde, bu yöntemler de verilerin boyutunu küçültmeye yönelik bir yaklaşım benimser. Doğrusal olmayan boyut indirgeme yöntemleri, verileri daha düşük boyutlu bir uzaya sıkıştırırken, veri setinin temel yapısını ve özelliklerini korumaya özellikle dikkat eder. Bu sayede, orijinal veri setindeki ilişkiler bozulmadan daha kompakt bir temsil elde edilir. Bu süreç, daha fazla bilgi kaybı yaşanmadan, verinin önemli bileşenlerinin düşük boyutlu bir temsiline ulaşmayı mümkün kılar. Böylece, yüksek boyutlu ve karmaşık veri setleri daha kolay işlenebilir hale gelir ve genellikle daha verimli analiz ve modelleme yapılabilir (Lee & Verleysen, 2007).

2.1. Oto Kodlayıcı

Oto Kodlayıcı (Auto Encoder - AE), derin öğrenme ve yapay sinir ağları temeline dayanan özel bir modeldir. AE'ler, verilen veri kümesini daha düşük bir boyuta indirger ve ardından çıktıyı orijinal veriye en yakın şekilde yeniden yapılandırır (Bank et al., 2021). Bu süreçte, veriyi düşük boyutlu bir temsile dönüştüren kısım "Encoder" olarak adlandırılırken, bu temsilden orijinal verilere yakın çözüm elde eden kısmı ise "Decoder" olarak tanımlanır. Sinir ağlarında "gizli katman" olarak bilinen yapı, oto kodlayıcılarda "Bottleneck (Dar Boğaz)" terimiyle anılmaktadır (Michelucci, 2022). Düşük boyutlu veri

temsillerinin oluşturulmasında bu dar boğaz önemli bir rol oynamaktadır. Oto kodlayıcıların genel yapısı Şekil 2 ile gösterilmektedir.



Şekil 2. Oto kodlayıcılar için genel bir yapı.

Oto kodlayıcılar, Şekil 2’te gösterildiği gibi encoder ve decoder bölümlerinden oluşmaktadır. Encode aşaması, Denklem 1 ile ifade edilir:

$$a = f_{w,b}(X) \quad (1)$$

Burada, X giriş verilerini, w ağırlık matrisini, b bias terimini, a ise gizli temsili simgeler. f , doğrusal veya doğrusal olmayan bir fonksiyon olarak tanımlanabilir. Verinin boyutunu küçültmek için kullanılan gizli katman a (dar boğaz), aynı zamanda çözümleme aşamasında da kullanılmalıdır. Çözümleme için Denklem 2 kullanılır:

$$\tilde{X} = g_{w',b}(a) \quad (2)$$

Burada, g doğrusal ya da doğrusal olmayan bir eşleme fonksiyonudur. Denklem (2)’de, a kodlama aşamasından elde edilen özellik vektörünü temsil eder. \tilde{X} ise yeniden yapılandırılmış ve yeniden boyutlandırılmış yeni öznelikleri ifade eder (Yıldırım et al.,

2018). AE, yapısal olarak bir yapay sinir ağı modeline dayandığı için, doğrusal yöntemlerle kıyaslandığında daha karmaşık yapılara sahip veri setlerini işleme konusunda daha verimli ve etkili sonuçlar sağlar (Almotiri et al., 2017).

Model, giriş katmanında yüksek boyutlu bir veri kümesi (örneğin bir görüntü veya vektör) alır. Bu veri, temel özellikleri çıkarılacak şekilde işlemek üzere hazırlanır. Sonrasındaki işlem basamakları şu şekildedir (Y. Wang et al., 2016):

1. $x \in R^n$ giriş verisi olsun. Encoder ağı kullanılarak z ile temsil edilecek düşük boyutlu bir temsile dönüştürülür:

$$z = f_{\theta}(x) = \sigma(W_d z + b_d) \quad (3)$$

Bu adım sayesinde veri boyutu düşürülerek önemli bilgilerin gizli bir gösterime aktarımı sağlanır. (Encoding)

2. Kodlanmış olan z , decoder yardımı ile yeniden giriş verisinin tahmini olan \hat{x} 'e dönüştürülür. Bunun için

$$\hat{x} = g_{\phi}(z) = \sigma(W_d z + b_d) \quad (4)$$

formülü kullanılır. (Decoding)

3. Orijinal olarak alınan x verileri ve yeniden yapılandırılmış olan (\hat{x}) verileri arasındaki fark, bir kayıp fonksiyonu L ile hesaplanır ve $L(x, \hat{x}) = \|x - \hat{x}\|^2$ ile model parametreleri optimize edilir. Bu süreç, stochastic gradient descent (SGD) veya türev algoritmalarla bulunan kayıp minimize edilene kadar tekrarlanır. (Model Eğitimi)

Oto kodlayıcılar, veri boyutunu düşürme ve özellik çıkarımı gibi görevlerde yaygın olarak kullanılan güçlü bir yapay sinir ağı modelidir. Veri sıkıştırma, gürültü giderme ve generatif modellerde temel bileşen olarak kullanılan oto kodlayıcılar, gizli katmanları sayesinde verinin önemli özelliklerini öğrenir. Doğrusal olmayan ilişkileri yakalama kapasitesi, oto kodlayıcıları özellikle karmaşık veri kümelerinde etkili

kılar. Bununla birlikte, modelin başarısı, ağ mimarisi, aktivasyon fonksiyonları ve seçilen kayıp fonksiyonuna bağlıdır. Doğru tasarlandığında, oto kodlayıcılar yüksek boyutlu verilerde güçlü bir temsil öğrenme aracı olarak ön plana çıkar.

Oto kodlayıcılar, veri boyutunu düşürme ve özellik çıkarımı gibi görevlerde yaygın olarak kullanılan güçlü bir yapay sinir ağı modelidir. Veri sıkıştırma, gürültü giderme ve generatif modellerde temel bileşen olarak kullanılan oto kodlayıcılar, gizli katmanları sayesinde verinin önemli özelliklerini öğrenir (Y. Wang et al., 2016). Ayrıca, modelin başarısı, ağ mimarisi, aktivasyon fonksiyonları ve seçilen kayıp fonksiyonuna bağlıdır. Doğru tasarlandığında, oto kodlayıcılar yüksek boyutlu verilerde güçlü bir temsil öğrenme aracı olarak ön plana çıkar.

2.2. Çekirdek Temel Bileşen Analizi

Çekirdek Temel Bileşen Analizi (Kernel Principal Component Analysis - K-PCA), doğrusal yöntemlerde kullanılan PCA'nın detaylandırılması ile doğrusal olmayan yapılarda kullanılabilen bir boyut indirgeme algoritmasıdır (Schölkopf et al., 1997). Doğrusal boyut indirgeme yöntemlerinin birçoğunda kullanılan varyansın en yüksek değerine ulaşma işlemi burada kullanılabilir değildir. Veriler doğrusal olmadığı için veri seti yüksek boyutlu bir uzaya çevrilir. Böylece veri içerisinde gizlenmiş olan karmaşık yapılar çok daha net bir şekilde ortaya çıkarılabilir. Gerekli detaylar öğrenildikten sonra veri setine PCA uygulayarak veri boyutunu istenilen düzeye tekrar düşürür. K-PCA'nın algoritmik adımları şu şekilde sıralanabilir (Han et al., 2022):

1. $X = \{x_1, x_2, \dots, x_N\}$ olmak üzere N tane örnek içeren bir veri seti olsun. Her bir örnek $x_2 \in R^d$ şeklinde yani d boyutlu bir vektördür. Kernel PCA, doğrusal olmayan veri yapılarını anlamlandırmak için bir çekirdek fonksiyon kullanır. Bu fonksiyon

sayesinde verilerin yüksek boyutlu bir uzaydaki görünümünü sağlanmış olur. K-PCA uygulanırken yaygın olarak kullanılan bazı çekirdek fonksiyonlar vardır. Bunlar;

- Polinomal Kernel: $k(x, y) = (x^T y + c)^p$
- Gaussian (RBF) Kernel: $k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$
- Sigmoid Kernel: $k(x, y) = \tanh(\alpha x + y + c)$

şeklinde. Bu formüllerde kullanılan c, p, σ, α değerleri çekirdek fonksiyonun hiperparametrelerin göstermektedir.

2. $K_{ij} = k(x_i, x_j)$ ile kernel fonksiyonları kullanılarak çekirdek matrisi oluşturulur. Burada elde edilen matrisin boyutu $N \times N$ biçimindedir, bu matris veriler arasındaki benzerliklerin temsil edildiği bir matristir. Kernel PCA verilerin daha yüksek boyutlu bir görünümde ifade edilmesi çekirdek fonksiyonlarının verileri dönüştürmesi sonucu gerçekleşir.

3. K çekirdek matrisinin merkezini bulmak yani merkezlemesini yapmak için

$$K_{centered} = K - 1K - K1 + 1K1 \quad (5)$$

şeklinde birim matris olan 1 ile matris çarpımı gerçekleştirilir ve her satırın ortalaması alınarak merkezleme işlemi sonlanır.

4. PCA'nın da klasik temel adımlarından biri olan özdeğer ve özvektörlerin hesaplanması basamağı K-PCA için de geçerlidir. Burada da PCA'da olduğu gibi en büyük özdeğere sahip olan özvektörler dikkate alınır. Seçilen özdeğer ve özvektörler kullanılarak verilerin doğrusal olmayan projeksiyonları elde edilmiş olur.

5. Elde edilen özdeğerler ile en önemli görülen k tane özvektör seçimi yapılır. Seçilen bu vektörler ile veri noktalarının düşük boyutlu bir gösterimi oluşturulmuş olur. Seçilen özvektörler, belirlenen yeni

boyutlu uzaya projeksiyon olarak temel alınır ve veri boyutunun indirgenmesi işlemi yapılmış olur.

Doğrusal olmayan veri yapılarında kullanıma uygun olan K-PCA verilerin özelliklerini çekirdek fonksiyonları sayesinde kolaylıkla bulabilmektedir. Bu yöntem ile veriler yüksek boyutlu bir uzaya projekte edilir ve sonrasında boyut indirgemeyi bu uzay içerisinde gerçekleştirmektedir (Y. Wang, 2024). Bu sayede, veri analizi, kümeleme ve sınıflandırma gibi işlemlerde doğrusal olmayan yapıların etkili bir şekilde işlenmesini mümkün kılar.

2.3. Çok Boyutlu Ölçeklendirme (Multidimensional Scaling - MDS)

Çok Boyutlu Ölçeklendirme (Multidimensional Scaling - MDS) algoritmasının temel amacı, yüksek boyutlu verileri kayıp yaşanmadan uygun bir boyuta indirgemektir; bu yaklaşım, birçok boyut indirgeme algoritmasıyla benzer bir hedef taşır. MDS algoritmasının temel çalışma mekanizması veri kümesindeki nesnelere arasındaki benzerlikleri ve farklılıkları görselleştirerek boyut indirgemeyi gerçekleştiren bir yöntemdir (Açar et al., 2014). MDS, nesnelere arasındaki yakınlık ölçütlerini düşük boyutlu bir uzaydaki noktalar arasındaki mesafelere dönüştürerek, verilerin daha kolay görselleştirilmesini sağlar. Bu yöntem, kullanıcıların yapılandırılmış varsayımlarını test etmesine veya verilerdeki gizli kalıpları keşfetmesine yardımcı olur. Çok boyutlu verileri daha düşük boyutlu bir uzaya sıkıştıran MDS, değişkenler arasındaki gizli ilişkileri temsil etmek için noktalar kullanır (Beatty & Manjunath, 1997). Noktalar arasındaki mesafelerin ölçülmesi için Öklid, Manhattan, Minkovski gibi literatürde sıklıkla kullanılan uzaklık ölçütlerini kullanır. Yapılan işlemler sonrasında benzer nesnelere yakın noktalarda görselleştirilirken, birbirinden farklı olan nesnelere farklı görselleştirme ile daha belirgin hale getirilir (Cambria et al., 2013).

Çok boyutlu ölçeklendirme algoritmasının temel işlem adımları aşağıda verilmektedir (Jia et al., 2022):

1. $D = [d_{ij}] \in R^{n \times n}$ mesafe matrisine göre $A = [a_{ij}] = -\frac{1}{2} d_{ij}^2$ matrisi oluşturulur.

2. $B = [b_{ij}] = [a_{ij} - \bar{a}_{i.} - \bar{a}_{.j} + \bar{a}_{..}]$ iç çarpım matrisi hesaplanır.

3. B 'nin $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ özdeğerleri ve özvektörleri bulunur. Burada kullanılan uzaklık matrisi Öklidyen matris olduğundan, özdeğerler negatiftir. Eğer negatif bir özdeğer varsa, matris Öklidyen olmamalıdır.

S_k , temel bileşen analizindeki kümülatif katkı oranı olmak üzere

$$S_k = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n |\lambda_i|} \quad (6)$$

değeri hesaplanır. Formülde verilen k değeri çok büyük olmamalı ve ayrıca kümülatif katkı oranından yüksek bir değere sahip olmamalıdır.

4. k değeri belirlendikten sonra, $\tilde{X} = E_k \Lambda_k^{1/2} E_k^T$ şeklinde yeniden düzenleme yapılır. Burada B 'nin ilk k özvektörlerinden oluşan matris E_k ile gösterilmektedir. Λ_k ile ise k özdeğerlerinden oluşan köşegen matris ifade edilir.

Çok boyutlu ölçeklendirmenin işlem adımları incelendiğinde temel bileşen analizi ile benzerlik gösterdiği açıkça görülmektedir. İki algoritma arasındaki fark çok boyutlu ölçeklemenin örneklere dayalı olması, temel bileşen analizinin ise değişkenlere dayalı olmasıdır (Dzidolikaitė, 2015).

2.4. *İzometrik Haritalar (İsometric Mapping - İzomap)*

İzometrik Haritalar (Isometric Mapping - Isomap), veri setindeki örneklerin düşük boyutlu manifoldlara sahip olduğunu varsayar ve bu örneklerin düşük boyutlu uzaydaki temsillerini izometrik haritalama yöntemiyle oluşturur. Çoğu boyut indirgeme algoritmasında örnekler arasındaki mesafeyi ölçmek için kullanılan doğrusal yapıları yansıtan Öklid mesafesi yerine izometrik haritalamada jeodezik mesafe kullanılır (Anowar et al., 2021). Bu algoritma, yüksek boyutlu verilerin içsel yapısını doğru bir şekilde yakalayarak, MDS yöntemini kullanarak bu jeodezik mesafeleri daha düşük boyutlu bir uzaya yansıtır. Bu sayede, veri noktalarının yerini değiştirmeden, verilerin içsel yapısını koruyarak daha doğru bir şekilde temsil edilmesi sağlanır. İzomap algoritması 3 adımdan oluşmaktadır (Ding et al., 2022):

1. Tüm veri noktalarının komşuluk grafiği G oluşturulur. Bu adımda, iki nokta arasındaki mesafe Öklid mesafesi olarak tanımlanır ve komşuluk ilişkisi ya k -en yakın komşu ya da ϵ -komşuluk olarak belirlenir.

2. Jeodezik mesafe matrisi DM hesaplanır. Bu matris, komşuluk grafiği G üzerindeki iki nokta arasındaki en kısa mesafeyi hesaplayarak, manifold üzerindeki iki nokta arasındaki jeodezik mesafeyi yeniden oluşturur.

3. Düşük boyutlu yerleştirme G oluşturulur. Jeodezik mesafe matrisi DM kullanılarak, MDS yöntemi ile veri noktaları düşük boyutlu bir uzaya indirgenir.

İzomap boyut indirgeme algoritmasını diğer boyut indirgeme algoritmalarından ayıran en belirgin özelliği doğrusal olmayan veri yapılarında kolaylıkla kullanılmasıdır. İzomap verinin içsel manifold yapısını koruyarak doğrusal olmayan ilişkileri de keşfetmeye olanak tanır. Bu özellik, özellikle verilerin karmaşık yapılar gösterdiği

durumlarda izomap etkili bir seçenek haline gelmektedir. Yani izomap yüksek boyutlu verilerin daha doğru ve anlamlı bir şekilde düşük boyutlu uzaya indirgenmesine olanak sağlar.

2.5. *Yerel Lineer Gömme*

Yerel Lineer Gömme (Locally Linear Embedding – LLE), verilerin yerel yapısını koruyarak yüksek boyutlu veri gruplarını düşük boyutlarda temsil etmeyi amaçlamaktadır (Roweis & Saul, 2000). LLE, yüksek boyutlu verilerin içyapılarını keşfetmek ve verinin manifold üzerindeki dağılımını göstermek için kullanılır. Her bir veri noktasına, kendi yakınında yani komşusu olan veri noktalarına doğrusal bir şekilde bağlanabilir. LLE'nin çalışma prensibinin temelinde bu mantık bulunmaktadır. Veri noktaları arasında sadece komşuların belirlediği doğrusal ilişkiler önemsenir (J. Wang, 2012). Bu özellik verilerin doğrusal olmayan bir manifold üzerinde bulunması durumunda LLE'yi etkili bir boyut indirgeme algoritması konumuna getirir. LLE algoritması üç temel adımdan oluşmaktadır:

1. Her veri noktası için, en yakın k komşusu belirlenir. Bu komşuluklar genellikle Öklid mesafesi veya başka bir mesafe ölçütü kullanılarak hesaplanır.

2. Her veri noktası, komşularının doğrusal kombinasyonu ile temsil edilir. Bu adımda, her veri noktasının komşularına olan ağırlıkları hesaplanır. Bu ağırlıklar, her noktanın komşularına nasıl doğrusal bir şekilde bağlandığını gösterir.

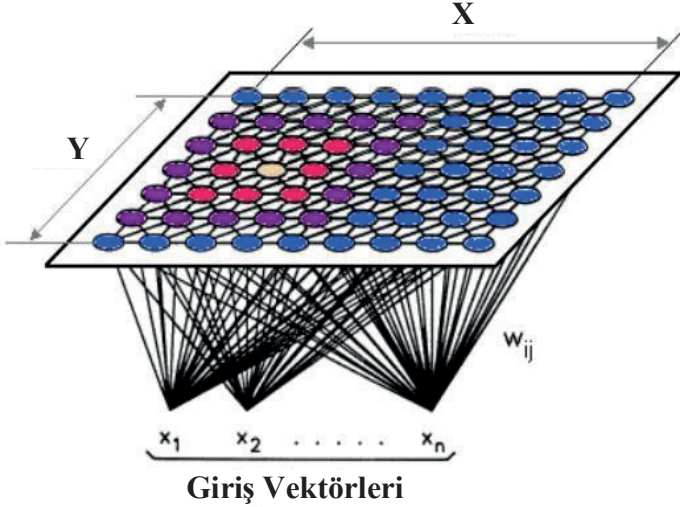
3. Son olarak, yüksek boyutlu uzayda elde edilen yerel ağırlık ilişkilerini koruyarak, veri noktaları düşük boyutlu bir uzaya projeksiyon yapılır. Bu adım, MDS gibi bir yöntemle yapılır ve verinin düşük boyutlu bir temsilini sağlar.

Verilerin büyük bir kısmı doğrusal olmayan bir yapıya sahipse boyut indirgeme yöntemi olarak LLE tercih edilebilir. LLE algoritması

verinin iç yapısını iyi bir şekilde koruduğu için doğrusal olmayan veri setlerinde veri kaybının yaşanmasını en aza indirir.

2.6. *Kendi Kendini Düzenleyen Haritalar*

Kohonen Map olarak da isimlendirilen Teuvo Kohonen tarafından ortaya koyulmuş bir çeşit sinir ağı yapısı olan Kendi Kendini Düzenleyen Haritalar (Self-Organizing Maps – SOM), boyut indirgeme, veri görselleştirme ve kümeleme için kullanılan bir yöntemdir (Kohonen, 1982). Yüksek boyutlu verileri daha düşük boyutlu bir haritaya projekte ederek birbirine yakın noktaları bir araya toplamayı hedefler. Diğer algoritmalarda olduğu gibi ana amaçlarından biri verilerin içsel yapılarını anlamlandırılmasını kolaylıkla sağlamaktır. Birbirine benzeyen veri noktaları bir arada olacak şekilde yerleştirilerek veri kümelerinin birbirleri ile ilişkileri görsel olarak incelenebilir hale getirilir. SOM yapısının mimari yapısına ait görsel Şekil 3’de verilmiştir (Eklavya, 2019).



Şekil 3. Kendi kendini düzenleyen haritalara ait mimari yapının gösterimi

SOM algoritmasının temel adımları şu şekildedir (Eklavya, 2019):

1. Hazırlanacak haritanın şekil ve boyut belirlenir. Girdi ve çıktı arasında gerekli olan başlangıç ağırlıkları atanır. (Başlatma)
2. Belirlenen haritanın boyutuna uygun olacak şekilde girdiden rastgele örnekler alınır. (Örnekleme)
3. Girdiye en yakın olan düğümün haritadaki yeri hesaplanır. (En Yakın Düğüm Hesaplama)
4. En yakın düğüme komşu olan düğümlerin ağırlıkları düzenlenir. (Ağırlık Güncelleme)
5. Başlangıçta belirlenen iterasyon eşiğine ulaşana kadar veya en küçük hata değeri elde edilene kadar Adım 2'den Adım 4'e kadar olan adımlar tekrarlanır. (Tekrar)

SOM, yüksek boyutlu verilerin boyutunun indirgenmesi, görselleştirilmesi ve kümelenmesi için etkili bir araçtır. Gözetimsiz öğrenme yaklaşımı ve esnek yapısı sayesinde çeşitli veri türlerinde başarılı bir şekilde uygulanabilir. Boyut indirgeme ve kümeleme algoritmalarının bir kombinasyonu olan SOM, veri analizinde güçlü bir yöntem sunar (Saraswati et al., 2018).

2.7. t-Dağıtık Stokastik Komşu Yerleştirme

t-Dağıtık Stokastik Komşu Yerleştirme (t-Distributed Stochastic Neighbor Embedding - t-SNE), özellikle karmaşık yapılar içeren verilerde, verilerin bulunduğu uzayın boyutunu indirgeyerek verilerin anlamlandırılmasını kolaylaştırmaktadır (Balamurali, 2023). t-SNE yönteminin temel amacı, yüksek ve düşük boyutlu uzaylardaki veriler arasındaki benzerlikleri, olasılık dağılımlarını kullanarak tanımlamaktır. Ayrıca olasılık işlemlerinden elde ettiği dağılımları optimize etmek de amaçlarından biridir (Anowar et al., 2021). t-SNE algoritmasının uygulama adımları aşağıdaki gibi sıralanmaktadır (Kang et al., 2021):

1. Yüksek boyutlu uzayda bulunan veri noktalarının birbirleri ile olan benzerliklerinin hesaplanması uygulanacak olan ilk adımdır. Bu işlemi gerçekleştirmek için Gaussian dağılımından faydalanılır. i ve j veri noktaları olmak üzere bu iki nokta arasındaki olasılık;

$$p_{i|j} = \frac{\exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma_i^2}\right)}{\sum_{k \neq i} \exp\left(-\frac{\|x_i - x_k\|^2}{2\sigma_i^2}\right)} \quad (7)$$

şeklinde hesaplanmaktadır. Burada σ_i t-SNE algoritmasında kullanılan bir hiperparametre olan etkili komşuluk boyutunu (perplexity parametresi) belirtir. Yani, seçilen veri noktası çevresindeki kaç verinin dikkate alınacağına bu parametre ile karar verilir. Burada σ_i , i . veri noktası için perplexity parametresine bağlı bir değerdir. Tüm noktalar arasında bulunan simetrik benzerlik ise p_{ij} ile ifade edilmektedir ve tanımı $p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n}$ ile verilmektedir.

2. Yüksek boyutlu uzaydaki benzerliklerden sonra bir de düşük boyutlu uzaydaki verilerin benzerliğinin hesaplanması gerekmektedir. Bu benzerlik Student-t dağılımı kullanılarak Denklem 8 formülü ile hesaplanır.

$$q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k \neq i} (1 + \|y_k - y_l\|^2)^{-1}} \quad (8)$$

Student-t dağılımı olasılık yoğunluk fonksiyonunun uç noktalarında bulunan olasılık değerlerini (uzun kuyrukları) hesaplama konusunda kullanıldığı için t-SNE algoritmasında da uzun kuyrukları, uzak noktaların etkisini azaltmaktadır.

3. Yüksek boyutlu p_{ij} ve düşük boyutlu q_{ij} benzerlikler arasındaki fark, Kullback-Leibler divergence (KL-divergence) ile hesaplanmaktadır. KL-divergence formülü;

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (9)$$

ile verilmektedir. Bu işlemin yapılmasındaki amaç $KL(P||Q)$ değerini minimize etmektir.

4. Gradient descent yöntemi ile düşük boyutlu koordinatlar y_i olarak güncellenir. Bu güncelleme için aşağıda verilen formül kullanılır:

$$y_i = y_i - \eta \frac{\partial KL(P||Q)}{\partial y_i} \quad (10)$$

Burada η ile verilen değer algoritmanın öğrenme hızını göstermektedir.

5. Adım 2 – Adım 4 arası, belirli bir iterasyon sayısına göre belirli bir eşik değerine ulaşıncaya kadar tekrar edilir. Burada dikkate alınan diğer bir parametre de hata değerinin düşürülmesidir. Algoritma optimum değerlere ulaşıp sonlandığında y_i ile elde edilen değer t-SNE algoritmasının çıktısı olarak değerlendirilir.

t-SNE algoritması, olasılık tabanlı bir boyut indirgeme algoritmasıdır. Yüksek boyutlu verilerin birbirleri ile olan benzerliklerinin olasılıksal değerlerini çıkararak düşük boyutlu uzaya optimize eder. Böylece verilerin düşük boyutlu uzayda görselleştirilmesini sağlar. Karmaşık yapıya sahip olan veriler için anlamlı sonuçlar çıkarılması bu algoritma ile mümkün hale gelmektedir.

2.8. Vektör Nicelemesini Öğrenme (Learning Vector Quantization - LVQ)

LVQ algoritması doğrudan bir boyut indirgeme algoritması olmamasına rağmen yüksek boyutlu veri uzaylarını prototip

vektörlerle göstererek daha düşük bir boyuta indirilmesini sağlamaktadır (Hammer et al., 2014). Bu algoritma uyguladığı prototip tekniği ile her sınıf için önemli noktaları koruyarak düşük bilgi kaybı olması durumunu üst düzeye çıkarabilir. Karmaşık ve doğrusal olmayan veri yapıları için kullanışlı bir boyut indirgeme süreci sunar (Hammer & Villmann, 2002). SOM algoritmasına benzer bir yapıya sahip olan sinir ağı temelli bir algoritmadır. Sınıf bölgelerini prototipler yardımıyla dolaşarak istatistiksel bir yapı altında desenleri çıkaran etkili bir yöntem olarak literatürde kendine yer edinmiştir (Nanga et al., 2021). LVQ algoritmasının temel işlem adımları aşağıdaki gibi sırasıyla verilebilir (Kohonen, 2001):

1. İşlenmek istenen veri seti $X = \{x_1, x_2, \dots, x_n\}$, bu veri setine ait etiket sınıfları $y = \{y_1, y_2, \dots, y_n\}$ ile gösterilsin. Veri setinde var olan her bir sınıf için bir $P = \{p_1, p_2, \dots, p_k\}$, (k : sınıf sayısı) prototip vektörü rastgele olacak şekilde seçilsin. Seçilen bu vektörler, veri setinin düşük boyutlu gösterimleri olmak üzere başlangıç noktası olarak kullanılacaktır.

2. x_i olarak belirlenen her bir veri noktasının tüm prototip vektörlerine olan uzaklığını bulmak için Öklik uzaklık ölçütü kullanılır:

$$d(x_i, p_j) = \|x_i - p_j\| = \sqrt{\sum_{m=1}^M (x_{im} - p_{jm})^2} \quad (11)$$

p_j en yakın prototip değeri olmak üzere, $d(x_i, p_j)$ 'nin minimum olduğu prototip değeri

$$p_{winner} = \arg \min_j d(x_i, p_j) \quad (12)$$

formülü ile hesaplanır.

3. Oluşan prototip vektörler, x_i veri noktaları ile ilişkili olan y_i sınıf etiketlerine göre güncellenir, α öğrenme oranı olmak üzere:

- Eğer $y_i = y(p_{winner})$ (Doğru):

$$p_{winner} \leftarrow p_{winner} + \alpha(x_i - p_{winner})$$

- Eğer $y_i \neq y(p_{winner})$ (Yanlış):

$$p_{winner} \leftarrow p_{winner} - \alpha(x_i - p_{winner})$$

4. Bir önceki adımdan elde edilen sonuçlara göre eğitim tamamlandığında, her bir prototip vektör sınıflarını temsil edecek şekilde optimum değerine ulaşmıştır. Bu aşamadan sonra başlangıçta X ile gösterilen veri noktaları, prototiplere olan uzaklıklarına göre ($d(x_i, p_j)$) boyutu indirgenerek düşük boyutlu bir uzayda ifade edilebilir hale getirilmiştir. Bu ifadenin gösterimi ise

$$z_i = \{d(x_i, p_1), d(x_i, p_2), \dots, d(x_i, p_k)\} \quad (13)$$

ile verilmektedir. Bu formülden yola çıkarak yüksek boyutlu x_i verisinin düşük boyutlu bir temsili z_i olarak elde edilmiştir, denilebilir.

3. Sonuçlar

Bu çalışmada doğrusal olmayan boyut indirgeme algoritmaları detaylı bir şekilde ele alınmış ve farklı yöntemlerin avantajları ile kullanım alanları karşılaştırmalı olarak incelenmiştir. İncelenen algoritmalar, özellikle karmaşık ve doğrusal olmayan veri yapılarına sahip veri setlerinin işlenmesinde önemli avantajlar sunmaktadır.

Oto kodlayıcılar (AE), yüksek boyutlu verilerin sıkıştırılması ve yeniden yapılandırılması konusunda etkili bir yaklaşım sunarken, çekirdek PCA (K-PCA) doğrusal olmayan veri yapılarında özdeğer ve özvektör tabanlı bir dönüşümle başarılı sonuçlar sağlamaktadır. Çok Boyutlu Ölçeklendirme (MDS) ve İzometrik Haritalar (Isomap) algoritmaları, veri noktaları arasındaki mesafeleri koruyarak düşük boyutlu temsiller oluştururken, Yerel Lineer Gömme (LLE), verilerin yerel komşuluk ilişkilerini dikkate alarak içsel yapıyı korumayı hedeflemektedir. Özellikle t-SNE yöntemi, karmaşık veri kümelerinin düşük boyutlu görselleştirilmesinde kullanıcılara büyük kolaylık sağlamaktadır.

Algoritmaların kullanımında, veri setinin yapısı ve hedeflenen çıktılar dikkate alınmalıdır. Örneğin, otoenkoderler ve t-SNE, veri sıkıştırma ve görselleştirme odaklı projelerde öne çıkarken, K-PCA ve LLE, sınıflandırma ve kümeleme gibi görevlerde daha uygun olabilir. İzometrik Haritalar ve Çok Boyutlu Ölçeklendirme ise jeodezik mesafelerin önemli olduğu durumlarda tercih edilmelidir.

Bu algoritmaların her biri, veri analizi süreçlerinde çeşitli avantajlar sunmaktadır ve doğru yöntem seçimi, sonuçların doğruluğu ile işlem verimliliğini doğrudan etkilemektedir. Bu nedenle, algoritmaların temel işleyiş mekanizmaları ve uygulama alanlarının iyi anlaşılması, veriye dayalı karar verme süreçlerinde kritik bir öneme sahiptir.

KAYNAKLAR

- Açar, N., Akyol, A., & Esen, S. K. (2014). *Marketing Higher Education In Turkey: A Multidimensional Scaling (Mds) Analysis Approach*.
- Almotiri, J., Elleithy, K., & Elleithy, A. (2017). Comparison Of Autoencoder And Principal Component Analysis Followed By Neural Network For E-Learning Using Handwritten Recognition. *2017 Ieee Long Island Systems, Applications And Technology Conference (Lisat)*, 1–5. <https://doi.org/10.1109/Lisat.2017.8001963>
- Anowar, F., Sadaoui, S., & Selim, B. (2021). Conceptual And Empirical Comparison Of Dimensionality Reduction Algorithms (Pca, Kpca, Lda, Mds, Svd, Lle, Isomap, Le, Ica, T-Sne). *Computer Science Review*, 40, 100378. <https://doi.org/10.1016/J.Cosrev.2021.100378>
- Balamurali, M. (2023). T-Distributed Stochastic Neighbor Embedding. In B. S. Daya Sagar, Q. Cheng, J. Mckinley, & F. Agterberg (Eds.), *Encyclopedia Of Mathematical Geosciences* (Pp. 1527–1535). Springer International Publishing. https://doi.org/10.1007/978-3-030-85040-1_446
- Bank, D., Koenigstein, N., & Giryas, R. (2021). *Autoencoders* (No. Arxiv:2003.05991). Arxiv. <http://arxiv.org/abs/2003.05991>
- Beatty, M., & Manjunath, B. S. (1997). Dimensionality Reduction Using Multi-Dimensional Scaling For Content-Based Retrieval. *Proceedings Of International Conference On Image Processing*, 2, 835–838 Vol.2. <https://doi.org/10.1109/Icip.1997.638626>
- Cambria, E., Mazzocco, T., & Hussain, A. (2013). Application Of Multi-Dimensional Scaling And Artificial Neural Networks For Biologically Inspired Opinion Mining. *Biologically Inspired Cognitive Architectures*, 4, 41–53. <https://doi.org/10.1016/J.Bica.2013.02.003>
- Ding, S., Keal, C. A., Zhao, L., & Yu, D. (2022). Dimensionality Reduction And Classification For Hyperspectral Image Based On Robust Supervised Isomap. *Journal Of Industrial And Production Engineering*, 39(1), 19–29. <https://doi.org/10.1080/21681015.2021.1952657>

- Dzidolikaitė, A. (2015). Genetic Algorithms For Multidimensional Scaling. *Mokslas – Lietuvos Ateitis / Science – Future Of Lithuania*, 7(3), Article 3. <https://doi.org/10.3846/Mla.2015.781>
- Eklavya. (2019, October 31). *Kohonen Self-Organizing Maps*. Medium. <https://towardsdatascience.com/kohonen-self-organizing-maps-a29040d688da>
- Hammer, B., Hofmann, D., Schleif, F.-M., & Zhu, X. (2014). Learning Vector Quantization For (Dis-)Similarities. *Neurocomputing*, 131, 43–51. <https://doi.org/10.1016/j.neucom.2013.05.054>
- Hammer, B., & Villmann, T. (2002). Generalized Relevance Learning Vector Quantization. *Neural Networks*, 15(8), 1059–1068. [https://doi.org/10.1016/S0893-6080\(02\)00079-5](https://doi.org/10.1016/S0893-6080(02)00079-5)
- Han, Y., Song, G., Liu, F., Geng, Z., Ma, B., & Xu, W. (2022). Fault Monitoring Using Novel Adaptive Kernel Principal Component Analysis Integrating Grey Relational Analysis. *Process Safety And Environmental Protection*, 157, 397–410. <https://doi.org/10.1016/j.psep.2021.11.029>
- Jia, W., Sun, M., Lian, J., & Hou, S. (2022). Feature Dimensionality Reduction: A Review. *Complex & Intelligent Systems*, 8(3), 2663–2693. <https://doi.org/10.1007/S40747-021-00637-X>
- Kang, B., García García, D., Lijffijt, J., Santos-Rodríguez, R., & De Bie, T. (2021). Conditional T-Sne: More Informative T-Sne Embeddings. *Machine Learning*, 110(10), 2905–2940. <https://doi.org/10.1007/S10994-020-05917-0>
- Kohonen, T. (1982). Self-Organized Formation Of Topologically Correct Feature Maps. *Biological Cybernetics*, 43(1), 59–69. <https://doi.org/10.1007/Bf00337288>
- Kohonen, T. (2001). Learning Vector Quantization. In T. Kohonen (Ed.), *Self-Organizing Maps* (Pp. 245–261). Springer. https://doi.org/10.1007/978-3-642-56927-2_6
- Lee, J. A., & Verleysen, M. (Eds.). (2007). *Nonlinear Dimensionality Reduction*. Springer. <https://doi.org/10.1007/978-0-387-39351-3>
- Mendez, M. A. (2023). Linear And Nonlinear Dimensionality Reduction From Fluid Mechanics To Machine Learning. *Measurement Science*

- And Technology*, 34(4), 042001. <https://doi.org/10.1088/1361-6501/Acaff6>
- Michelucci, U. (2022). *An Introduction To Autoencoders* (No. Arxiv:2201.03898). Arxiv. <http://arxiv.org/abs/2201.03898>
- Nanga, S., Bawah, A. T., Acquaye, B. A., Billa, M.-I., Baeta, F. D., Odai, N. A., Obeng, S. K., & Nsiah, A. D. (2021). Review Of Dimension Reduction Methods. *Journal Of Data Analysis And Information Processing*, 9(3), Article 3. <https://doi.org/10.4236/jdaip.2021.93013>
- Reddy, G. T., Reddy, M. P. K., Lakshmana, K., Kaluri, R., Rajput, D. S., Srivastava, G., & Baker, T. (2020). Analysis Of Dimensionality Reduction Techniques On Big Data. *Ieee Access*, 8, 54776–54788. Ieee Access. <https://doi.org/10.1109/Access.2020.2980942>
- Roweis, S. T., & Saul, L. K. (2000). Nonlinear Dimensionality Reduction By Locally Linear Embedding. *Science*, 290(5500), 2323–2326. <https://doi.org/10.1126/Science.290.5500.2323>
- Saraswati, A., Nguyen, V. T., Hagenbuchner, M., & Tsoi, A. C. (2018). High-Resolution Self-Organizing Maps For Advanced Visualization And Dimension Reduction. *Neural Networks*, 105, 166–184. <https://doi.org/10.1016/j.neunet.2018.04.011>
- Schölkopf, B., Smola, A., & Müller, K.-R. (1997). Kernel Principal Component Analysis. In W. Gerstner, A. Germond, M. Hasler, & J.-D. Nicoud (Eds.), *Artificial Neural Networks—Icann'97* (Pp. 583–588). Springer. <https://doi.org/10.1007/Bfb0020217>
- Wang, J. (2012). Locally Linear Embedding. In J. Wang (Ed.), *Geometric Structure Of High-Dimensional Data And Dimensionality Reduction* (Pp. 203–220). Springer. https://doi.org/10.1007/978-3-642-27497-8_10
- Wang, Y. (2024). Near-Optimal Quantum Kernel Principal Component Analysis. *Quantum Science And Technology*, 10(1), 015034. <https://doi.org/10.1088/2058-9565/Ad9176>
- Wang, Y., Yao, H., & Zhao, S. (2016). Auto-Encoder Based Dimensionality Reduction. *Neurocomputing*, 184, 232–242. <https://doi.org/10.1016/j.neucom.2015.08.104>

Yildirim, O., Tan, R. S., & Acharya, U. R. (2018). An Efficient Compression Of Ecg Signals Using Deep Convolutional Autoencoders. *Cognitive Systems Research*, 52, 198–211.
<https://doi.org/10.1016/j.cogsys.2018.07.004>