

Editör

Doç. Dr. Emin BORANDAĞ

gece
kitaplığı

**Bilgisayar
Bilimleri ve
Mühendisliği - Yapay
Zeka Alanında
Uluslararası
Araştırmalar ve
Değerlendirmeler**

**Ekim
2025**

Genel Yayın Yönetmeni / Editor in Chief • Eda Altunel

Kapak & İç Tasarım / Cover & Interior Design • Gece Kitaplığı

Birinci Basım / First Edition • © Ekim 2025

ISBN • 978-625-388-768-1

© copyright

Bu kitabın yayın hakkı Gece Kitaplığı'na aittir.

Kaynak gösterilmeden alıntı yapılamaz, izin almadan hiçbir yolla çoğaltılamaz. The right to publish this book belongs to Gece Kitaplığı. Citation can not be shown without the source, reproduced in any way without permission.

Gece Kitaplığı / Gece Kitaplığı

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

Ümit Apt No: 22/A Çankaya/ANKARA

Telefon / Phone: 0312 384 80 40

web: www.gecekitapligi.com

e-mail: gecekitapligi@gmail.com

Baskı & Cilt / Printing & Volume

Sertifika / Certificate No: 42488

BİLGİSAYAR BİLİMLERİ
VE MÜHENDİSLİĞİ - YAPAY
ZEKA ALANINDA
ULUSLARARASI
ARAŞTIRMALAR VE
DEĞERLENDİRMELER

EDİTÖR

DOÇ. DR. EMİN BORANDAĞ

İÇİNDEKİLER

Bölüm 1

Gıda Tedarik Sürecinde Yapay Zeka

Serel ÖZMEN AKYOL, Aysel GÜLBANDILAR —1

Bölüm 2

Gıda Güvenliğinde Yapay Zeka

Aysel GÜLBANDILAR, Serel ÖZMEN AKYOL —13

Bölüm 3

Gerçek Zamanlı Hava Kalitesi İzleme ve Analiz Sistemi

Engin OĞUZAY, Gökhan UÇKAN —33

Bölüm 4

Z-Skor Normalizasyonunun İris Veri Seti Sınıflandırma Performansına Etkisi

Bahadır ÇOKÇETİN, Ahmet ÇELİK—53

Bölüm 5

Veri İşlemede Çarpıklık Dağılımını Tespit Etme: Görsel Grafik Aracı Üzerinde Bir Uygulama

Ahmet ÇELİK, Bahadır ÇOKÇETİN—69

//

Bölüm 1

GIDA TEDARİK SÜRECİNDE YAPAY ZEKA

Serel ÖZMEN AKYOL¹, Aysel GÜLBANDILAR²

1 Dr. Öğr. Üyesi, Eskişehir Osmangazi Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü. sozmen@ogu.edu.tr ORCID No: 0000-0002-5344-4065

2 Doç. Dr., Eskişehir Osmangazi Üniversitesi, Ziraat Fakültesi Gıda Mühendisliği Bölümü. aysel.gulbandilar@ogu.edu.tr ORCID No: 0000-0001-9075-9923

GİRİŞ

Son yıllarda dünya nüfusunun artmasıyla birlikte gıda talebinde de bir artış yaşanmıştır. Bu talebin karşılanması, üretim geliştirme, gıda tedarik zincirinin yönetimi, kalitesinin iyileştirilmesi, belirlenmesi ve sınıflandırılması gibi amaçlarla sunduğu çözümler nedeniyle yapay zekâ gıda endüstrisinde teknolojik bir yaklaşım olarak yerini almıştır (Mavani et al., 2022; Sahni, Srivastava, & Khan, 2021) Yapay zeka, gerçek dünya problemlerinin matematiksel olarak modellenmesine imkan sağlayarak, gıda mühendisliği de dahil olmak üzere bir çok disiplinde etkili çözümler sunmaktadır (Arjona-Román et al., 2017)

Gıda mühendisliğinde yapay zeka, gıda üretimi, tedarik zinciri, güvenliği, kalitesi, teslimatı vb. hakkında düşünme biçimini değiştirmekte ve akıllı mobil uygulamalar çağının geçişine katkı sağlamaktadır. Yapay beyinler sayesinde, gıda veri tabanları oluşturulabilmekte ve verimli bir şekilde analiz edilebilmektedir. YZ tabanlı sistemler, RFID ve biyosensörler gibi teknolojilerle birleşerek tedarik zinciri boyunca gerçek zamanlı veri toplanmasına olanak tanımakta, böylece sürekli kalite kontrolü ve risk yönetimi sağlanmakta, gıda kaynaklı tehlikelerin erken tahmini ve yönetiminde proaktif çözümler sunmaktadır.

Günümüzde birçok çalışmada, araştırmaların hızlandırılması, analiz sürelerinin kısaltılması ve geliştirilen ürünlerin üretimi konularında modelleme programları yaygın olarak kullanılmaktadır. Bu teknolojiler kullanılarak; araştırma ve uygulamalarda elde edilen bilgilerin toplanması, depolanması, sınıflandırılması, yönetimi ve kullanımının etkinleştirilmesi sağlanmaktadır. Bu doğrultuda yapay zeka, özellikle geleneksel programlama için çok karmaşık olan veya önemli miktarda el emeği ve maliyet gerektirecek karmaşık kararlar almak için faydalı bir teknolojidir.

2000’li yıllarda gıda alanında basit makine öğrenmesi algoritmalarının kullanımı ile başlayan yapay zeka uygulamaları, görüntü analizi ve biyosensörler ile sistemin geliştirilmesi, gıda ürünlerinin geri çağırma takibi, yapay zeka ve nesnelerin interneti entegrasyonu sayesinde tedarik zincirinde gerçek zamanlı izlenebilirliği, doğal dil modelleri ile gıda tedarik sürecinde otomatik raporlama, sosyal medya ve geri bildirimlerin değerlendirilmesi gibi bir çok alana yayılmıştır. Devam eden çalışmalar ile yapay zekanın güncel teknoloji entegrasyonlarıyla akıllı karar sistemleri de dahil olmak üzere tedarik

sürecinde uçtan uca gıda güvencesi sağlayacak otonom yapay zeka modellerinin kullanımı söz konusu olacaktır.

Gıda Tedarik Sürecinde Yapay Zeka Uygulamaları

Gıda, birbiriyle küresel ölçekte etkileşime giren karmaşık bir tedarik zinciri ağı içinde üretilmektedir. Gıda maddeleri çiftlikten fabrikaya ve fabrikadan sofraya tüm tedarik süresince çeşitli sağlık tehlikeleriyle karşı karşıya kalabilmektedir. Dolayısıyla gıda sisteminin performansını, aralarında etkileşim bulunan uluslararası ve yerel gelişmeler etkilemektedir.

Gıda tedarik zincirini etkileyen birçok etken bulunmaktadır. Bunlar;

- artan dünya nüfusu ve değişen beslenme biçimleri,
- tarımsal üretim için ihtiyaç duyulan kaynaklardaki kıtlıklar (örn. verimli toprak, tatlı su, enerji),
- iklim değişikliği,
- azalan biyolojik çeşitlilik,
- yetersiz denetim ve
- rakip tarım sistemleri (Marvin et al., 2022)'dir.

Son yıllarda küreselleşen gıda ticaretinde yaşanan üretim ve tedarik zinciri hareketlerindeki artış, gıda güvenliğinde mikrobiyolojik salgınlara da katkıda bulunmaktadır. Mevcut durum, gıda şirketlerinin beklenmedik gıda güvenliği sorunlarından kaçınmak için yerel ve uluslararası pazarlarda üretilen gıdaların mikrobiyolojik güvenliğini artırma ihtiyacını da arttırmaktadır (Hussain & Dawson, 2013).

Gıda güvenliği tehlikeleri, gıdaları tüketicinin sağlığına zararlı hale getirebilecek süreçteki tüm tehlikeleri ifade etmektedir. Gıda tedarik zinciri boyunca gıda ürünleri, biyolojik tehlikeler (bakteri, virüsler ve parazitler), kimyasal tehlikeler (ağır metaller, pestisitler ve mikotoksinler) veya fiziksel tehlikeler (metal parçalar ve cam parçaları) olmak üzere çeşitli güvenlik tehlikeleriyle kontamine olabilmektedir (Bouzembrak, Camenzuli, Janssen, & van der Fels-Klerx, 2018; Lawley, Curtis, & Davis, 2012).

Piyasada satılan kontamine gıdalar, tüketildiğinde insanlarda ciddi sağlık hasarına neden olma potansiyeline sahiptir. Bunun yanı sıra piyasaya kontamine gıda tedarik eden şirketler, yüksek geri çağırma ve elden çıkarma maliyetleri, itibar zedelenmesi ve tüketicilerin şirkete

olan güveninin kaybolması riskini taşımakta ve bu durum düşük satış değerlerine sebep olabilmektedir (Hussain & Dawson, 2013). Tüketicilerin bilgilendirilmesi, yiyeceklerin raflardan kaldırılması ve davalar sonucunda ödenen tazminatlardan kaynaklanan gıda güvenliği olaylarının Amerika Birleşik Devletleri ekonomisi için yıllık tahmini maliyetinin, 7 milyar dolar civarında olduğu belirtilmektedir (Thomas et al., 2015). Diğer ülkelerin çoğu da benzer şekilde ekonomik kayıplar yaşamaktadır.

Gıda güvenliği tehlikeleri için risk bazında izleme programları artan bir şekilde oluşturulmaktadır. Bu tehlikelerin varlığı, insan ve hayvan sağlığı üzerinde potansiyel olarak zararlı etkileri olan gıda veya yem ürünlerinin güvenliğini etkileyebilmektedir. Gıda güvenliği yönetim sistemlerinin doğru işleyişini ve dolayısıyla gıda ve yem güvenliğini garanti altına almak için tüm gıda tedarik zinciri boyunca potansiyel gıda güvenliği tehlikelerini izlemek önemlidir (Focker, van der Fels-Klerx, & Oude Lansink, 2018; van Asselt, Noordam, Pikkemaat, & Dorgelo, 2018)plants (pests. Birçok gıda güvenliği riskinin varlığı ve gelişimi, iklim, ekonomi ve insan davranışı gibi gıda tedarik zincirinin içindeki ve dışındaki faktörlerin bir araya gelmesiyle oluşmaktadır. Bu faktörler ve tedarik zinciri arasındaki karmaşık etkileşimlerin neden-sonuç ilişkilerini ortaya çıkarmak ve gıda güvenliği risklerini en aza indirmek amacıyla etkili hafifletme eylemleri gerçekleştirebilmek için bütünsel, sistematik bir yaklaşım ihtiyaç duyulmaktadır (Bouzembrak & Marvin, 2019). Bu ihtiyaçlara yönelik olarak son dönemde, gıda tedarik zinciri sürecinde çok faktörlü ve aralarında karmaşık etkileşimlerin bulunduğu modellerde yapay zeka, makine öğrenmesi, kurumsal kaynak planlaması (ERP), radyo frekanslı tanımlama (RFID), blokzincir, büyük veri, nesnelerin interneti gibi güncel yöntemlerle daha etkin sonuçlar elde edilebilmektedir (Keleş & Ova, 2020).

Zhou ve ark. gıda tanıma, kalori tahmini, meyve, sebze, et ve su ürünlerinin kalite tespiti, gıda tedarik zinciri ve gıda kontaminasyonunu içeren gıda alanında derin öğrenmenin kullanımını tartışmışlardır [58]. Bouzembrak vd. (2018) ise yaptıkları çalışmada, tedarik zincirinin her seviyesinde (tedarikçiler, sınır kontrol noktaları, pazar ve tüketiciler) hangi tür bitki ve baharat ürünlerinin ve gıda güvenliği tehlikelerinin tercihli olarak izlenmesi gerektiğini tahmin etmek için bir yapay zeka destekli bir model geliştirmişlerdir (Bouzembrak et al., 2018). Çalışmada, 2005-2014 yılları arasında Hollanda ulusal izleme programı veri tabanından alınan gıda ve yemlerdeki kimyasal kirleticilere ilişkin bildirimler Bayes Ağı ile analiz edilerek bir hızlı uyarı sistemi geliştirilmiştir.

Geleneksel yöntemlerle karşılaştırıldığında gıda ürünlerinin duyu-sal ve kalite özelliklerini tahribatsız, hızlı ve verimli bir şekilde değer-

lendirmeleri sebebiyle makine öğrenmesi yöntemleri gıda tedarik zinciri sürecinde etkili bir şekilde kullanılmaktadır. Alanda makine öğrenmesi yöntemlerinin potansiyelini göstermek adına, 2017 – 2022 yılları arasında literatürde yer alan gıda tedarik zincirinde sınıflandırma, paketleme, nakliye, depolama ve satış ve hiperspektral görüntüleme süreci için mevcut makine öğrenimi uygulamalarının incelendiği kapsamlı bir çalışma sunulmuştur (Kang et al., 2022). Benzer bir çalışmada makine öğrenimi, yapay zeka ve gıda tedarik zinciri uygulamaları değerlendirilerek, geliştirilen yeni modellerin işletmelere rekabet avantajı sağlama, mali kayıpları azaltma ve verimliliğini destekleme yeteneklerine sahip olduğu ifade edilmiştir (Diwane, Oza, Nerlekar, & Chougule, 2023).

Gıda tedarik zincirinde yapay zeka, şeffaflığı ve izlenebilirliği iyileştirerek gıda güvenliği, kalitesi ve israfına ilişkin problemleri başarılı şekilde ele alması sebebiyle tercih edilmektedir. Ayrıca gıda tedarik sürecinde yapay zeka teknolojilerinin kullanımının yiyecek ve içecek tedarik zinciri için sürdürülebilir bir finansman akışı sağladığı da bilinmektedir (Olan, Liu, Suklan, Jayawickrama, & Arakpogun, 2022). Hindistan’da Gıda Tedarik Zincirinde bu yöntemin kullanımını etkileyen kritik başarı faktörlerinin incelendiği çalışmada, teknoloji organizasyonu, çevre ve insan teorisine dayanan kavramsal bir çerçeve kullanılmıştır. SWARA tekniği ile elde edilen sonuçlara göre, teknoloji hazırlığı, güvenlik, mahremiyet, müşteri memnuniyeti, fayda, talep değişkenliği, yasal uyumluluk, rakip baskısı ve ortaklar arasında bilgi paylaşımı parametrelerinin varlığı ve değerlendirilmesinin yapay zeka yöntemlerinin kullanılmasında etkili olduğu belirtilmiştir (Dora, Kumar, Mangla, Pant, & Kamal, 2022).

Gıda tedarik zinciri sürecinin başka bir paydaşı ise tedarikçilerdir. Perakende sektöründe, yüksek satış hacmi olan ve tüketici tarafından günlük olarak tüketilen et, süt ürünleri gibi taze gıda ürünlerinde tedarikçi seçimi maliyet, verimlilik ve kaliteyi etkileyen çok kriterli bir problemdir. Bu sebeple tedarikçi seçiminde çok kriterli karar verme yöntemlerinden faydalanılmaktadır. Muğla bölgesinde yürütülen çalışmada, uzman görüşü ve literatürden faydalanılarak tedarikçi seçiminde öne çıkan kriterler AHP (Analytic Hierarchy Process) yöntemi ile belirlenmiştir. Analiz sonucunda en önemli ana kriterlerin kalite, sevkiyat, maliyet, firma profili ve servis olduğu belirtilmiştir (Tursun & Özkoç, 2019).

Gıda tedarik sürecinde üretim, hazırlama ve dağıtım aşamalarında her an kontaminasyon meydana gelmesi riski bulunmaktadır. Bu riskleri bertaraf etmek amacıyla, Radyo Frekansı Tanımlama (RFID) ve Raspberry Pi tabanlı cihazları kullanılarak IoT tabanlı gözetim programı oluşturulmuştur (Rao et al., 2022)preparation and distribution process, contamination can occur at any time. The perishable food supply there-

fore needs to pay close attention to food efficiency and safety. The health and quantity of food was assured by the traceability system, as it provides comprehensive food data. The IoT-based surveillance program proposed in this study employed Radio Frequency Identification (RFID). Tüm üretim ve nakliye aşamasında RFID okuyucu tarafından izlenen ürünlerin Raspberry Pi ile nem ve sıcaklık bilgileri alınarak, belirlenen eşik değerinin aşılması durumunda cihazın uyarı vermesini sağlayacak, gelişmiş bir uyarı modelinin oluşturulmasında makine öğrenmesi yöntemi kullanılmıştır. Sonuçların, önerilen yöntemin ürünleri takip edebildiğini ve önceki yaklaşımlara kıyasla sensör bilgilerini doğru ve başarılı bir şekilde tahmin edebildiği ifade edilmiştir.

Doğası gereği karmaşık, çok miktarda ve genellikle yetersiz kalite ve yapıya sahip olan çoklu veri setlerinin entegrasyonunu ve analizini ifade eden büyük veri (big data - BD), gıda işletmelerinde ortaya çıkan risk ve operasyonel zorlukları ele almakta kullanılmaktadır. Tedarik zinciri yönetiminde bu güncel teknolojinin faydalarının tartışıldığı çalışma sonucunda, anlamlı bilgi ve tahminler üretmesi, karar destek yapısı sunması, verimlilik performansını iyileştirmesi gibi özellikleri nedeniyle tedarik sürecinde etkili kullanılması önerilmiştir (Rejeb, Keogh, & Rejeb, 2022)

Popülerliği her geçen gün daha da artan blokzincir teknolojisi, süreçlerin izlenebilirliğini kolaylaştırması, güvenilirliği ve verimliliği artırması nedeniyle tedarik zincirlerinin yeniden tasarlanmasında sıklıkla kullanılmaktadır (Yıldızbaşı & Üstünyer, 2019). Blokzinciri teknolojisi, başkası tarafından müdahale edilemeyen ve bu nedenle güvenilir olduğu varsayılan, dijital olarak imzalanan bir hesap hareketi kayıdır (Dabbagh, Sookhak, & Safa, 2019). Kripto para birimi işlemlerinin yapılmasına olanak sağlamak amacıyla oluşturulan blokzincir teknolojisini diğer bilgi sistemlerinden ayıran en önemli özellikler; merkezi olmayan yapısı, güvenliği, denetlenebilirliği ve akıllı uygulama olmasıdır (Saber, Kouhizadeh, Sarkis, & Shen, 2019). Yıldızbaşı ve Üstünyer (2019), blokzincir teknolojisinin temel prensipleri ve bir tedarik zincirinde bu teknolojiye nasıl yararlanılacağı üzerine fikirlerini sunarak, Türkiye'nin gündeminde olan hal yasası uygulaması kapsamında sebze ve meyvelerin bir aracıya ihtiyaç duymadan tedarik zinciri yönetim sürecine blokzinciri teknolojisine katkı sağladığını değerlendirmişlerdir. Blokzinciri teknolojisine sağladığı içerik bilgisinin şeffaflığı, doğrulama süreci, tarımsal gıdaların güvenliği, ürün takip izlenebilirliğine katkısı sebebiyle farklı tedarik zincirlerine uyarlanabilir olduğunu belirtmişlerdir (Yıldızbaşı & Üstünyer, 2019).

Tarımda tedarik zinciri yönetiminde, müşteri sadakatini ve memnuniyetini artırmak ve gıda güvenliğini sağlamak için izlenebilirlik önemli bir unsurdur. Tedarik sürecinde üreticiden tüketiciye tedarik zinciri boyunca şeffaflığın ve kalite güvencesinin garanti edildiği yeni bir yaklaşıma ihtiyaç duyulmaktadır. Bu nedenle mevcut tedarik zinciri modellerinde, tedarik zincirinin birimleri arasındaki iletişim eksikliği, ürünün lojistik hareketleri ve menşei hakkında bilgi olmaması gibi dezavantajları bertaraf etmek için güncel teknolojilerle çözüm aranmaktadır. Bunlardan biri olan Blok zincir (blockchain) teknolojisi kullanılarak tedarik zinciri sürecine şeffaflık ve izlenebilirlik kazandırılması, işlem kayıtlarının takibinin sağlanması bu sayede tüm tedarik zinciri için güvenliğin artırılması amaçlanmaktadır. Ehsan vd. (2022), çeşitli çiftçiler ve paydaşlar arasında sistemin bütünlüğünü ve şeffaflığını sağlayan blok zinciri tabanlı, tamamen merkezi olmayan bir izlenebilirlik modeli sunmuşlardır. Oluşturulan yeni modelin, geleneksel tedarik zincirinin dezavantajlarının çoğunu ortadan kaldırdığı ifade edilmiştir (Ehsan et al., 2022).

Geleneksel tarımsal gıda tedarik zinciri modeline alternatif olarak kullanılabilecek blok zincir destekli akıllı tarımsal gıda tedarik zinciri modellerinin değerlendirildiği çalışmada ise, bu popüler yöntemin gıda tedarik zinciri uygulamalarında dünyadaki gelişiminin incelenerek ve ülkemiz açısından uygulanabilirliği tartışılmıştır (Kılıç, 2023).

Gıda tedarik zinciri sürecinde blokzinciri teknolojisinin kullanıldığı ampirik çalışmada, tedarik, ön işleme, lojistik, depolama, envanter yönetimi, dağıtım, perakendecilik, işleme ve pazarlama faaliyetleri gibi farklı tedarik zinciri faaliyetleri için hazırlanmış gıda endüstrisi yöneticilerinden elde edilen anket verileri analiz edilmiştir. Analiz sonuçlarına göre, blokzinciri teknolojisinin maliyet azaltma, izlenebilirlik, zaman tasarrufu, değişmezlik, kimlik doğrulama ve çalışma kanıtı gibi avantajlarının yanı sıra mevcut çalışanların bilgi eksikliği, sınırlı ölçeklenebilirlik, kullanımda karmaşıklık ve yüksek başlangıç maliyeti gibi sınırlılıkları olduğu paylaşılmıştır (David, Kumar, & Paul, 2022).

Blokzinciri ve gıda tedarik zincirine ilişkin yapılmış başka bir deneysel çalışmada sinyal teorisine dayanarak, gıda ürünlerinin izlenmesinde blokzincir kullanımının tüketicilerin ürün kalitesi algısı ve satın alma davranışlarını nasıl etkilediği değerlendirilmiştir (Treiblmaier & Garas, 2023). Buna göre, gıda tedarik zincirlerinde blokzincir kullanımının tüketicilerin algılanan gıda kalitesini güçlendirmeye yardımcı olduğu belirtilmiştir.

Sonuç ve Değerlendirme

İşletmeler, gelişen küresel ticaret nedeniyle artan rekabette geri kalmamak için satış cirosunu ve buna bağlı olarak karlılıklarını yükselterek, maliyet verimliliklerini artırması gerekmektedir. Bu nedenle firmalar sürekli olarak; periyodik raporlama, ölçüm, değerlendirmeler yaparak, sonuçlara göre yönetmek zorundadırlar. Bu süreçleri etkin işletebilmeleri için gerek işletme içi gerekse dış tedarikçi ve müşterilerine ait parametrelerin yönetimi için ise gelişen bilgi teknolojilerinden faydalanılması gerekmektedir (Sayın & Demirel, 2020).

Gıda tedarik süreci, küresel ölçekte karmaşık ve çok katmanlı bir yapıdır. Bu süreçte karşılaşılan sağlık ve güvenlik riskleri, uluslararası ve yerel faktörlerin etkileşimi ile şekillenmektedir. Artan nüfus, kaynak kıtlıkları, iklim değişikliği ve teknolojik gelişmeler gibi etkenler, gıda tedarik zincirinin performansını doğrudan etkileyerek, gıda güvenliğinin sağlanmasında çok disiplinli ve bütünsel yaklaşımlar gerektirmektedir.

Artan nüfusa paralel olarak, gıda talebinin karşılanması, tedarik zinciri yönetimi, gıda ayırma, üretim geliştirme, gıda kalitesini iyileştirme ve uygun endüstriyel hijyen alanlarında yapay zeka uygulamalarından yararlanılmaktadır (Mavani et al., 2022). Son yıllarda yapay zeka, makine öğrenmesi, blokzincir, nesnelerin interneti ve büyük veri gibi ileri teknolojilerin gıda tedarik sürecine entegrasyonu, hem izlenebilirlik ve şeffaflığı artırmış hem de risk yönetimi ve kalite kontrol süreçlerinde önemli iyileştirmeler sağlamıştır. Özellikle blokzinciri teknolojisi, tedarik zincirindeki verilerin değişmez ve güvenilir kayıt altına alınmasını mümkün kılmakta, bu da tüketici güveninin artırılmasına ve işletmelerin rekabet avantajı elde etmesine katkıda bulunmaktadır.

2000'li yıllarda gıda kalite denetiminde basit makine öğrenmesi algoritmalarının kullanımı ile başlayan yapay zeka uygulamaları, görüntü analizi ve biyosensörler ile sistemin geliştirilmesi, gıda ürünlerinin geri çağırma takibi, yapay zeka ve nesnelerin interneti entegrasyonu sayesinde tedarik zincirinde gerçek zamanlı izlenebilirliği, doğal dil modelleri ile gıda güvenlik sürecinde otomatik raporlama, sosyal medya ve kullanıcı geri bildirimlerin değerlendirilmesi gibi bir çok alana yayılmıştır. Tedarikçi seçiminde kalite, maliyet ve lojistik gibi çok kriterli faktörlerin dengelenmesi kritik olup, buna yönelik analitik ve karar destek sistemlerinin kullanımı giderek önem kazanmaktadır.

Gıdaların sınıflandırılması, gıda kalitesi, diyet sağlığı ve insan güvenliği denetiminin yapay zeka yöntemleriyle yürütülmesiyle ilgili tüm çalışmalar, gelecekte gıda tedarik zinciri sürecindeki sorunları daha kolay çözüme kavuşturulabilecektir. Bu yöntemler sayesinde iş süreçleri kolay-

laşacak ve bu süreçler daha verimli hale gelebilecektir (Khoshhal, Dakhel, Etemadi, & Zeresghi, 2010; Musa, Abdullah, & Al-Haiqi, 2016; Tao, Wu, Zhang, & Sun, 2014)extraction temperature, solvent-to-solid ratio and the solvent composition on the extraction yields of total phenolics and total anthocyanins were taken into account. The extraction process was simulated and optimized by means of artificial neural network (ANN).

Gelecekteki araştırmaların, özellikle küçük ölçekli üreticiler için veri uyumunu iyileştirmeye, model yorumlanabilirliğini artırmaya ve erişilebilirliği sağlamaya odaklanması büyük önem taşımaktadır (Balakrishnan et al., 2025). Sonuç olarak, gıda tedarik sürecinin sürdürülebilir, güvenli ve etkin yönetimi için teknolojik yeniliklerin stratejik olarak benimsenmesi ve tedarik zinciri paydaşları arasında güçlü iş birliği mekanizmalarının geliştirilmesi gerekmektedir. Bu doğrultuda, bütünsel sistem modellerinin geliştirilmesi ve politika yapıcılarının bu sürece yön verecek düzenleyici çerçeveleri oluşturması, gıda güvenliğinde risklerin minimize edilmesi ve tedarik zincirinin dayanıklılığının artırılması açısından elzemdir.

REFERANSLAR

- Arjona-Román, J. L., Hernández-García, R. P., Navarro-Limón, I., Coria-Hernández, J., Rosas-Mendoza, M. E., & Meléndez-Pérez, R. (2017). Heat Capacity Prediction During Pork Meat Thawing: Application of Artificial Neural Network. *Journal of Food Process Engineering*, 40, 1–8. <https://doi.org/10.1111/jfpe.12399>
- Balakrishnan, P., Anny Leema, A., Jothiaruna, N., Assudani, P. J., Sankar, K., Kulkarni, M. B., & Bhaiyya, M. (2025). Artificial intelligence for food safety: From predictive models to real-world safeguards. *Trends in Food Science and Technology*, 163(105153), 1–15. <https://doi.org/10.1016/j.tifs.2025.105153>
- Bouzembrak, Y., Camenzuli, L., Janssen, E., & van der Fels-Klerx, H. J. (2018). Application of Bayesian Networks in the Development of Herbs and Spices Sampling Monitoring System. *Food Control*, 83, 38–44. <https://doi.org/10.1016/j.foodcont.2017.04.019>
- Bouzembrak, Y., & Marvin, H. J. P. (2019). Impact of drivers of change, including climatic factors, on the occurrence of chemical food safety hazards in fruits and vegetables: A Bayesian Network approach. *Food Control*, 97(2019), 67–76. <https://doi.org/10.1016/j.foodcont.2018.10.021>
- Dabbagh, M., Sookhak, M., & Safa, N. S. (2019). The Evolution of Blockchain: A Bibliometric Study. *IEEE Access*, 7, 19212–19221. <https://doi.org/10.1109/ACCESS.2019.2895646>
- David, A., Kumar, C. G., & Paul, P. V. (2022). Blockchain Technology in the Food Supply Chain: Empirical Analysis. *International Journal of Information Systems and Supply Chain Management*, 15(3), 1–12. <https://doi.org/10.4018/IJISSCM.290014>
- Diwane, R. B., Oza, K. S., Nerlekar, N. A., & Chougule, A. M. (2023). Machine Learning in Food Supply Chain Management – A Review. *International Research Journal of Humanities and Interdisciplinary Studies*, 4(5), 56–65.
- Dora, M., Kumar, A., Mangla, S. K., Pant, A., & Kamal, M. M. (2022). Critical Success Factors Influencing Artificial Intelligence Adoption in Food Supply Chains. *International Journal of Production Research*, 60(14), 4621–4640. <https://doi.org/10.1080/00207543.2021.1959665>
- Ehsan, I., Irfan Khalid, M., Ricci, L., Iqbal, J., Alabrah, A., Sajid Ullah, S., & Alfakih, T. M. (2022). A Conceptual Model for Blockchain-Based Agriculture Food Supply Chain System. *Scientific Programming*, 2022, 1–15. <https://doi.org/10.1155/2022/7358354>
- Focker, M. M., van der Fels-Klerx, H. J. I., & Oude Lansink, A. G. J. M. A. (2018). Systematic Review of Methods to Determine the Cost-Effectiveness of Monitoring Plans for Chemical and Biological Hazards in the Life Sciences. *Comprehensive Reviews in Food Science and Food Safety*, 17(3), 633–645. <https://doi.org/10.1111/1541-4337.12340>

- Hussain, M. A., & Dawson, C. O. (2013). Economic impact of food safety outbreaks on food businesses. *Foods*, 2(4), 585–589. <https://doi.org/10.3390/foods2040585>
- Kang, Z., Zhao, Y., Chen, L., Guo, Y., Mu, Q., & Wang, S. (2022). Advances in Machine Learning and Hyperspectral Imaging in the Food Supply Chain. *Food Engineering Reviews*, 14(4), 596–616. <https://doi.org/10.1007/s12393-022-09322-2>
- Keleş, B., & Ova, G. (2020). Gıda Tedarik Zinciri Yönetiminde Bilgi Teknolojileri Kullanımı. *Adnan Menderes Üniversitesi Ziraat Fakültesi Dergisi*, 17(1), 137–143. <https://doi.org/10.25308/aduziraat.658756>
- Khoshhal, A., Dakhel, A. A., Etemadi, A., & Zereskhi, S. (2010). Artificial neural network modeling of apple drying process. *Journal of Food Process Engineering*, 33, 298–313. <https://doi.org/10.1111/j.1745-4530.2009.00435.x>
- Kılıç, S. (2023). Geleceğin Tedarik Zinciri Modelinin Bugünden Tartışılması : Çiftçi Dostu Blok Zincir Destekli Akıllı Tarımsal Gıda Tedarik Zinciri Modeli Sabiha KILIÇ. *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 13(1), 344–367. <https://doi.org/10.18074/ckuiibfd.1184397>
- Lawley, R., Curtis, L., & Davis, J. (2012). *The food safety hazard guidebook*. Royal Society of Chemistry.
- Marvin, H. J. P., Bouzembrak, Y., van der Fels-Klerx, H. J., Kempenaar, C., Veerkamp, R., Chauhan, A., ... Tekinerdogan, B. (2022). Digitalisation and Artificial Intelligence for sustainable food systems. *Trends in Food Science and Technology*, 120(2022), 344–348. <https://doi.org/10.1016/j.tifs.2022.01.020>
- Mavani, N. R., Ali, J. M., Othman, S., Hussain, M. A., Hashim, H., & Rahman, N. A. (2022). Application of Artificial Intelligence in Food Industry—a Guideline. *Food Engineering Reviews*, 14(1), 134–175. <https://doi.org/10.1007/s12393-021-09290-z>
- Musa, K. H., Abdullah, A., & Al-Haiqi, A. (2016). Determination of DPPH free radical scavenging activity: Application of artificial neural networks. *Food Chemistry*, 194, 705–711. <https://doi.org/10.1016/j.foodchem.2015.08.038>
- Olan, F., Liu, S., Suklan, J., Jayawickrama, U., & Arakpogun, E. O. (2022). The Role of Artificial Intelligence Networks in Sustainable Supply Chain Finance for Food and Drink Industry. *International Journal of Production Research*, 60(14), 4418–4433. <https://doi.org/10.1080/00207543.2021.1915510>
- Rao, U. V. A., Nalinipriya, G., Jangirala, S., Poonguzhali, I., Azahad, S., & Samatha, B. (2022). Improved Food Supply Chain Forecasting Management Using IoT and Machine Learning. *2022 IEEE 2nd Mysore Sub Section International Conference*, 18–21. <https://doi.org/10.1109/MysuruCon55714.2022.9972630>
- Rejeb, A., Keogh, J. G., & Rejeb, K. (2022). Big data in the food supply chain: a literature review. *Journal of Data, Information and Management*, 4(1), 33–47. <https://doi.org/10.1007/s42488-021-00064-0>

- Saberi, S., Kouhizadeh, M., Sarkis, J., & Shen, L. (2019). Blockchain Technology and Its Relationships to Sustainable Supply Chain Management. *International Journal of Production Research*, 57(7), 2117–2135. <https://doi.org/10.1080/00207543.2018.1533261>
- Sahni, V., Srivastava, S., & Khan, R. (2021). Modelling Techniques to Improve the Quality of Food Using Artificial Intelligence. *Journal of Food Quality*, 2021, 1–10. <https://doi.org/10.1155/2021/2140010>
- Sayın, A. A., & Demirel, R. (2020). Tedarik Zinciri Yönetiminde Bilgi Teknolojilerinin Önemi - Gıda Firma Uygulaması. *Turkish Studies-Information Technologies and Applied Sciences*, 15(1), 109–126. <https://doi.org/10.29228/turkishstudies.41540>
- Tao, Y., Wu, D., Zhang, Q. A., & Sun, D. W. (2014). Ultrasound-assisted extraction of phenolics from wine lees: Modeling, optimization and stability of extracts during storage. *Ultrasonics Sonochemistry*, 21, 706–715. <https://doi.org/10.1016/j.ultsonch.2013.09.005>
- Thomas, M. K., Vriezen, R., Farber, J. M., Currie, A., Schlech, W., & Fazil, A. (2015). Economic cost of a listeria monocytogenes outbreak in Canada, 2008. *Foodborne Pathogens and Disease*, 12(12), 966–971. <https://doi.org/10.1089/fpd.2015.1965>
- Treiblmaier, H., & Garaus, M. (2023). Using Blockchain to Signal Quality in the Food Supply Chain: The Impact on Consumer Purchase Intentions and the Moderating Effect of Brand Familiarity. *International Journal of Information Management*, 68(2023), 102514. <https://doi.org/10.1016/j.ijinfomgt.2022.102514>
- Tursun, S. S., & Özkoç, H. (2019). Taze Gıda Ürünlerinde Tedarikçi Seçim Kararı Üzerine Karma Bir Araştırma. *Girişimcilik İnovasyon ve Pazarlama Araştırmaları Dergisi*, 3(6), 75–94. <https://doi.org/10.31006/gipad.632710>
- van Asselt, E. D., Noordam, M. Y., Pikkemaat, M. G., & Dorgelo, F. O. (2018). Risk-based monitoring of chemical substances in food: Prioritization by decision trees. *Food Control*, 93(2018), 112–120. <https://doi.org/10.1016/j.foodcont.2018.06.001>
- Yıldızbaşı, A., & Üstünyer, P. (2019). Tarımsal Gıda Tedarik Zincirinde Blokzincir Tasarımı: Türkiye’de Hal Yasası Örneği. *Bartın Orman Fakültesi Dergisi*, 21(21), 458–465. <https://doi.org/10.24011/barofd.584025>

Bölüm 2

GIDA GÜVENLİĞİNDE YAPAY ZEKA

Aysel GÜLBANDILAR¹, Serel ÖZMEN AKYOL²

1 Doç. Dr., Eskişehir Osmangazi Üniversitesi, Ziraat Fakültesi Gıda Mühendisliği Bölümü. aysel.gulbandilar@ogu.edu.tr ORCID No: 0000-0001-9075-9923

2 Dr. Öğr. Üyesi, Eskişehir Osmangazi Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü. sozmen@ogu.edu.tr ORCID No: 0000-0002-5344-4065

GİRİŞ

Sağlıklı yaşam için güvenli gıda temini kritik öneme sahiptir ve gıda güvenliği, toplumsal sağlık, ekonomi ve sürdürülebilirlik açısından büyük bir gerekliliktir. Yapay zeka, sağladığı maliyet ve zaman avantajı ile gerçek dünya problemlerinin çözümünde matematiksel modellemeye göre üstünlük sunmaktadır (Ekici, Simsek, Ozturk, Sagdic, & Yetim, 2014; Marini, 2009). Bu sebeple üretkenlik ve sürdürülebilirlik hedeflerine verimli ve etkili bir şekilde ulaşmak için gıda mühendisliği de dâhil olmak üzere birçok disiplinde uygulanmaktadır (Sahni et al., 2021).

Son yıllarda gıda güvenliği alanında yapay zeka (YZ), bitki hastalıklarının erken teşhisi, patojen tespiti, raf ömrü tahmini, kalite kontrolü ve sahtecilik önleme gibi uygulamalarda sıklıkla kullanılmaktadır. Gıda endüstrisinde YZ, bilgisayarla görme, IoT, bulut bilişim ve blokzincir entegrasyonları ile gıda kalitesi ve güvenliğini izlemede otomatik, hızlı ve güvenilir yaklaşımlar geliştirilmesine olanak vermektedir. Bu teknolojiler, gıda ürünlerinin izlenebilirliğini artırarak sahteciliği azaltmakta ve tüketicinin güvenliğini güçlendirmektedir. Yapay zekâ, gıda güvenliğinde risklerin erken tespiti, kalite kontrolü, izlenebilirlik ve sürdürülebilirlik alanlarında devrim yaratmakta, sektörün dijital dönüşümünde kritik bir rol oynamaktadır. Bu sayede gıda sektöründe hem üreticiler hem tüketiciler için güvenli ve kaliteli gıda sistemleri oluşturulmaktadır.

Gıda Bilimleri ve Mühendisliğinde Yapay Zeka Uygulamaları

Yapay zekâ uygulamalarından biri olan yapay sinir ağları, insan beyninin çalışma yapısını taklit ederek mevcut verileri analiz edip, bu verilerden farklı öğrenme algoritmaları ile yeni bilgiler oluşturan matematiksel modellerdir (Öztürk & Şahin, 2018). Yapay sinir ağı (YSA), sinir hücreleri veya nöronlardan oluşan biyolojik sinir sisteminin basitleştirilmiş bir modelidir. Sinir ağı yapısı ise, giriş katmanı, gizli katman ve çıkış katmanlarında farklı nöron sayısı ve transfer fonksiyonları kullanılarak oluşturulmaktadır (Özmen Akyol, 2022).

YSA geçmiş verilerden sonuçlar çıkarma yeteneği ile eksik bilgilerle bile yanıtları tahmin edebilen bir sistem olması sayesinde, geleneksel yaklaşımların uygulanmasıyla çözülmesi zor olan gıda güvenliği ve kalite analizleri, gıda görüntü analizi ve çeşitli termal ve termal olmayan gıda işleme operasyonlarının modellenmesi gibi çeşitli gıda mühendisliği uygulamalarında kullanılmaktadır (Aghbashlo, Hosseinpour, & Mujumdar, 2015; Bhagya Raj & Dash, 2022).

Kabli vd. (2020) laktik asit bakterilerinde (LAB) farklı parametrelerden etkilenebilen EPS üretiminin optimizasyonunda yapay sinir ağları

ve bulanık modelleme yaklaşımlarından yararlanmışlardır (Kabli et al., 2020). Leuconstoc mesenteroides DL1 suşundan bir EPS'nin yapısını karakterize etmeyi ve entegre YSA ve bulanık modelleme yaklaşımları kullanarak giriş parametreleri olan inkübasyon süresi, sıcaklığı, sakaroz konsantrasyonu ve ilk levan konsantrasyonunun etkilerini belirleyerek EPS üretimini optimize etmeyi amaçlamışlardır. Bulanık model, girdi-çıkı parametrelerinin temel öğelerinin tanınmasına ve sistem tanımlaması için kullanılan YSA'ların gücüne dayalı olarak geliştirilmiştir.

YSA gıda mühendisliğinde uygulamaları olan; ekstraksiyon, ekstrüzyon, kurutma, filtrasyon, konserve, fermentasyon, fırınlama, süt işleme ve kalite değerlendirme gibi süreçlerin modellenmesinde de uygulanmaktadır (Bhagya Raj & Dash, 2022). Ayrıca, çeşitli gıda ürünlerinin depolama sırasında fiziksel, kimyasal, fonksiyonel ve duyuşal özelliklerinin ve belirli saklama koşulları altında raf ömrünün tahmin edilmesinde de bu yöntemden yararlanılmaktadır. Bu süreçlerin çok yönlü olması nedeniyle, duyuşal analizden izlenebilirliğe ve elektronik burun sonuçlarının işlenmesinden tüketici tercihlerinin haritalanmasına kadar farklı gıda alanlarında uygulamalara sahiptir (Y. Huang, Kangas, & Rasco, 2007).

Ekstraksiyon, katıdan sıvıya, sıvıdan sıvıya ve asit-baz ekstraksiyon süreci olmak üzere birçok gıda maddesinde kullanılan bir ayırma yöntemidir. Ekstraksiyon sürecini etkileyen farklı parametreler, partikül boyutu, solvent bileşimi, solventin katıya oranı, ekstraksiyon sıcaklığı ve ekstraksiyon süresidir. Ekstraksiyon uygulamalarında da yaygın kullanıma sahip olan YSA ile ekstraksiyonun modellenmesi için çıktı değişkenlerini; proteinler, pektinler, vitaminler, pigmentler, uçucu yağlar, aroma bileşikler, tat bileşikler gibi ham maddelerden oluşan farklı çözünür bileşenler oluşturmaktadır (Bhagya Raj & Dash, 2022).

En yüksek meyve suyu veriminin elde edilebileceği süreç koşullarının elde edilmesi için genetik algoritma (GA) ile birleştirilmiş yapay sinir ağı (YSA) ve yanıt yüzey yöntemi (RSM) kullanıldığı çalışmada, selüloz enzimi kullanılarak Chironji meyve suyu ekstraksiyonu, çeşitli selüloz konsantrasyonlarında, kuluçka sıcaklığı ve sürelerinde gerçekleştirilmiştir (Pradhan, Abdullah, & Pradhan, 2021). Hem RSM hem de YSA-GA yönteminden elde edilen optimize edilmiş koşullar kullanılarak ekstrakte edilen meyve suyunun çeşitli fizikokimyasal özelliklerini analiz etmenin yanı sıra, Chironji meyve suyundan maksimum verim elde edilmesi sağlanmıştır.

Vakumlu kurutma, mikrodalga vakumlu kurutma ve çeşitli bitki materyallerinin dondurularak kurutulmasında YSA yönteminin kurutma işlemi ve kurutulmuş ürün tepkilerinin tahmin edilmesi ve değeren-

dirmesinde kullanışlı olduğu yapılan çalışmalarda belirtilmektedir (Inyang & Obunikut, 2022). YSA'lar, kurutma işleminde kalınlık, sıcaklık, hız, nem içeriği, kuruma hızı gibi çoklu parametrelerin tahmininin yanı sıra, süreç kontrolü ve simülasyonda ve gıda güvenliği, muhafaza ve kalite kontrol uygulamalarında da karmaşık görevleri modellemek için etkin bir şekilde kullanılmaktadır.

Ülkemizde yapılan bir çalışmada geleneksel (sıcak hava ile), elektrohidrodinamik (EHD) ve EHD-sıcak hava kombinasyonu kurutma yöntemleriyle kurutulan kayısının farklı kurutma parametrelerinin kuruma hızı ve ürünün nem oranı üzerine etkileri YSA ile modellenmiştir. Farklı transfer fonksiyonları ve öğrenme algoritmaları uygulanarak kurutma yöntemleri için en iyi model performansını oluşturan YSA tespit edilmiştir. EHD yöntemi ile tarımsal ürün kurutmanın YSA esaslı yöntemlerle modellenebileceği belirtilmiştir (Kurtulmuş, Polat, & İzli, 2020).

Çoklu parametrelerin bulunduğu modellerdeki tahmin yeteneği ile YSA fırıncılık alanında da yaygın kullanıma sahiptir. Fırıncılık; ekmek, kek ve bisküvi gibi unlu mamullerin üretiminde son ve önemli adımları içermektedir. Pişirme işlemi sırasında, ürünlerde nihai genel kalitelerinden sorumlu olan eş zamanlı ve birleşik fiziksel, kimyasal ve biyokimyasal değişiklikler meydana gelmektedir (Purlis, 2011). Pişirme sürecini etkileyen parametreler; unun kalitesi (unun gluten içeriği, unun nem içeriği, un-su emilimi), işlem parametreleri (pişirme sıcaklığı, pişirme süresi) ve hamur hazırlama sırasında kullanılan farklı işlemler (çırpma, katlama, yoğurma, karıştırma, harmanlama ve silme) olarak ifade edilmektedir. Pişmiş ürünün kalitesi ise, nem içeriği, gözeneklilik, genleşme oranı, doku, duyu kalite ve ürünün raf ömrü açısından değerlendirilmektedir. Buna göre YSA'nın formülasyonuna dayalı olarak çeşitli gıda maddelerinin dielektrik sabiti ve dielektrik kayıp faktörünü tahmin etmek için uygulanabileceği belirtilmektedir.

Etkili gıda saklama yöntemlerinde biri olan konserve teknolojisinde de YSA modellerinden faydalanılmaktadır. Konserve, işleme ve muhafaza için gerekli ve güvenli bir tekniktir. Reçel, jöle, turşu, sebze, et, kümes hayvanları ve deniz ürünleri gibi gıda maddeleri hava geçirmez şekilde kapatılmış kutularda paketlenmekte ve mikroorganizmaları yok etmek için ısı işlem uygulanmaktadır. Konserve, raf ömrünün uzamasının yanı sıra; ürünün duyu özelliklerinin korunması konusunda da avantajlı bir yöntemdir (Enujiugh & Akanbi, 2010).

Konserve işlemini; çözelti özellikleri (çözelti tipi, konsantrasyon), gıda malzemelerinin ısıtma süresi, işleme parametreleri (sıcaklık, süre), kap için kullanılan malzeme tipi ve gıda malzemesi özellikleri (nem

içeriği, pH, termal yayılım) gibi faktörler etkilemektedir. Konserve sürecinde, ürünün kabulü, nihai ürünün pH değeri, mikrobiyal büyüme, duyuusal değerlendirme ve ürünün kalite parametreleri ile oluşturulan YSA modeli, termal olarak işlenmiş konserve gıdaların güvenliği ve kalitesinin tahminde kullanılmıştır (Kseibat, Mittal, & Basir, 2004). YSA modellemesi ile başlangıç gıda sıcaklığının çıktılar üzerinde ihmal edilebilir bir etkiye sahip olduğu ve yanıtları tahmin etmek için beş girdi parametresi arasında en az etkiye sahip olduğu belirtilmiştir. Dört giriş nöronlu YSA modelinin MRE'si, işlem sıcaklığının tahmin edilmesinde %0.3, işlem süresinin tahmin edilmesinde %6.7 ve minimum kalite bozulmasının tahmin edilmesinde ise %2.8 olarak bulunmuştur. Çalışma, YSA'nın konserve sürecini izlemek ve modellemek ve ısıl işlem görmüş konserve gıda ürünlerinin güvenliğini ve tutarlılığını tahmin etmek için uygulanabileceğini ortaya koymuştur.

Reolojik ölçümler, kalite değerlendirmesi, viskozite profili oluşturma, ekstrüder ve karıştırıcılarda stabilite gibi mühendislik hesaplamaları, süreç tasarımı ve gıda ürünlerinin dağılım stabilitesinin standartlaştırılması için gerekli olan bir parametredir (Bhagya Raj & Dash, 2022). Gıda maddesinin akış davranışını tanımlayan reolojik özelliklerin tahmininde de YSA kullanılmaktadır. Buğday unu hamurunun farinografik özellikleri ile kimyasal bileşimi arasındaki ilişkiyi tahmin etmek için yapay sinir ağı (YSA) yönteminin kullanıldığı çalışmada elde edilen model sayesinde farinografik deneylerin en aza indirilerek, ürün kalitesinin iyileştirilmesi, zaman ve maliyetin azaltılmasının mümkün olabileceği belirtilmektedir (Razmi-Rad, Ghanbarzadeh, Mousavi, Emam-Djomeh, & Khazaei, 2007).

Fermantasyon, gıdada biyokimyasal değişikliklerin meydana geldiği, anaerobik koşullar altında mikroorganizmalar (mayalar veya bakteriler) kullanılarak bir alkol veya aside dönüştürüldüğü metabolik bir süreçtir. Bu süreçte enzimler, özellikle amilazlar, proteazlar veya lipazlar, polisakkaritlerin, proteinlerin ve lipidlerin daha basit toksik olmayan ürünlere hidrolizine neden olmakta ve reolojik ve dokusal özellikleri ve amino asitler, antioksidanlar, vitaminler ve mineraller gibi besin içeriğini iyileştirmektedir. Fermantasyon sürecini etkileyen faktörler; hammadde tipi, partikül boyutu, mikroorganizma türü, substrat bileşimi, şeker tipi, besleme/su oranı, karıştırma hızı, sıcaklık, oksijen seviyesi, pH ve fermantasyon süresidir (Bhagya Raj & Dash, 2022). Yapay zeka yöntemlerinin fermantasyon sürecinde uygulandığı çalışmalardan birinde; gıda koruyucusu olarak onaylanmış bir bakteriyosin olan nisin üretimini sağlamak ve fermantasyon ortamını optimize etmek için yapay sinir ağı-genetik algoritma yöntemi (YSA-GA) ile birleştirilerek yanıt yüzey yöntemi (RSM)'nden yararlanılmıştır (Guo et al., 2010) the path of steepest ascent

method (PSA. Çalışmada fermentasyon ortamındaki önemli bileşenleri tanımlamak için Plackett-Burman tasarımı (PBD) kullanılmış, sonrasında, optimal konsantrasyonlarına yaklaşmak için PSA yönteminden yararlanılmıştır. Daha fazla optimizasyon için ise; Box-Behnken tasarım deneyleri uygulanmıştır. Özellikle, test edilen değişkenlerin nisin titreşi (NT) üzerindeki bireysel etkisini ve karşılıklı etkileşim etkisini belirlemek için bir RSM modeli, NT tahmini için bir YSA modeli ve YSA modeline dayalı optimum çözümleri aramak için ise; GA modellerinden yararlanılmıştır.

Raf ömrü çalışmaları, gıda bozulmalarının üstesinden gelmek ve tüketicilere bir süre boyunca yüksek kaliteli ürün sağlamak için önemli bilgiler sağlamaktadır. Depolama sırasında gıda kalitesi öncelikle pH, su aktivitesi, titre edilebilir asitlik, serbest yağ asitleri, peroksit değeri, koruyucular gibi içsel faktörlerden ve depolama koşulları, bağıl nem, işleme ve depolama sırasındaki zaman sıcaklık profili gibi dış faktörlerden etkilenmektedir. Bunlara ek olarak, ambalaj malzemesinin bariyer özelliği (nem, oksijen veya uçucu bileşiklerin difüzyonu) gıda malzemesinin raf ömrünün modellenmesinde hayati bir rol oynamaktadır. Bu parametreler, depolama sırasında gıda maddelerinin mikrobiyal sayısını, duyuşal özelliklerini (renk, doku ve aroma) ve besin değerlerini etkilemektedir (Siripatrawan & Jantawat, 2009). Yapılan bir çalışmada 7-8 °C'de depolanan işlenmiş peynirin raf ömrünü tahmin etmek için gecikmeli tek katmanlı ve çok katmanlı YSA modelleri geliştirilmiştir. YSA modelinde girdi parametreleri olarak, çözünür nitrojen, pH, standart tabak sayısı, maya ve küf sayısı ve spor sayısı, çıktı parametresi olarak da duyuşal skor belirlenmiştir. Analiz sonuçlarına göre, gecikmeli YSA modellerinin işlenmiş peynirin raf ömrünün tahmin edilmesinde uygulanabilir, iyi bir yöntem olduğu belirtilmiştir (Goyal, 2012).

Çoklu doğrusal regresyon ve YSA yöntemlerinin 8 °C, 20 °C ve 30 °C'de depolama sırasında sürülebilir Gouda peynirinin genel kalitesinin tahmin edilmeye çalışıldığı araştırmada ise; daha yüksek belirleme katsayılarına ve daha düşük RMSE değerlerine sahip YSA'lara dayalı modellerin daha doğru sonuçlar ürettiği belirtilmiştir (Stangierski, Weiss, & Kaczmarek, 2019).

Birçok araştırmacı, gıda mühendisliği problemlerinde modelleme ve optimizasyon, görüntü ve örüntü tanıma için yapay zeka yöntemlerini kullanmıştır. Yanıt yüzey yöntemi (RSM), simülasyon, modelleme ve süreçleri optimize etme gibi mühendislikte daha geniş uygulamaları olan, bağımsız değişkenler arasındaki ilişkiyi araştıran matematiksel ve istatistiksel araçların bir kombinasyonudur (Abdullah, Pradhan, Aflah, & Mishra, 2020). Gıda işlemede, yanıt yüzey yöntemi (RSM) yaygın olarak kul-

lanılan bir modelleme türüdür, ancak YSA'nın yüksek performans, daha mükemmel tahmin yeteneği ve doğrusal olmayan problemleri modelleme kabiliyeti ile alternatif bir modelleme yöntemi olduğu belirtilmektedir (S. M. Huang, Kuo, Chen, Liu, & Shieh, 2017).

Gıda mühendisliğinde yapay zeka yöntemlerinin uygulandığı başka bir alan da kapsülleme uygulamalarıdır. Enkapsülasyon, probiyotikler gibi canlı hücreleri ve biyoaktif bileşikleri gıdalara vermek için kullanılan bir taşıyıcı veya duvar materyali içinde aktif ajanları hapsetme işlemidir. Gıda ürünlerinde antioksidanlar, vitaminler, mineraller, aroma bileşikleri ve oleoresinler, katı ve sıvı yağlar, renklendiriciler ve enzimler kapsül-lenebilmektedir. Huang ve arkadaşları yaptıkları çalışmada; zayıf suda çözünürlüğünü ele almak için piceidin'in sulu bir solüsyon olarak bir lipozomal formüle ultrason destekli kapsüllemesini gerçekleştirmiştir (S. M. Huang et al., 2017). Süreç, RSM ve YSA yaklaşımları kullanılarak optimize edilmiştir. Çalışma sonucunda, YSA'nın tahmin doğruluğunun RSM yönteminden daha üstün olduğu ifade edilmiştir.

Benzer şekilde tahmine dayalı mikrobiyal modelleme yaklaşımı, gıda mikrobiyolojisinde güçlü bir araç olarak dikkat çekmektedir. Bu modelleme yaklaşımı sayesinde, farklı koşullarda mikroorganizmaların davranışları açıklanabilmektedir. Mikrobiyal tepkiler, tanımlanmış ve kontrollü koşullar altında ölçülerek; sonuçlar daha sonra enterpolasyon yoluyla yeni koşullara verilen yanıtları tahmin edebilen matematiksel denklemler şeklinde özetlenmektedir. YSA ile tahmine dayalı modelle-menin başarılı bir şekilde uygulanması, uygun modellerin geliştirilmesine bağlı olarak değişmektedir (Raoufy, Gharibzadeh, Radmehr, Khaksar, & Hosseini, 2010) Pakistan and Afghanistan. Although artificial neural network is a sophisticated novel method in predictive microbiology, no study has been fulfilled on *Z. multiflora*. Therefore, the present study is undertaken to predict the effects of *Z. multiflora* essential oil (EO).

Son yıllarda yapay zeka yöntemlerinden biri olan derin öğrenme teknikleri kullanılarak gıda ve bitki üretim alanında hastalık ve zararlı tespiti, tür tespiti ve sınıflandırılması gibi birçok çalışma yapılmıştır. Derin öğrenme, ön işleme gerek duyulmadan evrişimli yapay sinir ağı (CNN) modeli oluşturulabilmekte ve sınıflandırma işlemlerini diğer ma-kine öğrenmesi (machine learning- ML) yöntemlerine göre daha hızlı ve daha doğru bir şekilde yapılabilmesi sebebiyle gıda alanında da derin öğ-renme uygulamaları yer almaktadır (Zhou, Zhang, Liu, Qiu, & He, 2019).

Üzüm çeşitlerinin sahip olduğu ampelografik özellikler kullanılarak Corint, Merlot, Tayfi, Michele Palieri, Narince üzüm çeşitlerinin derin öğrenme teknikleri ile sınıflandırılmıştır (Terzi, Özgüven, & Yağcı, 2023).

Çalışmada 15 katmandan meydana gelen yeni bir CNN modeli oluşturularak 1028 adet görüntü analiz edilmiş ve CNN modelinin %96,10 sınıflandırma başarımları ile beş farklı üzüm türünü sınıflandırdığı ifade edilmiştir.

Gıda tanıma için derin CNN'lerin kullanıldığı başka bir çalışmada derin evrişimli sinir ağı (DCNN) uygulanmış ve gıda görüntülerinden yemek türlerinin, gıda bileşenlerinin ve pişirme yöntemlerinin tanımlanması sağlanmıştır. DCNN'in, büyük verilerde görüntü tanıma yaşanan zorlukların bertaraf edilmesi amacıyla kullanılan son teknoloji olması sebebiyle kullanıldığı bildirilmektedir. Özellikle geleneksel yaklaşımlara göre bu yöntemin avantajının, veri kümeleri için optimum özellik temsillerini uyarlamalı olarak tahmin edebilen bir teknik olduğu da belirtilmektedir (Zhang, Lu, & Zhang, 2016).

Çok görevli CNN kullanarak, hem gıda kategorisi hem de gıda kalorisi tahmininin gerçekleştirildiği çalışmada, büyük veri kullanılarak, çevrimiçi yemek tarifi sitelerinden yiyecek görüntüleri ve kalori açıklamalarının yer aldığı bir veritabanı oluşturulmuştur (Ege & Yanai, 2017). Çalışmada, sunulan gıdaların miktarlarını tahmin etmekten ziyade, aynı gıda kategorileri içinde içeriklere, çeşnilere, soslara veya nasıl pişirileceğine bağlı olarak gıda kalorilerindeki farklılıkların tahmin edilmesi amaçlanmıştır. Bu sayede sadece gıda kategorisi sınıflandırması ile çözülemeyen gıda kalori tahmininin, görüntü analizi ile yapılması sağlanmıştır.

Yapay zeka destekli makine öğrenmesi yöntemleri görüntü işleme alanında kullanılarak, endüstriyel ortamda gıda sınıflandırması problemlerine de sıklıkla uygulanmaktadır. Sınıflandırma problemlerinin çözümü için geliştirilmiş başarılı makine öğrenimi algoritmalarından biri olan Destek Vektör Makineleri (support vector machine – SVM), birçok sınıflandırma probleminin çözümünde başarıyla uygulanmaktadır. Optimizasyon temelli olduğundan sınıflandırma ve genelleme performansı, hesaplama karmaşıklığı ve kullanılabilirlik açısından diğer tekniklere göre daha başarılı bir algoritma olması sebebiyle literatürdeki yerini almıştır (Ru et al., 2017). SVM'in en önemli avantajı, sınıflandırma problemini kareli optimizasyon problemine dönüştürüp çözmesi böylece problemin çözümüne ilişkin öğrenme aşamasında işlem sayısını azaltarak diğer teknik/algoritmalarla göre daha hızlı çözüme ulaşmasıdır. Farklı badem türlerinin sınıflandırılması için çok sınıflı destek vektör makineleri ve yapay sinir ağları olan denetimli öğrenme yöntemlerinden yararlanıldığı çalışmada, sınıflandırma algoritmalarının karşılaştırmalı analizi sonuçlarına göre, sınıflandırma için temel olarak yapay sinir ağları yerine destek vektör makinesi kullanıldığında; badem sınıflandırmasında daha

yüksek doğruluk seviyelerine ulaşıldığı ifade edilmiştir (Halac, Sokic, & Turajlic, 2017).

Gıda Güvenlik Sürecinde Yapay Zeka Uygulamaları

Sağlıklı yaşam için dengeli ve düzenli beslenmeye ihtiyaç vardır. Sağlıklı beslenmede en önemli basamak ise güvenli gıda teminidir. Bu sebeple gıda güvenliği önem arz etmektedir. Gıda güvenliği, toplumsal gelişmeyi etkileyen aynı zamanda ülke ekonomisi ve insanların hayatını da etkileyen önemli bir sorundur. Gıda güvenliği, gıda kaynaklı hastalık ve zararlılardan korunmak için gıdanın kullanım amacına uygun olarak hazırlanması, bozulmadan tüketicilere ulaşması ve onlara zarar vermeden tüketilmesinin sağlanmasıdır. Gıdalarda meydana gelebilecek fiziksel, kimyasal, biyolojik ve her türlü bozulma etkenlerini ortadan kaldırmak için alınan önlemlerin bütünü olarak da ifade edilmektedir (Bouzemrak, Camenzuli, Janssen, & van der Fels-Klerx, 2018).

Son yıllarda insan sağlığını ciddi şekilde tehdit eden çok sayıda gıda güvenliği olayı yaşanmıştır. Gıda kaynaklı hastalıkları önlemek ve gıda güvenliğini sağlamak için patojenik mikroorganizmaların hızlı ve etkili tespiti, halk sağlığı açısından önemlidir. Bu amaçla, gıda güvenliğinin sağlanması ve risklerinin yönetilmesine ilişkin erken uyarı analizi ve risk kontrolü büyük önem taşımaktadır (Geng et al., 2021). Yapay zekanın gıda güvenliğindeki uygulamaları, ürün hastalıkları teşhisi ve verim tahmini gibi tarımsal optimizasyondan akıllı tedarik zinciri tahminine ve gerçek zamanlı izlemeye kadar çeşitlilik göstermektedir (Belwal, Huang, & Zhao, 2025). Derin öğrenme ve açıklanabilir yapay zeka gibi gelişmiş makine öğrenimi modelleri, patates yaprağı ve mısır hastalıkları üzerine yapılan çalışmalarda da kanıtlandığı gibi, bitki hastalıklarının erken tespiti ve sınıflandırılmasında yüksek doğruluk göstermektedir. Bu yapay zeka destekli teşhis araçları, zamanında müdahaleleri ve ürün korumasını kolaylaştırarak tarımsal verimliliği artırmakta ve daha güvenilir gıda tedarikine katkıda bulunmaktadır.

YZ, gıda güvenliği alanında risk değerlendirme ve kontrol stratejilerini önemli ölçüde iyileştirmiştir. Son araştırmalar, kirletici madde tespiti, raf ömrü tahmini, sahteciliğin önlenmesi ve gerçek zamanlı kalite izlemede makine öğrenimi ve bilgisayarlı görme teknolojilerinin kullanımını vurgulamaktadır. Örneğin, transformatör tabanlı ve diğer gelişmiş makine öğrenimi modelleri, gıda güvenliği riskini sistematik olarak değerlendirerek potansiyel tehlikelerin hızlı ve doğru bir şekilde sınıflandırılmasını ve tahmin edilmesini sağlamak için kullanılmıştır (Dimitrakopoulou & Garre, 2025). Yapay zekâ destekli izleme sürecinde, RFID ve biyosensör-

ler gibi sensör teknolojileriyle birlikte, tedarik zinciri ve üretim süreçleri boyunca sürekli, veri odaklı kalite kontrolü sağlanmaktadır.

Geleneksel olarak, gıda kalitesi, yeni analitik tekniklerin geliştirilmesine yönelik acil ihtiyaç göstermektedir. Dördüncü Sanayi Devrimi (Endüstri 4.0), bir dizi dijital teknolojinin ve diğer yenilikçi teknolojik gelişmelerin ortaya çıkmasıyla hızla ilerleme göstermektedir. “Gıda Kalitesi 4.0”, gıda kalitesinin hızlı, güvenilir ve objektif bir şekilde değerlendirilmesini sağlamak için gıda analizinde Endüstri 4.0 teknolojilerinin kullanımına atıfta bulunan yeni bir kavramdır. Gıda Kalitesi 4.0; gıda endüstrisindeki en gelişmiş teknolojileri kullanarak gıda analizinin artan dijitalleşmesi ve otomasyonu ile karakterize edilmektedir. Özellikle yapay zeka, bilgisayarla görü, bulut bilişim, blok zinciri, nesnelerin İnterneti (IoT), akıllı sensörler, robotik ve 3D yazıcı gibi dijital yenilikler GGYS’de tüm bu amaçlara ulaşmak için otomatik, tahribatsız ve ekonomik bir yaklaşım sağlamaktadır (Bouzembrak, Klüche, Gavai, & Marvin, 2019; Hassoun et al., 2023; Sahni et al., 2021). Blok zincir ve Nesnelerin İnterneti (IoT) teknolojilerinin yapay zeka ile entegrasyonu sayesinde gıda ürünlerinin kaynağından tüketiciye kadar orijinallliğini ve güvenliğinin izlenebilirlik ve şeffaflığı sağlanmaktadır (Yang, Jiao, Zouyi, Diao, & Xia, 2025). Bu sayede, küresel gıda tedarik zincirleri sahtecilik ve kontaminasyon gibi risklere karşı güçlendirilirken, sürdürülebilir ve etik üretim uygulamalarının doğrulanmasını da desteklenmektedir (Sari et al., 2025).

Günümüzün gıda sistemlerinin karmaşıklığı ve gelişmişliği, tüketicileri gıda kaynaklı hastalıklardan korumak için en son teknolojileri kullanarak en yüksek gıda güvenliği standartlarını sürdürmeyi ve gıda şirketlerinin geri çağırma maliyetlerini ve olası itibar zararlarını azaltmayı gerektirmektedir. Gıda güvenliği yönetim sistemleri (GGYS) (Food safety management system - FSMS), gıda güvenliği tehlikelerini önlemek, riskleri ve geri çağırımları en aza indirmek için prosedürleri ve izlemeyi içeren entegre bir sistemdir (Kudashkina, Corradini, Thirunathan, Yada, & Fraser, 2022) while reducing the risk of food companies from losing millions of dollars in recall costs and possible reputational damage. Food safety management systems (FSMSs. Modern gıda endüstrisi; tüketicilerin daha iyi gıda kalitesi, güvenliği ve raf ömrüne yönelik artan talepleri, daha fazla ürün çeşitliliği ve yeşil/çevre dostu/sürdürülebilir üretimin benimsenmesi ile çok rekabetçi ve dinamik olarak gelişen bir ortamdır. Bununla birlikte, geleneksel işleme teknolojileri, gıda ürünlerindeki yapısal ve konformasyonel değişikliklerden (örneğin, lipid oksidasyonu ve protein denatürasyonu) dolayı görünüm, renk, tat ve doku gibi duyu kalite özelliklerini etkileyebilmektedir. Bu nedenle, yüksek kaliteli gıda ürünlerine yönelik sürekli artan tüketici taleplerini karşılamak için gıda araştırmacıları ve gıda endüstrisi, yenilikçi işleme ve analitik teknikler

dahil olmak üzere sürekli olarak daha gelişmiş çözümler ve teknolojiler arayışı içerisindeyler (Hassoun et al., 2023).

Yapay Zeka, sorunları ortaya çıkmadan önce belirleyen öncü göstergeler geliştirmek için verileri kullanarak proaktif veri odaklı bir yaklaşım sağlamaktadır. Özellikle doğru verileri toplayan firmaların gıda güvenliği sorunlarının nerelerde ortaya çıkabileceğine dair tahminlerde bulunmalarına olanak sağladığı için yapay zekanın öncü göstergeler oluşturmada doğal bir çözüm olduğu belirtilmekte ve bu göstergeler sayesinde gıda güvenliği sorunlarını tahmin etmede etkili bir çözüm aracı olarak kullanılmaktadır (Kudashkina et al., 2022)while reducing the risk of food companies from losing millions of dollars in recall costs and possible reputational damage. Food safety management systems (FSMSs).

Gıda güvenliği ve kalitesinde, temel olarak gıdanın duyuusal özellikleri, raf ömrü ve tazeliği ile ilgili bir dizi özelliğin yanında; mikrobiyolojik ve teknolojik parametrelerle ilişkili diğer özelliklerin değerlendirilmesi gerekmektedir (Ren, Fang, Yang, & Han, 2022). Değerlendirmede, sıcaklık noktaları veya mikrobiyal yük sayıları gibi verileri analiz etmek için doğrudan yapay zekanın kullanıldığı algoritmalar mevcuttur.

Son zamanlarda gıda güvenliğinde, temel olarak gıda kaynaklı patojenlere ve gıda kaynaklı hastalıklara odaklanan makine öğrenmesi yöntemlerinin kullanıldığı uygulamalara yer verilmektedir. Kullanıcı yanıtlarından elde edilen çevrimiçi verilerden yararlanan bir yöntem olan kitle kaynak kullanımı, makine öğrenimi yaklaşımlarıyla birleştiğinde, gıda güvenliği risk analizi ve risk sürecini yürütmek için yeni bir araç sağlamaktadır (Wang, Bouzemrak, Oude Lansink, & van der Fels-Klerx, 2022).

Modelleme; son zamanlarda uygulamalı mikrobiyoloji alanında ve güvenli gıda üretiminde önemli bir araç haline gelmiştir. Mikrobiyal model geliştirme ve tahmine dayalı mikrobiyoloji alanındaki büyük ilerlemelerin birçoğunun kökeni gıda güvenliği araştırmalarına dayanmaktadır. Mikrobiyal büyümenin matematiksel modelleri, gıda güvenliği araştırmalarında en yaygın kullanılan yaklaşımlardır. Bayes ağları, Monte Carlo simülasyon modellemesi, etmen ve bireysel tabanlı modelleme, yapay sinir ağları gibi alternatifler ve fiziksel süreç modelleri de özellikle risk değerlendirme çalışmaları için daha popüler hale gelmektedir (Esser, Leveau, & Meyer, 2015).

Gıda güvenliğinin değerlendirilmesinde fizikokimyasal parametrelerle ilgili olarak renk; et ve et ürünleri, yağlar veya bal gibi ürünlerde tüketici kabulü üzerinde büyük etkisi olan en önemli parametrelerden birisidir (Milovanovic et al., 2020; Pirmoradi, Mostafaei, Naderloo, & Ja-

vadikia, 2022). Gıda renginin gıda kalitesini tespit etme yaklaşımını etkilediğine inanılmaktadır. Bu nedenle, bilimsel araştırmalar gıda renginin korunmasına yönelmiştir. Aynı nedenle, yasaklanmış renklendirici maddeler, birçok sağlık tehlikesine neden olmalarına rağmen, yaygın bir şekilde gıda katkı maddeleri olarak kullanılmaktadır (Dixit, Khanna, & Das, 2013)an attempt has been made to understand the exposure assessment of food colors through 2 major groups, sweets and savories, at a national level so as to evolve a scientific yardstick to fix levels of colors in commodities based on technological and safety requirement. A vast majority of colored food commodities (83.6%.

Bhargava ve ark. (2021), meyve ve sebzelerin kalitesinin renk, doku, boyut, şekil ve kusurlarına göre değerlendirilmesine yönelik çeşitli yöntemlerin ayrıntılı bir incelemesini sunmuştur. Bu yöntemler ön işleme, segmentasyon, özellik çıkarma ve sınıflandırma adımlarını içermektedir. Meyve ve sebzelerin kalitesini incelemek için araştırmacılar tarafından önerilen farklı algoritmalara dayalı olarak kritik bir karşılaştırma yapılmıştır (Bhargava & Bansal, 2021). Derin sinir ağlarının popüler mimarileri tartışılmış ve derin öğrenmenin gıda kategorisinin zorluklarını ve sorunlarını çözmek için bir veri analiz aracı olarak kullanılabileceği bulunmuştur.

Gıda güvenliği kapsamında ele alınan başka bir husus olan gıda tağşişi; ekonomik kazanç için gıda alıcılarını yanıltmaya yönelik yapılan aldatıcı bir eylemdir. Gıda kalitesinin veya besin değerinin azalması nedeniyle halk sağlığı üzerinde risk oluşturmaktadır. Genel olarak gıda sahtekarlığı ise, değiştirme, ekleme ve çıkarma şeklinde gerçekleşmektedir (Taylan et al., 2021). Gıda sahtekarlığının temel özellikleri gıda yasalarına uymama ve/veya tüketiciyi yanıltma, kasıtlı dolandırıcılık ve finansal kazanç sağlamaktır. Gıda sahtekarlığı ve tağşişini tespit etmek için endüstrilerde kullanılan yöntemler oldukça pahalı, karmaşık yöntemlerdir ve kalite değerlendirme yöntemleri için uzmanlaşmış altyapı ve yoğun el emeği gerektirmektedir. Yapay zeka sayesinde, gıda ürünlerinde sahtekarlık ve tağşiş tespit etmek için son kullanıcı tarafından kullanılabilecek düşük maliyetli otomatik sistemler sunulmaktadır.

Gıda bileşenleri ile yaygın olarak ilişkilendirilen gıda sahtekarlığı, yağ, balık, bal, süt ve süt ürünleri, et ürünleri, tahıl bazlı gıdalar, meyve suları, şarap ve alkollü içecekler, organik gıdalar, baharatlar, kahve, çay ve bazı yüksek oranda işlenmiş gıdalar üzerinde gerçekleştirilmektedir (Hong et al., 2017). Beslenmemiz üzerinde önemli yeri olan bal üzerinde ciddi taklit ve tağşiş söz konusudur. Bal tağşişi aynı zamanda endüstri için de önemli ekonomik kayıplara neden olmaktadır ve tespiti güçtür (Shao et al., 2022). Yapılan bir araştırmada, bal tağşişinin belirlenmesinde

Uyarlanabilir Bulanık Sinirsel Çıkarım Sistemi (ANFIS), Yapay Sinir Ağı (YSA) ve yanıt yüzey yöntemi duyarlılık analizleri kullanılmış; sonuçlar önerilen yöntemlerin geleneksel yöntemlere göre baldaki sahtekarlığı daha yüksek doğruluk ve hızla tespit etmenin mümkün olabileceği göstermiştir (Pirmoradi et al., 2022).

İnsan diyetinin önemli bir bileşenini oluşturan tahıl ürünlerinde bulunan ağır metal tehlikelerini kontrol etmek için risk faktörlerinin ve erken uyarı sistemlerinin sistematik olarak değerlendirmesinin yapıldığı bir çalışmada bir risk değerlendirme modeli oluşturulmuş ve bu model K-means++ algoritması ile birleştirilerek risk düzeyi sınıflandırması gerçekleştirilmiştir. Daha sonra ise, gıda güvenliği risklerine ilişkin erken uyarı sağlayan, çok adımlı bir risk düzeyi tahmini yapmak için derin öğrenme modelleri kullanılmıştır. Sonuçların, önerilen modelin diğer modellere göre daha üstün olduğunu, 7 günlük tahminde %90,47'lik genel doğruluk gösterdiğini ve böylece gıda denetimi departmanının temel gereksinimini karşıladığını göstermiştir (Z. Wang et al., 2022).

Gıda kaynaklı hastalıkların ortaya çıkmasının ve yayılmasının önlenmesi, halk sağlığının iyileştirilmesi için önemli bir ön koşuldur. Gıda kaynaklı patojenlerin tam genom dizilimi kullanılarak rutin gözetimi ve izlenmesi, rutin salgın araştırmaları ve vaka kontrol çalışmalarının yapılması; gıda kaynaklı hastalık epidemiyolojisini incelemek için fırsatlar yaratmaktadır. Salmonella typhimurium 'un genomik kaynak tahmini için 1998–2013 yılları arasında Amerika Birleşik Devletleri'nde yapılan bir çalışmada 8 büyük zoonotik salgından 7'sinin makine öğrenimi yöntemi ile doğru bir şekilde belirlendiği bildirilmektedir (L. Zhou et al., 2019).

Gıda güvenliği izleme süreci, prosedürlerin doğru bir şekilde yürütüldüğünün ve yönetmeliklere uygunluğunun doğrulanması için gıda güvenliği tehlikelerinin varlığına yönelik yapılan düzenli denetimler mekanizması olarak tanımlanmaktadır. Başarılı sonuçlar nedeniyle gıda güvenliği izleme sürecinde makine öğrenmesi yöntemlerinin kullanımı son yıllarda artış göstermiştir (Bouzembrak et al., 2018; Wang, Bouzembrak, Lansink, & van der Fels-Klerx, 2022). Süreçte hangi parametrelerin izleneceğini (hangi gıda ürünleri ve hangi gıda güvenliği tehlikeleri dâhil) ve nasıl yapılacağını (kaç numunenin dahil edileceği) açıklayan bir izleme planı ve bunların nasıl hazırlanıp analiz edileceği gıda güvenliği izlemesinin uygulanması için önemlidir. Burada gıda güvenliği izleme, yapay zekanın yapısı itibarıyla geçmiş veriler analiz edilerek bu öğrenmeden hareketle gelecekteki gıda güvenliği olaylarını veya sonuçlarını tahmin etmeye çalışan model tabanlı bir süreç olarak tanımlanmaktadır. Gıda güvenliği izleme planlarının ve özellikle ürünlerin ve tehlikelerin tanımlanmasına ilişkin süreci değerlendirilmesi gereken, erken uyarı ve

tahmine dayalı modelleme yaklaşımlarından faydalanabilir (Liu, Manstretta, Rossi, & Van Der Fels-Klerx, 2018).

Han vd. (2019), gıda kalitesi güvenliğini tahmin etmek için çoklu gri ilişkisel bir model geliştirmiştir (Han et al., 2019). Önerilen yöntem kullanılarak sterilize süt kalitesi ve güvenliğine ilişkin risk derecelendirme sonuçları elde edilmiş ve sonuçların risk değerlendirmesinde yüksek doğruluğa sahip olduğu belirtilmiştir. Çalışmada; risk kontrol önlemlerinin önceden formüle edilebileceği veya ayarlanabileceği ve gıda üretim, denetim ve yönetim departmanları için önerilerde bulunabileceği bildirilmektedir.

Sonuç ve Değerlendirme

Gıda işleme, emek isteyen uzun süreli bir süreçtir. Yapay zeka sayesinde işleme sürecinde uygun olmayan ürünler tespit edilerek, süreç kısaltılabilmekte, çalışanlarda değişiklik yapılarak üretim artışı sağlanabilmektedir. Bu sayede ekonomik kayıplar da azaltılabilmektedir. Ayrıca bu tekniklerin uygulanmasıyla daha hızlı karar verme, görme ve koku alma duyularının kullanılması da sağlanabilmektedir.

Gıdaların sınıflandırılması, gıda tedarik zincirinin yönetilmesi, çalışanların kişisel hijyen prosedürlerinin takip edilmesi, yeni ürünlerin geliştirilmesi gibi uygulamalar; halk sağlığını da olumlu yönde etkileyecektir. Besin hijyeni açısından yapay zeka uygulamaları; gıda üretiminde bozulmaya neden olan gıda kaynaklı patojenleri belirlemeye ve gıda kaynaklı hastalıkların varlığını azaltmaya yardımcı olabilmektedir. Devam eden çalışmalar ile yapay zekanın güncel teknoloji entegrasyonlarıyla akıllı karar sistemleri de dahil olmak üzere uçtan uca gıda güvenliği güvencesini sağlayacak otonom yapay zeka modellerinin kullanımı söz konusu olacaktır (Kim, Koo, Kim, & Om, 2025).

Yapay zekâ, gıda sektöründe yenilik ve sürdürülebilirliği teşvik ederek dünya çapında daha akıllı, daha güvenli ve daha dayanıklı gıda sistemlerinin önünü açmaktadır Yapay zeka uygulamaları gıda güvenliği risklerini ortaya çıkmadan önce azaltmak ya da ortadan kaldırmak; gıda tedarik zincirinin sağlıklı işleyişine yardımcı olmak gibi birçok fayda sağlayabilen uygulamalardır. Gelecekte diğer alanlarda olduğu gibi gıda endüstrisi alanında da güven veren tahmin ve çözüm araçları olarak başarılı ve etkin bir şekilde daha yaygın kullanılacaktır.

REFERANSLAR

- Abdullah, S., Pradhan, R. C., Aflah, M., & Mishra, S. (2020). Efficiency of tannase enzyme for degradation of tannin from cashew apple juice: Modeling and optimization of process using artificial neural network and response surface methodology. *Journal of Food Process Engineering*, 43, 1–10. <https://doi.org/10.1111/jfpe.13499>
- Aghbashlo, M., Hosseinpour, S., & Mujumdar, A. S. (2015). Application of Artificial Neural Networks (ANNs) in Drying Technology: A Comprehensive Review. *Drying Technology*, 33, 1397–1462. <https://doi.org/10.1080/07373937.2015.1036288>
- Belwal, T., Huang, R., & Zhao, H. (2025). Editorial: Defining the role of Artificial Intelligence (AI) in the food sector and its applications. *Frontiers in Nutrition*, 12(1596699), 1–3. <https://doi.org/10.3389/fnut.2025.1596699>
- Bhagya Raj, G. V. S., & Dash, K. K. (2022). Comprehensive study on applications of artificial neural network in food process modeling. *Critical Reviews in Food Science and Nutrition*, 62(10), 2756–2783. <https://doi.org/10.1080/10408398.2020.1858398>
- Bhargava, A., & Bansal, A. (2021). Fruits and Vegetables Quality Evaluation Using Computer Vision: A Review. *Journal of King Saud University - Computer and Information Sciences*, 33(3), 243–257. <https://doi.org/10.1016/j.jksuci.2018.06.002>
- Bouzembrak, Y., Camenzuli, L., Janssen, E., & van der Fels-Klerx, H. J. (2018). Application of Bayesian Networks in the Development of Herbs and Spices Sampling Monitoring System. *Food Control*, 83, 38–44. <https://doi.org/10.1016/j.foodcont.2017.04.019>
- Dimitrakopoulou, M.-E., & Garre, A. (2025). AI's Intelligence for Improving Food Safety: Only as Strong as the Data that Feeds It. *Current Food Science and Technology Reports*, 3(15), 1–13. <https://doi.org/10.1007/s43555-025-00060-0>
- Dixit, S., Khanna, S. K., & Das, M. (2013). All India Survey for Analyses of Colors in Sweets and Savories: Exposure Risk in Indian Population. *Journal of Food Science*, 78(4), T642–T647. <https://doi.org/10.1111/1750-3841.12068>
- Ege, T., & Yanai, K. (2017). Simultaneous Estimation of Food Categories and Calories with Multi-task CNN. *Proceedings of the 15th IAPR International Conference on Machine Vision Applications, MVA 2017*, 198–201. Nagoya, Japan: MVA Organization. <https://doi.org/10.23919/MVA.2017.7986835>
- Enujiugha, V. N., & Akanbi, C. T. (2010). Optimization of canning process conditions for fermented African oil bean (*Pentaclethra macrophylla* Benth) seeds in three different media. *Agricultural Engineering International: CIGR Journal*, 12(2), 95–103.
- Esser, D. S., Leveau, J. H. J., & Meyer, K. M. (2015). Modeling Microbial Growth and Dynamics. *Applied Microbiology and Biotechnology*, 99(21), 8831–8846. <https://doi.org/10.1007/s00253-015-6877-6>

- Geng, Z., Liu, F., Shang, D., Han, Y., Shang, Y., & Chu, C. (2021). Early warning and control of food safety risk using an improved AHC-RBF neural network integrating AHP-EW. *Journal of Food Engineering*, 292(2021), 110239. <https://doi.org/10.1016/j.jfoodeng.2020.110239>
- Goyal, G. (2012). Shelf Life Estimation of Processed Cheese by Artificial Neural Network Expert Systems. *Journal of Advanced Computer Science & Technology*, 1(1), 32–41. <https://doi.org/10.14419/jacst.v1i1.10>
- Guo, W. L., Zhang, Y. B., Lu, J. H., Jiang, L. Y., Teng, L. R., Wang, Y., & Liang, Y. C. (2010). Optimization of fermentation medium for nisin production from *Lactococcus lactis* subsp. *lactis* using response surface methodology (RSM) combined with artificial neural network-genetic algorithm (ANN-GA). *African Journal of Biotechnology*, 9(38), 6264–6272. <https://doi.org/10.5897/AJB10.256>
- Halac, D., Sokic, E., & Turajlic, E. (2017). Almonds classification using supervised learning methods. *ICAT 2017 - 26th International Conference on Information, Communication and Automation Technologies, Proceedings*, 1–6. IEEE. <https://doi.org/10.1109/ICAT.2017.8171603>
- Han, Y., Cui, S., Geng, Z., Chu, C., Chen, K., & Wang, Y. (2019). Food quality and safety risk assessment using a novel HMM method based on GRA. *Food Control*, 105, 180–189. <https://doi.org/10.1016/j.foodcont.2019.05.039>
- Hassoun, A., Jagtap, S., Garcia-Garcia, G., Trollman, H., Pateiro, M., Lorenzo, J. M., ... Câmara, J. S. (2023). Food quality 4.0: From Traditional Approaches to Digitalized Automated Analysis. *Journal of Food Engineering*, 337(2023), 11216. <https://doi.org/10.1016/j.jfoodeng.2022.111216>
- Hong, E., Lee, S. Y., Jeong, J. Y., Park, J. M., Kim, B. H., Kwon, K., & Chun, H. S. (2017). Modern Analytical Methods for the Detection of Food Fraud and Adulteration by Food Category. *Journal of the Science of Food and Agriculture*, 97(12), 3877–3896. <https://doi.org/10.1002/jsfa.8364>
- Huang, S. M., Kuo, C. H., Chen, C. A., Liu, Y. C., & Shieh, C. J. (2017). RSM and ANN modeling-based optimization approach for the development of ultrasound-assisted liposome encapsulation of piceid. *Ultrasonics Sonochemistry*, 36, 112–122. <https://doi.org/10.1016/j.ultsonch.2016.11.016>
- Huang, Y., Kangas, L. J., & Rasco, B. A. (2007). Applications of Artificial Neural Networks (ANNs) in food science. *Critical Reviews in Food Science and Nutrition*, 47(2), 113–126. <https://doi.org/10.1080/10408390600626453>
- Inyang, U. E., & Obunikut, M. E. (2022). Artificial Neural Network and Their Applications in Food Materials: A Review. *Engineering and Technology Journal*, 07(04), 1235–1248. <https://doi.org/10.47191/etj/v7i4.06>
- Kabli, M., Yilmaz, M. T., Taylan, O., Kaya, Y., İspirli, H., Basahel, A., ... Dertli, E. (2020). An integrated neural-fuzzy methodology for characterisation and modelling of exopolysaccharide (EPS) production levels of *Leuconostoc mesenteroides* DL1. *Computers and Industrial Engineering*, 148(2020), 106619. <https://doi.org/10.1016/j.cie.2020.106619>
- Kseibat, D. S., Mittal, G. S., & Basir, O. A. (2004). Predicting safety and quality of thermally processed canned foods using a neural network. *Transactions*

- of the Institute of Measurement and Control*, 26(1), 55–68. <https://doi.org/10.1191/0142331204tm0104oa>
- Kudashkina, K., Corradini, M. G., Thirunathan, P., Yada, R. Y., & Fraser, E. D. G. (2022). Artificial Intelligence Technology in Food Safety: A Behavioral Approach. *Trends in Food Science & Technology*, 123(2021), 376–381. <https://doi.org/10.1016/j.tifs.2022.03.021>
- Kurtulmuş, F., Polat, A., & İzli, N. (2020). Yapay Sinir Ağları Kullanarak Kayısının Farklı Kurutma Yöntemleriyle Kurutulmasında Kuruma Hızı ve Nem Oranı Parametrelerinin Modellenmesi. *ÇOMÜ Ziraat Fakültesi Dergisi*, 8(2), 261–269. <https://doi.org/10.33202/comuagri.733166>
- Liu, C., Manstretta, V., Rossi, V., & Van Der Fels-Klerx, H. J. (2018). Comparison of three modelling approaches for predicting deoxynivalenol contamination in winter wheat. *Toxins*, 10(267), 1–15. <https://doi.org/10.3390/toxins10070267>
- Milovanovic, B., Djekic, I., Miocinovic, J., Djordjevic, V., Lorenzo, J. M., Barba, F. J., ... Tomasevic, I. (2020). What is the Color of Milk and Dairy Products and How is it Measured? *Foods*, 9(11), 1–17. <https://doi.org/10.3390/foods9111629>
- Özmen Akyol, S. (2022). Estimating Breast Cancer Class Using Artificial Neural Network and Logistic Regression Methods. *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 3(1), 26–31. <https://doi.org/10.53608/estudambilisim.1052793>
- Öztürk, K., & Şahin, M. E. (2018). Yapay Sinir Ağları ve Yapay Zekâ'ya Genel Bir Bakış. *Takvim-i Vekayi*, 6(2), 25–36.
- Pirmoradi, M., Mostafaei, M., Naderloo, L., & Javadikia, H. (2022). Modeling Honey Adulteration by Processing Microscopic Images Using Artificial Intelligence Methods. *Journal of Agricultural Science and Technology*, 24(2), 365–378.
- Pradhan, D., Abdullah, S., & Pradhan, R. C. (2021). Chironji (Buchanania lanzan) fruit juice extraction using cellulase enzyme: modelling and optimization of process by artificial neural network and response surface methodology. *Journal of Food Science and Technology*, 58(3), 1051–1060. <https://doi.org/10.1007/s13197-020-04619-8>
- Purlis, E. (2011). Bread baking: Technological considerations based on process modelling and simulation. *Journal of Food Engineering*, 103(1), 92–102. <https://doi.org/10.1016/j.jfoodeng.2010.10.003>
- Raoufy, M. R., Gharibzadeh, S., Radmehr, B., Khaksar, R., & Hosseini, H. (2010). Predicting the combined effect of zataria multiflora essential oil, PH and temperature on the growth of Staphylococcus Aureus using artificial neural networks. *Journal of Food Safety*, 30, 318–329. <https://doi.org/10.1111/j.1745-4565.2009.00209.x>
- Razmi-Rad, E., Ghanbarzadeh, B., Mousavi, S. M., Emam-Djomeh, Z., & Khazaei, J. (2007). Prediction of rheological properties of Iranian bread dough from chemical composition of wheat flour by using artificial neural

- networks. *Journal of Food Engineering*, 81(4), 728–734. <https://doi.org/10.1016/j.jfoodeng.2007.01.009>
- Ren, Q. S., Fang, K., Yang, X. T., & Han, J. W. (2022). Ensuring the Quality of Meat in Cold Chain Logistics: A Comprehensive Review. *Trends in Food Science and Technology*, 119(2022), 133–151. <https://doi.org/10.1016/j.tifs.2021.12.006>
- Ru, G., Crescio, M. I., Ingravalle, F., Maurella, C., Gregori, D., Lanera, C., ... Rosat, L. (2017). Machine Learning Techniques applied in risk assessment related to food safety. *EFSA Supporting Publications*, 14(7). <https://doi.org/10.2903/sp.efsa.2017.en-1254>
- Shao, Y., Shi, Y., Xuan, G., Li, Q., Wang, F., Shi, C., & Hu, Z. (2022). Hyperspectral Imaging for Non-destructive Detection of Honey Adulteration. *Vibrational Spectroscopy*, 118(2022), 103340. <https://doi.org/10.1016/j.vibspec.2022.103340>
- Siripatrawan, U., & Jantawat, P. (2009). Artificial neural network approach to simultaneously predict shelf life of two varieties of packaged rice snacks. *International Journal of Food Science and Technology*, 44(1), 42–49. <https://doi.org/10.1111/j.1365-2621.2007.01631.x>
- Stangierski, J., Weiss, D., & Kaczmarek, A. (2019). Multiple regression models and Artificial Neural Network (ANN) as prediction tools of changes in overall quality during the storage of spreadable processed Gouda cheese. *European Food Research and Technology*, 245(11), 2539–2547. <https://doi.org/10.1007/s00217-019-03369-y>
- Taylan, O., Cebi, N., Yilmaz, M. T., Sagdic, O., Ozdemir, D., & Balubaid, M. (2021). Rapid Detection of Green-pea Adulteration in Pistachio Nuts Using Raman Spectroscopy and Chemometrics. *Journal of the Science of Food and Agriculture*, 101(4), 1699–1708. <https://doi.org/10.1002/jsfa.10845>
- Terzi, İ., Özgüven, M. M., & Yağcı, A. (2023). Detection of Some Grape Varieties with Deep Learning Techniques. *Turkish Journal of Agriculture - Food Science and Technology*, 11(1), 125–130. <https://doi.org/https://doi.org/10.24925/turjaf.v11i1.125-130.5722>
- Wang, X., Bouzembrak, Y., Lansink, A. G. J. M. O., & van der Fels-Klerx, H. J. (2022). Application of Machine Learning to the Monitoring and Prediction of Food Safety: A Review. *Comprehensive Reviews in Food Science and Food Safety*, 21(1), 416–434. <https://doi.org/10.1111/1541-4337.12868>
- Wang, X., Bouzembrak, Y., Oude Lansink, A., & van der Fels-Klerx, H. J. (2022). Application of Machine Learning to the Monitoring and Prediction of Food Safety: A Review. *Comprehensive Reviews in Food Science and Food Safety*, 21(1), 416–434. <https://doi.org/10.1111/1541-4337.12868>
- Zhang, X. J., Lu, Y. F., & Zhang, S. H. (2016). Multi-Task Learning for Food Identification and Analysis with Deep Convolutional Neural Networks. *Journal of Computer Science and Technology*, 31(3), 489–500. <https://doi.org/10.1007/s11390-016-1642-6>

Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y. (2019). Application of Deep Learning in Food: A Review. *Comprehensive Reviews in Food Science and Food Safety*, 18(6), 1793–1811. <https://doi.org/10.1111/1541-4337.12492>

Bölüm 3

GERÇEK ZAMANLI HAVA KALİTESİ İZLEME VE ANALİZ SİSTEMİ

Engin OĞUZAY¹, Gökhan UÇKAN²

¹ Engin OĞUZAY, enginoguzay@maltepe.edu.tr, Department of Computer Programming, Maltepe University, Istanbul, Türkiye (ORCID: (0000-0003-2030-2981)),

² Gökhan UÇKAN, guckan@pau.edu.tr, Department of Computer Engineering, Pamukkale University, Denizli, Türkiye (ORCID:0009-0004-7938-0401),

1. Giriş

Bu bölüm, simüle edilmiş gerçek dünya senaryolarında Hava Kalitesi İndeksini (AQI) ölçmek için zamana dayalı bir kablosuz sensör ağı (WSN) kullanarak küresel bir sistem geliştirmeyi amaçlayan gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesini araştırıyor. Proje, iklim araştırmacılarının teknik yöntemleri ve gerçek dünya sensörleri ile entegre edildiğinde gerçek zamanlı veriler sunabilen simüle edilmiş bir ortam sağlıyor. Bu bölümün amaçları doğrultusunda, sistemin yeteneklerini ve potansiyel sonuçlarını göstermek için veriler simüle edilir.

İklim değişikliğini çevreleyen devam eden küresel tartışma, hava kirliliğinin çevresel bozulmaya ve halk sağlığı endişelerine katkıda bulunan önemli bir faktör olduğunun altını çiziyor. Artan farkındalık ve teknolojik gelişmelere rağmen, araştırmacılar, çalışmalarını için kapsamlı, yüksek kaliteli veri tabanlarının olmaması nedeniyle genellikle zorluklarla karşı karşıya kalmaktadır.

Doğru, gerçek zamanlı hava kalitesi verileri, etkili azaltma stratejileri ve politikaları geliştirmek için çok önemlidir. Çalışma; küresel bir WSN aracılığıyla hava kalitesi verilerinin sürekli olarak toplanması, işlenmesi ve yayılması için sağlam bir çerçeve oluşturmayı amaçlamaktadır. WSN tarafından toplanan verileri toplamak, sıkıştırmak ve birleştirmek için gelişmiş dağıtılmış grafik algoritmaları kullanılır. Bu algoritmalar, sensör düğümleri tarafından üretilen büyük hacimli verileri verimli bir şekilde yöneterek minimum veri kaybı ve yüksek doğruluk sağlar. İşlenen veriler, çok çeşitli analitik araçları ve kullanıcı sorgularını destekleyen ölçeklenebilir ve kolay erişilebilir merkezi bir veritabanında daha da rafine edildiği ve saklandığı havuz düğümüne iletilir. Kullanıcı dostu bir web uygulaması arayüzü, hava kalitesi verilerini görselleştirmek ve analiz etmek için birincil erişim noktası görevi görür. Bu girişim, gerçek zamanlı küresel hava kalitesi verileri sağlayarak araştırmacıları, genel halkı, çevre kurumlarını, politika yapımcıları ve sağlık kuruluşlarını hava kalitesi koşulları ve potansiyel sağlık riskleri hakkında etkileşimli haritalar, trend analizi araçları ve uyarı sistemleri ile güçlendirmeyi amaçlamaktadır. Hava kirliliğini ele almak ve küresel ısınmayla mücadele etmek, kapsamlı bir inceleme için bol miktarda kaynak ve veri gerektirir. Bu çalışma, mevcut kaynaklardaki boşluğu dolduran entegre, gerçek zamanlı bir hava kirliliği izleme sistemi oluşturmak için ultra modern teknolojiden yararlanıyor.

Toplanan kapsamlı veriler, kapsamlı araştırmaları kolaylaştırır, tahmine dayalı modellerin geliştirilmesini destekler ve hava kalitesi sorunları hakkında halkın farkındalığını ve eğitimini artırır. Bu çalışma,

daha sağlıklı bir ortam yaratmayı amaçlayan daha geniş çabalara önemli ölçüde katkıda bulunur.

Hava kirliliği modelleri ve eğilimleri hakkında ayrıntılı bilgiler sunarak, hedefli müdahaleleri ve politikaları destekler. Ayrıca, sistemin ölçeklenebilir mimarisi, gelecekteki teknolojik gelişmelere uyarlanabilirlik ve daha fazla bölgeyi ve kirleticiyi içerecek şekilde genişlemeyi sağlar.

Sonuç olarak, gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesi, gelecek nesiller için çevreyi ve halk sağlığını korumak için sınırlar ve disiplinler arası işbirliğini teşvik ederek hava kirliliğine karşı küresel mücadelede bir mihenk taşı olmayı amaçlamaktadır.

1.1. Arka plan

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesi, bir kablo-suz sensör ağı (WSN) kullanarak doğru, gerçek zamanlı bir hava kirliliği izleme sistemi geliştirmeyi amaçlamaktadır. Bu proje, hava kalitesi verilerinin toplanmasını, işlenmesini ve erişilebilirliğini geliştirmek için ultra modern araç ve tekniklerden yararlanmaktadır. Son on yılda, hava kirliliğinde önemli bir artış oldu ve küresel ölçekte insan sağlığına, hayvanlara ve bitkilere zarar verdi. Bu zararlı etkileri ele almak için çeşitli araştırma ekipleri, gelişmiş makine öğrenimi tekniklerini kullanarak hava kalitesini tahmin etmeye çalışıyor. Bununla birlikte, doğru tahminlere ulaşmada önemli bir zorluk, hava kalitesiyle ilgili yüksek kaliteli veri kümelerinin sınırlı mevcudiyetinden kaynaklanmaktadır.

1.2. Sorun Bildirimi

Artan farkındalık ve teknolojik gelişmelere rağmen, araştırmacılar, çalışmaları için kapsamlı, yüksek kaliteli veri tabanlarının olmaması nedeniyle genellikle zorluklarla karşı karşıya kalmaktadır. Doğru, gerçek zamanlı hava kalitesi verileri, etkili azaltma stratejileri ve politikaları geliştirmek için çok önemlidir. SICE'nin "Akıllı Yeşil Hava İzleme Sistemi" [1] ve Perfect Pollucon Services'in IoT tabanlı hava kirliliği izleme sistemi [2] gibi mevcut sistemler, değerli bilgiler sağlar, ancak kapsam ve ölçeklenebilirlik açısından sınırlıdır.

1.3. Hedefler

Çalışmasının temel hedefleri şunlardır:

- Küresel olarak hava kirliliği faktörlerini ölçmek ve izlemek için zaman odaklı bir WSN geliştirilmesi;
- Veri işleme ve toplama için dağıtılmış grafik algoritmaları uygulanması;
- İşlenmiş verileri depolamak için bir veritabanı mimarisi oluşturulması;
- Veri erişilebilirliği için kullanıcı dostu bir web uygulaması arayüzü oluşturulması;
- Araştırmacılara ve halka gerçek zamanlı hava kalitesi bilgileri sağlanması.

1.4. Çalışmanın Kapsamı

Bu projenin kapsamı, küresel bir WSN sisteminin oluşturulmasına ve uygulanmasına kadar uzanmaktadır. Bu projenin temel taşı, çeşitli bölgelere dağıtılmış 237 sensör düğümünden oluşan, titizlikle tasarlanmış bir WSN'dir. Sensör düğümleri, ortalama Hava Kalitesi İndeksini ölçerek hava kalitesini sürekli olarak izler. WSN, veri toplama ve iletim verimliliğini optimize ederek sağlam performans ve güvenilirlik sağlamak için tasarlanmıştır.

1.5. Projenin Önemi

Bu projenin önemi, hava kirliliği araştırmaları için kaynaklardaki mevcut boşluğu doldurma potansiyelinde yatmaktadır. Merkezi bir veritabanı ve kullanıcı dostu bir arayüz sunarak, araştırmacılar için verimli veri erişimini kolaylaştırmayı ve halkı dünya çapında hava kalitesi koşulları hakkında eğitmeyi amaçlamaktadır. Toplanan kapsamlı veriler, kapsamlı araştırmaları kolaylaştırabilir, tahmine dayalı modellerin geliştirilmesini destekleyebilir ve hava kalitesi sorunları hakkında halkın farkındalığını ve eğitimini artırabilir.

1.6. Beklenen Sonuçlar

Başarılı bir uygulamanın ardından, bu proje aşağıdaki sonuçlara ulaşmayı öngörmektedir:

- Hava kalitesi ölçümü için dünya çapında işlevsel bir WSN simülasyonu.

- Kapsamlı hava kirliliği verilerini barındıran sağlam bir veritabanı.
- Kullanıcı etkileşimi için erişilebilir bir web uygulaması arayüzü.
- Küresel hava kalitesi koşullarına ilişkin gelişmiş farkındalık ve anlayış.

1.7. Metodolojiye Genel Bakış

Metodoloji, WSN tarafından toplanan verileri toplamak, sıkıştırmak ve birleştirmek için gelişmiş dağıtılmış grafik algoritmalarının kullanılması içerir. Bu algoritmalar, sensör düğümleri tarafından üretilen büyük hacimli verileri verimli bir şekilde yöneterek minimum veri kaybı ve yüksek doğruluk sağlar. İşlenen veriler, daha da rafine edildiği ve ölçeklenebilir ve kolay erişilebilir bir merkezi veritabanında saklandığı havuz düğümüne iletilir. Bu veritabanı, çok çeşitli analitik araçları ve kullanıcı sorgularını destekler. ASP.NET Web Uygulaması (.NET Framework) ile oluşturulan web uygulaması üzerinden erişilebilen harita, web servisleri ile uygulanacaktır. Bu yaklaşım, her tür kullanıcıya açık, kapsamlı ve kullanıcı dostu bir arayüz sağlar.

2. Metodoloji

2.1. Sistem Mimarisi

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesi, hava kalitesi verilerinin verimli bir şekilde toplanmasını, işlenmesini ve sunulmasını sağlamak için birkaç temel modül halinde yapılandırılmıştır. Sistem mimarisi üç ana katmana ayrılmıştır: Kablosuz Sensör Ağı (WSN), Veri Toplama ve Sıkıştırma, Veri Görselleştirme ve Sunum. Bu katmanlar, projenin hedeflerine ulaşmak için sorunsuz bir şekilde etkileşime girerek hava kalitesi izleme için kapsamlı ve kullanıcı dostu bir platform oluşturur.

2.2. Kablosuz Sensör Ağı (WSN) Tasarımı

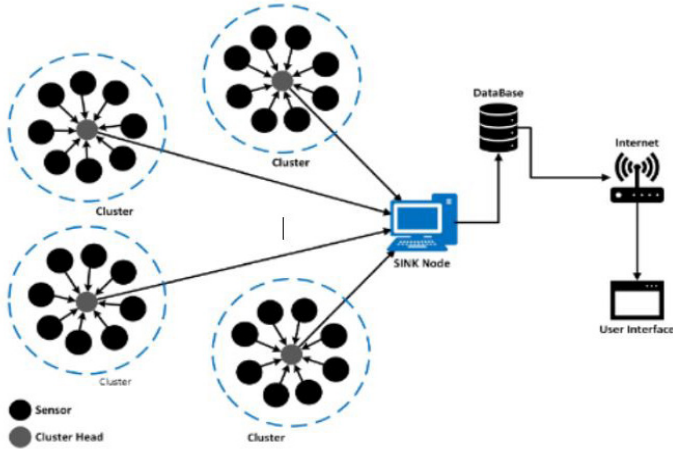
Kablosuz Sensör Ağları (WSN'ler), uzak konumlardan gerçek zamanlı veri toplama olanağı sunarak son birkaç on yılda önemli ölçüde gelişti. Başlangıçta askeri uygulamalar için geliştirilen WSN'ler artık çevresel izleme, sağlık hizmetleri ve endüstriyel otomasyonda yaygın olarak kullanılmaktadır. WSN'lerin birincil avantajı, uzak konumlardan gerçek zamanlı veri sağlama yetenekleridir ve bu da onları hava kalitesini izlemek için ideal hale getirir.

Projedeki WSN, PM2.5 ve CO2 seviyeleri gibi hava kalitesi parametrelerini izleyen, mekansal olarak dağıtılmış otonom sensörlerden oluşur.

Her sensör düğümü veri toplama, işleme ve diğer düğümlerle iletişim kurma yeteneğine sahiptir. Sistem tasarımı şunları içerir:

- **Sensör Düğümleri:** Bu cihazlar çevreden veri toplar ve hava kalitesi parametrelerini ölçmek için stratejik olarak çeşitli konumlara yerleştirilir.
- **Küme Başlıkları:** Birden fazla sensörden veri toplayan ve havuz düğümüne ileten düğümler. Bölgelerindeki belirli sensör düğümü gruplarını yönetirler.
- **Sink Node:** Küme başkanlarından veri toplayan ve depolama ve analiz için veritabanına ileten merkezi düğüm.

Aşağıdaki Şekil 1’de Kablosuz Sensör Ağ Modülünün yapısı gösterilmektedir:



Şekil 1. Kablosuz Sensör Ağı Mimarisi

Sensör düğümleri, her biri bir Küme Baş Düğümü (CHNode) tarafından yönetilen kümeler halinde düzenlenmiştir. CH Düğümleri, ilgili sensör düğümlerinden veri toplar ve merkezi toplama noktası olarak hizmet veren Sink Düğümüne iletir. Bu hiyerarşik yapı, verimli veri iletimi sağlar ve bireysel sensör düğümleri üzerindeki yükü azaltır. [3]

2.3. Veri Toplama Yöntemleri

Ham veriler Havuz Düğümüne ulaştığında, Veri Toplama ve Sıkıştırma Modülü tarafından işlenir. Bu modül, verileri verimli bir şekilde toplamak ve sıkıştırmak için Düşük Enerjili Uyarlanabilir Kümeleme

Hiyerarşisi (LEACH) protokolünü ve çeşitli kümeleme algoritmalarını kullanır.

- **Toplama:** Kapsamlı bir veri kümesi oluşturmak için birden çok sensör düğümünden gelen verileri birleştirir, fazlalığı azaltır ve daha sonraki işlemlere hazırlar.
- **Sıkıştırma:** İletim ek yükünü en aza indirmek için veri boyutunu küçültür.

Hem toplanan hem de sıkıştırılan işlenmiş veriler daha sonra depolama için DatabaseModule'e iletilir. Bu yöntem, minimum veri kaybı ve yüksek doğruluk sağlayarak sistemi sağlam ve güvenilir hale getirir.

2.4. Veri Tabanı Tasarımı

2.4.1. Veritabanı Modülü

Veritabanı Modülü, projedeki sensör ağı tarafından toplanan verilerin depolanması ve yönetilmesi için çok önemlidir. Sistem, Bölgeler, Ülkeler, Sensörler, SensorData, SensorHierarchy, AggregatedData ve UserAccounts dahil olmak üzere yapılandırılmış tabloları korumak için bir MS-SQL veritabanı kullanır.

Veritabanı Yapısı: Veritabanı yapısı, bölgelerin, ülkelerin organizasyonunu ve şehirlerdeki sensörlerin konuşlandırılmasını yansıtır. Verilerin sistematik olarak kategorize edilebildiği ve verimli bir şekilde erişilebildiği hiyerarşik bir düzenleme sağlar. Bu organize yapı, sensör ağı tarafından üretilen çok büyük miktarda veriyi yönetmek için gereklidir.

Veri Yönetimi: Veritabanı modülü, hem ham hem de işlenmiş verileri işlemek için tasarlanmıştır. Bu özellik, gerçek zamanlı verilerin toplandığı sırada depolanmasının yanı sıra analiz için kullanılan işlenmiş ve toplu verilere olanak tanır. Her iki veri türünün de depolanması, geçmiş verilerin alınabilmesini ve analiz edilebilmesini sağlayarak zaman içindeki hava kalitesi eğilimleri hakkında değerli bilgiler sağlar.

Web Servisleri: Modül, veri erişimini ve manipülasyonunu kolaylaştıran ASMX web servislerini içerir. Bu web hizmetleri, veritabanı ve diğer sistem bileşenleri arasında bir arayüz görevi görerek verimli veri alımı, depolama ve güncellemeler sağlar. Sistem, bu web hizmetlerini kullanarak verilerin tutarlı ve güvenilir bir şekilde yönetilmesini sağlayabilir.

Örneğin, SensorData tablosu ReadingID, SensorID, ReadingValue, Timestamp ve Interval gibi çeşitli özniteliklere sahip tek tek sensör okumalarını yakalar. Veritabanı mimarisi, belirli SQL sorgu örneklerine dal-

madan verileri etkili bir şekilde yönetmek için gelişmiş sorguları ve veri işleme işlemlerini destekler.

2.5. Web Uygulama Arayüzü

Web uygulama modülü, ASP.NET çerçevesi kullanılarak geliştirilen, proje sisteminin kullanıcıya dönük bileşenidir. Bu modül, gerçek zamanlı ve geçmiş hava kalitesi verilerini görüntüleyen dinamik web sayfaları sağlar.

Ön Uç Kullanıcı Arayüzü: Kullanıcıların sistemle etkileşime girmesine, hava kalitesi ölçümlerini görüntülemesine ve raporlara erişmesine olanak tanır.

Veri Sunumu: ASMX web servisleri aracılığıyla Veritabanı Modülünden veri alır ve sezgisel ve erişilebilir bir şekilde sunar.

Kullanıcı Özellikleri: Zaman içindeki hava kalitesi eğilimlerini gösteren panoları, zaman çizelgesi çizelgelerini, bölge karşılaştırmalarını ve kapsamlı raporları içerir.

Web uygulaması arayüzü, kullanıcılar için birincil erişim noktası görevi görür ve hava kalitesi verilerini görselleştirmek ve analiz etmek için kullanıcı dostu bir platform sunar. Bu girişim, gerçek zamanlı küresel hava kalitesi verileri sağlayarak araştırmacıları, halkı, çevre kurumlarını, politika yapıcıları ve sağlık kuruluşlarını hava kalitesi koşulları ve potansiyel sağlık riskleri hakkında etkileşimli haritalar, trend analiz araçları ve uyarı sistemleri ile güçlendirmeyi amaçlamaktadır.

3. Kablosuz Sensör Ağı (WSN)

Sistemin iş akışı, yapılandırılmış ve katmanlı bir yaklaşımla hava kalitesi verilerini verimli bir şekilde toplamak, işlemek, depolamak ve görselleştirmek için tasarlanmıştır.

3.1. İş Akışı Açıklaması

3.1.1. Kablosuz Sensör Ağı Simülasyonu

Simülasyon, sensör simülasyon verilerinin başlatılmasıyla başlar. Bu adım, gerçek dünyadaki hava kirliliği ölçümlerini taklit eden sentetik veriler oluşturmak için bir simülatör çalıştırmayı içerir. Oluşturulan sensör verileri toplanır ve çeşitli konumlara dağıtılan çeşitli sensör düğümlerinden gerçek zamanlı hava kalitesi verileri simüle edilir. Toplanan sensör verileri, kapsamlı bir veri kümesi oluşturmak için birden çok sensörden gelen verileri birleştiren, fazlalığı azaltan ve daha sonraki

işlemlere hazırlayan toplamaya tabi tutulur. Toplanan veriler daha sonra merkezi veritabanında saklanır.

3.1.2. Uygulama Katmanında Veri Hazırlama

Sistem veri türünün geçerliliğini kontrol eder. Bu doğrulama adımı, verilerin gerekli format ve standartları karşıladığından emin olur. Geçerli veriler, okunması ve yazılması kolay, hafif bir veri değişim biçimi olan JSON türüne yeniden biçimlendirilir ve bu da onu API etkileşimleri için uygun hale getirir. Yeniden biçimlendirilmiş hava kirliliği verileri API bağlantısı için hazırlanır. Sistem, ek veri almak veya toplanan verileri daha fazla analiz için göndermek için harici API'lere bağlanır. Veriler API aracılığıyla birbirine bağlanarak harici sistemler veya hizmetlerle sorunsuz entegrasyon ve etkileşim sağlar.

3.1.3. Veri Görselleştirme ve Sunum Katmanı

Bağlı veriler görselleştirme için bağlanır. Veri bağlama, veri kaynaklarını kullanıcı arabirimi bileşenlerine bağlama işlemidir ve verilerin web uygulamasında görüntülenmesini kolaylaştırır. Sistem, verileri bir harita üzerinde görselleştirerek hava kalitesi verilerinin coğrafi bir temsilini sağlar.

Araştırmacılar ve halk da dahil olmak üzere kullanıcılar, gerçek zamanlı hava kalitesi verilerini görüntülemek için web sitesine erişir ve harita sayfasına gider. Oturum açmış kullanıcılar, gösterge tablosu ve veri yorumlama araçları gibi ek özelliklere erişebilir. Araştırmacılar, zaman çizelgesi çizelgeleri ve bölge karşılaştırmaları dahil olmak üzere derinlemesine analiz araçlarına erişebilir. Doğrulanmış kullanıcılar bu araçları kullanarak ayrıntılı veri analizi yapabilir. Analiz edilen veriler, gerçek zamanlı ve geçmiş hava kalitesi verilerini görüntüleyen dinamik web sayfaları aracılığıyla sunulur. Bu, zaman içindeki hava kalitesi eğilimlerini gösteren panoları ve kapsamlı raporları içerir.

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi, bu yapılandırılmış iş akışını izleyerek, hava kalitesi verilerinin doğru bir şekilde toplanmasını, verimli bir şekilde işlenmesini ve etkili bir şekilde görselleştirilmesini sağlayarak araştırmacılara, geliştiricilere, kuruluşlara ve genel halka değerli bilgiler sağlar.

3.2. FIFO Tampon ve Senkronizasyon Mekanizmaları

Sensör düğümleri ve küme baş düğümleri arasındaki iletişimi yönetmek için simülasyon bir FIFO (İlk Giren İlk Çıkar) arabelleği kullanır. Semaforlar ve muteksler gibi senkronizasyon mekanizmaları, arabellek üzerinde iş parçacığı güvenli işlemler sağlar. WriteToFifo ve ReadFrom-

Fifo yöntemleri, sırasıyla FIFO arabelleğine ve FIFObuffer'dan veri yazma ve okuma işlemlerini yönetir. Bu yöntemler, veri akışını izlemek, verimli ve doğru veri işlemeyi sağlamak için günlüğe kaydetmeyi içerir. [4]

3.3. Küme Başkanı Seçimi ve Dinamik Görev Değişikliği

Kablosuz Sensör Ağı (WSN) simülasyonu, hava kalitesi verilerini gerçek zamanlı olarak toplamak ve işlemek için sensör düğümlerinin davranışını ve küme kafa düğümleriyle etkileşimini modeller. C# dilinde uygulanan simülasyon, gerçekçi ve verimli bir simülasyon elde etmek için çoklu iş parçacığı, senkronizasyon mekanizmaları ve e-posta bildirimleri gibi gelişmiş özelliklerden yararlanır. Simülasyon, önceden tanımlanmış sayıda sensör düğümü (NUM_SENSORS) ve küme başı düğümü (NUM_CH_NODES) ile başlatılır. Sensör düğümlerine, farklı coğrafi alanlarda dağıtımlarını simüle etmek için benzersiz kimlikler ve bölge kimlikleri atanır. Küme başı düğümleri, kendi bölgelerindeki belirli sensör düğümü gruplarını yönetir. TheSensorDataConfig sınıfı, simülasyonda sensör kimlikleri olarak kullanılan ülke kimliklerini alır. Bu yapılandırmalar izleme amacıyla günlüğe kaydedilir.

3.3.1. LEACH (Düşük Enerji Uyarlamalı Kümeleme Hiyerarşisi)

Düşük Enerji Uyarlamalı Kümeleme Hiyerarşisi (LEACH), enerji verimliliğini artırmak ve ağ ömrünü uzatmak için kablosuz sensör ağları (WSN'ler) için özel olarak tasarlanmış, önde gelen bir küme tabanlı hiyerarşik yönlendirme algoritmasıdır. LEACH, sensör düğümlerini kümeler halinde düzenler ve her kümedeki bir düğüm küme başı (CH) olarak işlev görür. CH'ler, küme üyelerinden veri toplamaktan ve baz istasyonuna (havuz) iletmekten sorumludur [7].

LEACH'de, alınan sinyal gücüne göre kümeler oluşturulur ve yönlendirme kararları vermek için CH'ler seçilir. Bir kümedeki tüm düğümler doğrudan CH ile iletişim kurar. Enerji dağılımını dengelemek için, CH'ler periyodik olarak döndürülür. Bir düğüm, bir tur için belirli bir olasılığa sahip bir CH haline gelir ve bir sonraki tur için yeni bir CH seçilir. Bu olasılığa dayalı seçim, bir CH olmanın enerji yükünün zaman içinde tüm düğümlere eşit olarak dağılmasını sağlar [3] [6].

Kararlı durum aşamasında:

Veri İletimi: Küme üyeleri, verilerini CH tarafından oluşturulan bir Zaman Bölmeli Çoklu Erişim (TDMA) programına göre CH'ye iletir. CH, üyelerinden aldığı verileri toplar, sıkıştırır ve baz istasyonuna iletir. Bu, baz istasyonuna iletilen veri miktarını azaltarak enerji tasarrufu sağlar [3].

LEACH protokolünün çalışması : İki ana aşamadan oluşur: kurulum aşaması ve kararlı durum aşaması. Kurulum aşamasında, düğümler önceden tanımlanmış bir olasılığa dayalı olarak bir CH olup olmayacağına karar verir. Düğümler 0 ile 1 arasında rastgele bir sayı üretir ve bu sayı bir eşik değerinden küçükse, o tur için bir CH olurlar. Eşik değeri $T(n)$, her düğümün P turları üzerinden bir CH olma olasılığının eşit olmasını sağlamak için hesaplanır, Burada P , ağdaki istenen CH yüzdesidir. CH'ler daha sonra diğer düğümlere rollerini belirten bir reklam mesajı yayınlar. CH olmayan düğümler reklamı alır ve alınan sinyalin gücüne göre hangi CH'ye katılacağına karar verir. Daha sonra seçilen CH'ye [7] bir birleştirme isteği gönderirler.

Girişim Azaltma: LEACH, kümeler arasındaki paraziti en aza indirmek için Kod Bölmeli Çoklu Erişim (CDMA) kullanır, çünkü her küme farklı bir CDMA kodları seti kullanır [5].

LEACH'in Avantajları:

Enerji Verimliliği: LEACH, CH rolünü döndürerek ve veri toplamayı kullanarak enerji tüketimini önemli ölçüde azaltır ve ağ ömrünü uzatır.

Ölçeklenebilirlik: Protokol ölçeklenebilirdir ve çok sayıda düğüme sahip ağlarda kullanılabilir.

Kendi Kendine Organizasyon: LEACH kendi kendini organize eder, yani herhangi bir merkezi kontrol veya ağ hakkında küresel bilgi gerektirmez [5].

LEACH'in Dezavantajları:

Rastgele CH Seçimi: CH'lerin rastgele seçimi bazen CH'lerin ağ üzerinde eşit olarak dağılmadığı optimal olmayan kümelemeye neden olabilir. Bu, bazı CH'lerin aşırı yüklenmesine ve diğerlerinin yetersiz kullanılmasına neden olabilir.

Enerji Eşitsizliği: Düşük artık enerjiye sahip düğümlerin, yüksek artık enerjiye sahip düğümlerle CH olarak seçilme olasılığı aynıdır, bu da topremature düğüm arızasına yol açabilir [7]. Geleneksel LEACH'in sınırlamalarını ele almak için çeşitli iyileştirmeler ve varyantlar önerilmiştir. Örneğin, geliştirilmiş bir enerji tasarruflu LEACH protokolü, düğüm kapsama alanını, enerji tüketim oranını ve baz istasyonuna olan mesafeyi dikkate alan yeni bir noderank (NR) parametresi sunar. Bu, CH'lerin enerji seviyelerine ve kapsama alanlarındaki komşu sayısına göre seçilmesini sağlar. Ek olarak, güvene dayalı bir güvenli röle seçim mekanizması, hatalı davranan düğümleri yönlendirme yolundan kaldırarak veri iletiminin güvenliğini artırabilir [7].

3.3.2. HEED (Hibrit, Enerji Verimli, Dağıtık Kümeleme Protokolü)

Hibrit, Enerji Verimli, Dağıtılmış Kümeleme (HEED) protokolü, kablolu sensör ağları (WSN'ler) alanında önemli bir ilerlemedir. HEED, kavramlar üzerine inşa edilmiştir.

HEED, artık enerji ve küme içi iletişim maliyetinin bir kombinasyonuna dayalı olarak küme başkanlarını periyodik olarak seçen çok sekmeli bir kümeleme protokolüdür. Bu hibrit yaklaşım, küme kafası seçimi için yalnızca artık enerjiye dayanan protokollere kıyasla ağ genelinde daha dengeli bir enerji tüketimi sağlar. [8]

HEED, değişen enerji seviyelerine ve ağ topolojisine uyum sağlamak için CH'leri periyodik olarak yeniden seçer. Bu periyodik yeniden seçim, enerji yükünün düğümler arasında daha eşit bir şekilde dağıtılmasına yardımcı olur ve böylece ağ ömrünü uzatır. [9]

HEED'in Operasyonel Aşamaları:

1. Başlatma Aşaması: Düğümler, artık enerjilerine dayalı olarak bir CH olma olasılığıyla başlar. Her düğüm, diğer tüm düğümlerle olan iletişim maliyetini hesaplar ve bu bilgileri yayınlar.

2. Tekrarlama Aşaması: Bu aşamada, henüz bir kümenin parçası olmayan düğümler, her yinelemede CH olasılıklarını ikiye katlar ve sonunda en düşük iletişim maliyetiyle CH kümesine katılır. Bu yinelemeli işlem, tüm düğümler CH'ler veya bir kümenin üyeleri olana kadar devam eder.

3. Sonlandırma Aşaması: CH'ler sonlandırıldıktan sonra kümeler oluşturulur. CHnode olmayanlar, seçtikleri CH'lere birleştirme istekleri gönderir ve CH'ler, LEACH'de kullanılan yaklaşıma benzer şekilde küme üyeleri için bir TDMA programı oluşturur [10].

HEED'in Avantajları:

Enerji Verimliliği: HEED, hem artık enerjiyi hem de iletişim maliyetini göz önünde bulundurarak, ağ genelinde daha dengeli bir enerji tüketimi sağlar ve enerji tükenmesi nedeniyle erken düğüm arızaları olasılığını azaltır.

Ölçeklenebilirlik: HEED ölçeklenebilirdir ve çok sayıda düğüme sahip ağları verimli bir şekilde yönetebilir, bu da onu çok çeşitli WSN uygulamaları için uygun hale getirir.

Geliştirilmiş Ağ Ömrü: CH'lerin mevcut ağ koşullarına göre periyodik olarak yeniden seçilmesi, enerji yükünün eşit olarak dağıtılmasına yardımcı olur ve böylece genel ağ ömrünü uzatır.

HEED'in Dezavantajları:

Yeniden Seçimde Ek Yük: Periyodik yeniden seçim süreci, sık sık yeniden seçimlerin gerekli olduğu son derece dinamik ağlarda bir dezavantaj olabilecek ek iletişim yükü getirir.

Karmaşıklık: CH seçimi için ikili kriterler (artık enerji ve iletişim maliyeti), LEACH gibi daha basit protokollere kıyasla protokole karmaşıklık katar. [9] 10]

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesinin WSN simülasyonu, HEED protokolünden çeşitli öğeleri içerir. HEED'e benzer şekilde, algoritma verimli kümelemeyi sağlamak için hem düğümlerin enerji seviyelerini hem de iletişim maliyetini göz önünde bulundurarak CH seçimi için hibrit bir yaklaşım kullanır. Bu, dengeli enerji tüketiminin sağlanmasına ve ağ ömrünün uzatılmasına yardımcı olur.

3.3.3. Simülasyon Uygulaması

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi algoritması, bu yönleri entegre ederek, veri yönetiminde enerji verimliliğini ve sağlamlığı artırmak için HEED'in güçlü yönlerinden yararlanır ve WSN'lerdeki bazı yaygın zorlukları ele alır.

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesindeki WSN simülasyonu, LEACH ve HEED öğelerini birleştiren özel bir algoritma kullanarak sensör düğümlerini ve etkileşimlerini modeller. Bu hibrit yaklaşım, verimli enerji kullanımı ve sağlam veri toplama sağlamak için her iki protokolün güçlü yönlerinden yararlanır.

Başlatma: Simülasyon, sensör düğümlerinin başlatılmasıyla başlar. Her sensör düğümüne, farklı coğrafi alanlarda dağıtımlarını taklit eden benzersiz bir kimlik ve bir bölge kimliği atanır. Küme başı (CH) düğümleri, belirli sensör düğümü gruplarını yönetmek için tasarlanmıştır. HEED'e benzer şekilde, CH'lerin seçimi, ağ genelinde verimli kümeleme ve dengeli enerji tüketimi sağlamak için artık enerji ve iletişim maliyetinden etkilenir.

Veri Oluşturma: Başlatıldıktan sonra, sensör düğümleri bölge kimliklerine göre rastgele hava kalitesi indeksi (AQI) değerleri üretmeye başlar. Bu değerler, gösteri amacıyla çeşitli bölgelerde farklı kirlilik seviyelerini simüle eder. LEACH'deki veri oluşturma sürecine benzeyen bu

yaklaşım, sensör düğümlerinin CH'lerin toplayabileceği ilgili çevresel verileri sağlamasını sağlar.

Havuz Düğümü İşlemleri: Gerçek zamanlı hava kalitesi izleme ve analiz sistemi simülasyonundaki havuz düğümü, CH düğümlerinden toplanan verileri almaktan sorumludur. ReceiveData yöntemi bu verileri toplar ve merkezi bir veritabanını algılayıcı okumalarıyla güncelleştirir. Bu merkezi yaklaşım, toplanan verilerin sistematik olarak saklanması ve daha fazla analiz için kolayca erişilebilir olmasını sağlar. HEED'deki işleme benzer şekilde, verilerin bulunduğu yer CH'ler tarafından işlendikten sonra lavabo düğümüne iletilir, bu yöntem verimli veri yönetimi ve yedekli iletimlerde minimum enerji harcaması sağlar.

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi simülasyonu, hem LEACH hem de HEED'den öğeleri entegre ederek, enerji verimliliği ve sağlam veri yönetimi arasında bir denge sağlar. HEED'de görüldüğü gibi CH'lerin periyodik olarak yeniden seçilmesi, enerji yükünün düğümler arasında daha eşit bir şekilde dağıtılmasına yardımcı olur ve böylece ağ ömrünü uzatır. Bu arada, LEACH'ten ödünç alınan veri toplama ve sıkıştırma teknikleri, havuz düğümüne iletilen veri miktarının en aza indirilmesini, enerji tasarrufu sağlanmasını ve iletişim ek yükünün azaltılmasını sağlar.

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesi, kümeleme tekniklerini entegre ederek, verimli ve sağlam bir WSN oluşturmak için hem LEACH hem de HEED'in güçlü yönlerinden yararlanır. CH'lerin ilk seçimi ve dinamik olarak yeniden atanması, dengeli enerji tüketimi ve verimli veri iletimi sağlarken, veri toplama ve sıkıştırma, depolama ve iletişim verimliliğini optimize eder.

Avantajlar ve Sınırlamalar

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesindeki WSN simülasyonu ve kümeleme yöntemi çeşitli avantajlar sunar:

Enerji Verimliliği: Sistem, küme başı rollerini döndürerek ve küme başlıklarını dinamik olarak yeniden atayarak enerji tasarrufu sağlar ve ağ ömrünü uzatır.

Veri Bütünlüğü: Toplama ve sıkıştırma, minimum veri kaybı ve yüksek doğruluk sağlar.

Ölçeklenebilirlik: Sistemin mimarisi, daha fazla bölgeyi ve kirleticiyi içerecek şekilde genişletmeyi destekler.

Ancak, dikkate alınması gereken sınırlamalar vardır:

Simülasyon Kısıtlamaları: Bu bölümde kullanılan veriler simüle edilmiştir ve bu da gerçek dünyadaki değişkenliği önemli ölçüde yakalayabilir.

Karmaşıklık: Küme başlıklarının dinamik olarak yeniden atanması, sisteme karmaşıklık katar.

Gelecekteki Geliştirmeler

Gelecekteki çalışmalar, simülasyonlar yerine gerçek dünya sensörlerini birleştirmeye ve veri toplama algoritmalarını geliştirmeye odaklanabilir. Ek olarak, sistem daha fazla bölgeyi kapsayacak şekilde ölçeklendirilebilir ve küresel hava kalitesinin kapsamlı bir görünümünü sağlamak için diğer çevresel izleme ağlarıyla entegre edilebilir.

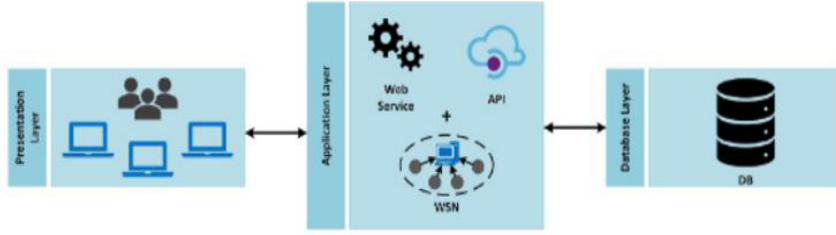
Algorithm 1 Pseudocode for WSN Simulation

```
1: procedure MAIN
2:   sensorIds ← GetCountryIDs(NUM_SENSORS)
3:   simulation ← WSNSimulation(sensorIds, logFilePath)
4:   PRINTSENSORIDS(sensorIds)
5:   SIMULATION.RUNSIMULATION
6:   SIMULATION.DATACONVERSION
7:   SIMULATION.SENDEMAILNOTIF
8: end procedure
9: procedure WSNSimulation(sensorIds, logFilePath)
10:  thisSensorIds ← sensorIds
11:  sinkNode ← new SinkNode(NUM_CH_NODES)
12: end procedure
13: procedure RUNSIMULATION
14:  INITIALIZECLUSTERHEADIDS
15:  sensors ← CREATESENSORNODES(sensorIds)
16:  chNodes ← CREATECLUSTERHEADNODES
17:  sensorCompletionEvent ← new ManualResetEvent(false)
18:  startEvent ← new ManualResetEvent(false)
19:  completedSensorCount ← 0
20:  sensorThreads ← CREATESENSORTHREADS(sensors, startEvent, sensorCompletionEvent)
21:  startEvent.Set()
22:  chThreads ← CREATECLUSTERHEADTHREADS(chNodes)
23:  sensorCompletionEvent.WaitOne()
24:  WAITFORTHEADS(chThreads)
25: end procedure
```

Algoritma 1. Kablosuz Sensör Ağı Simülasyonu

3.4. Proje Mimarisi

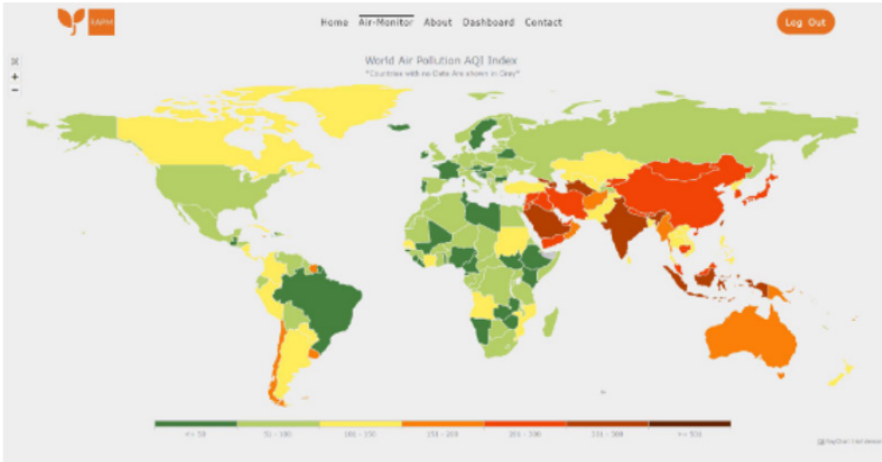
Bu projenin uygulanması, Veritabanı Katmanı, Uygulama Katmanı ve Sunum Katmanı'ndan oluşan Üç Katmanlı bir Mimariye dayanmaktadır. Bu mimari, her katmanın kendi altyapısı üzerinde çalışabilmesi gibi çeşitli avantajlar sunar, bu da ayrı geliştirme ekiplerinin aynı anda çalışmasına ve diğerlerini etkilemeden her katmanı gerektiği gibi güncellemesine veya ölçeklendirmesine olanak tanır. Projenin ayrıntılı mimarisi aşağıdaki blok diyagramda görülebilir :



Şekil 2. 3 Katmanlı Mimari

3.4.1. Sunum Katmanı

Bu, son kullanıcının uygulama ile etkileşime girdiği uygulamanın kullanıcı arayüzü ve iletişim katmanıdır. Proje, ASP. NETFramework, Web Uygulaması arayüzü için dinamik web sayfaları üretmek üzere tasarlanmış bir sunucu tarafı web uygulaması çerçevesidir. [11]



Şekil 3. Kullanıcı Arayüzü Harita Görselleştirme

3.4.2. Uygulama Katmanı

Orta katman olarak da bilinen bu hayati bileşen, sunum katmanından alınan bilgileri işler ve veri katmanıyla etkileşime girer. WSN'deki her sensör düğümü benzersiz bir şekilde tanımlanır ve Küme Baş Düğümleri (CH Düğümleri) altında stratejik olarak gruplandırılır. Küme, verileri merkezi bir Havuz Düğüme iletmeyen önce ilgili sensör düğümlerinden toplayarak veri tıkanıklığını azaltır ve iletim verimliliğini artırır. Bu katman, sunu katmanından verileri işler ve veri katmanıyla iletişimi kolaylaştırır. Ayrıca, veritabanından alınan işlenmiş verileri harici API'le-

rin beklediği biçime dönüştürür ve dönüştürülen verilerle API çağrılarını yapar.

```
Updating ClusterHeadID: sensor UA {AG, 2} ==> {AM, 3}
CH {AM, 3} received data: | sensor UA ID: UA Value: 39 ClusterHeadID: {AM, 3} Finished: True |
***Reading from FIFO: ID: BU, Value: 125, ClusterHeadID: {AU, 5}, Finished: True***

Updating ClusterHeadID: sensor BU {AU, 5} ==> {AM, 3}
CH {AM, 3} received data: | sensor BU ID: BU Value: 125 ClusterHeadID: {AM, 3} Finished: True |
***Reading from FIFO: ID: VU, Value: 178, ClusterHeadID: {AM, 3}, Finished: True***
CH {AM, 3} received data: | sensor VU ID: VU Value: 178 ClusterHeadID: {AM, 3} Finished: True |

All data received by CH {AM, 3}.

-->Writing to FIFO: System.Collections.Generic.Dictionary`2[System.String,System.Int32]<--
CH {AM, 3} done signal sent.
Aggregated Data for CH {AM, 3}:
  Sensor UA: 39
  Sensor BU: 125
  Sensor VU: 178
***Reading from FIFO: ID: BY, Value: 48, ClusterHeadID: {AO, 1}, Finished: True***

Updating ClusterHeadID: sensor BY {AO, 1} ==> {AD, 4}
CH {AD, 4} received data: | sensor BY ID: BY Value: 48 ClusterHeadID: {AD, 4} Finished: True |
***Reading from FIFO: ID: BW, Value: 38, ClusterHeadID: {AD, 4}, Finished: True***
CH {AD, 4} received data: | sensor BW ID: BW Value: 38 ClusterHeadID: {AD, 4} Finished: True |
***Reading from FIFO: ID: UZ, Value: 288, ClusterHeadID: {AG, 2}, Finished: True***

Updating ClusterHeadID: sensor UZ {AG, 2} ==> {AD, 4}
CH {AD, 4} received data: | sensor UZ ID: UZ Value: 288 ClusterHeadID: {AD, 4} Finished: True |

All data received by CH {AD, 4}.
```

Şekil 4. WSN Örnek İşlemler

3.4.3. Veritabanı Katmanı

Bu katman, uygulama tarafından işlenen bilgileri depolar ve yönetir. Veritabanı katmanı ile iletişim tipik olarak web hizmetleri, özellikle bu projedeki ASMX web hizmetleri aracılığıyla kolaylaştırılır. ASMX web hizmetleri, uygulama katmanı ile veritabanı katmanı arasında arabirim görevi görür ve HTTP üzerinden Basit Nesne Erişim Protokolü'nü (SOAP) kullanarak mesajlar gönderir. Web hizmeti, veri alma, kayıtları güncelleştirme veya sorguları yürütme gibi çeşitli veritabanı işlemlerini gerçekleştirmek için uygulama katmanı tarafından çağrılabilen yöntemleri kullanıma sunar. Uygulama katmanı, standart HTTP protokollerini kullanarak, veritabanı etkileşimi için sağlanan yöntemlere özel API çağrılarını yaparak ASMX web hizmeti ile iletişim kurar. [11]

Gerçek zamanlı veri akışını yönetmek için sistem bir İlk Giren İlk Çıkar (FIFO) tampon mekanizması kullanır. FIFO arabelleği, eşzamanlı veri erişimini yönetmek için semaforlar ve muteksler kullanılarak senkronize edilir ve sensör verilerinin düzenli bir şekilde işlenmesini ve iletilmesini sağlar. Tamponun performansı titizlikle günlüğe kaydedilir, sistem işlemleri hakkında bilgi sağlar ve hızlı sorun gidermeye olanak tanır.

4. Sonuçlar

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesiyle ilgili çalışma, WSN teknolojisini kullanarak gerçek zamanlı bir hava kirliliği izleme sistemini başarıyla geliştirdi. Bu proje, gerçek zamanlı veri toplama ve analizi için ölçeklenebilir ve verimli bir çözüm sağlayarak çevresel izleme alanına önemli katkılar sağlama potansiyelini göstermektedir.

5. Tartışma

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesi, gerçek zamanlı küresel hava kalitesi izleme için Kablosuz Sensör Ağlarının (WSN'ler) kullanılmasının fizibilitesini ve etkinliğini başarıyla göstermiştir. Veri toplama ve sıkıştırma için dağıtılmış grafik algoritmalarının kullanımıyla birlikte 237 sensör düğümünün konuşlandırılması, WSN'lerde verimli veri yönetimi ve iletimi hakkında değerli bilgiler sağlamıştır. Hiyerarşik ağ yapısı, veri tıkanıklığını önemli ölçüde azaltarak sağlam performans ve güvenilirlik sağlar.

Sensör kalibrasyonu ve ağ bağlantısı sorunları gibi zorluklarla karşılaşmasına rağmen, proje bu engellerin üstesinden geldi ve büyük ölçekli WSN dağıtımlarının karmaşıklığını ve potansiyelini vurguladı. Ölçeklenebilir bir veritabanının ve kullanıcı dostu bir web uygulaması arayüzünün entegrasyonu, veri erişilebilirliğini artırdı, araştırmacılar, politika yapıcılar ve halk arasında hava kalitesi sorunlarının daha iyi bilinmesini ve anlaşılmasını kolaylaştırdı.

Genel olarak, gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesi çevresel izlemeye önemli katkılarda bulunmuş ve bu alanda gelecekteki ilerlemeler için sağlam bir temel sağlamıştır.

6. Son Sözler

Gerçek zamanlı hava kalitesi izleme ve analiz sistemi projesinin ileri teknikler kullanılarak geliştirilmesi önemli bir başarıdır.. Projenin önemi, hava kirliliği araştırmaları için kaynaklardaki mevcut boşluğu doldurma potansiyelinde, verimli veri erişimini ve dünya çapında hava kalitesi koşulları hakkında eğitimi kolaylaştırmak için merkezi bir veri tabanı ve kullanıcı dostu bir arayüz sunmasıdır.

Projenin hiyerarşik bir sensör ağının yenilikçi kullanımı, kümeleme ve veri sıkıştırma teknikleriyle birleştiğinde, verilerin güvenilir ve verimli bir şekilde iletilmesini sağlar. Sağlam veritabanı tasarımı ve ASMX web hizmetleri, sistemin hava kalitesi verilerini etkin bir şekilde yönetme ve yayma yeteneğini daha da geliştirir.

7. Kaynakça

- [1] Env Air Quality Eng; <https://www.scribd.com/document/617163684/ENV-AIR-QUALITY-ENG-12>.
- [2] “Mükemmel Kirlenici Hizmetleri”; <https://www.ppsthane.com/air-kalite-izleme-test>.
- [3] K. Erciyes, Bilgisayar Ağları için Dağıtık Grafik Algoritmaları, İzmir: Springer Londra, Heidelberg, New York, Dordrecht, 2013.
- [4] P. D. K. ERCİYES, “Ek B.2”; Bilgisayar için Dağıtık Grafik Algoritmaları içinde Ağlar.
- [5] R. I. Tandel, “Uluslararası Bilgisayar Bilimleri ve Enformasyon Dergisi Teknolojiler,” Kablosuz Sensör Ağında Leach Protokolü: Bir Anket, vol. 7(4), s. 3, 2016.
- [6] M. Homaei, “LEACH (Düşük Enerji Uyarlamalı Kümeleme Hiyerarşisi protokolü),” MATLAB Merkezi; <https://www.mathworks.com/matlabcentral/fileexchange/48162-leach-low-energy-adaptive-clustering-hierarchy-protocol>.
- [7] G. S. C. P. Bhukya Suresh, “LEACH Kullanarak Enerji Verimli Güvenli Bir Yönlendirme Şeması IoT ağları için WSN’de protokol,” Ölçüm: Sensörler, vol. 30, 2023.
- [8] D. R. A. Sasikumar M, “Heterojen HEED Protokolünün Performans Değerlendirmesi; Kablosuz Sensör Ağları,” Uluslararası Bilgisayar ve İletişim Mühendisliği İleri Araştırma Dergisi , vol. 3, no. 2, p. 4, 2014.
- [9] S. S. B. K. Satish Chand, “Kablosuz Sensör Ağları için Heterojen HEED Protokolü,” Springer Bilimi, s. 23, 2014.
- [10] A. F. M. S. S. Kareem Hameed Ali, “PEGASIS, HEED ve Kablosuz Sensör Ağlarında LEACH Protokolleri,” Celal Bayar Üniversitesi Fen Bilimleri Dergisi, cilt 19, no. 1, s. 8, 2023.
- [11] “Tüketmek Bir ASP.NET Web Hizmet (ASMX)”; <https://learn.microsoft.com/enus/pre ious versions/xamarin/xamarin-forms/data-cloud/web-services/asmx>.
- [12] “Nedir Bu Üç Katmanlı Mimari”; <https://www.ibm.com/topics/three-tier-architecture>

//

Bölüm 4

Z-SKOR NORMALİZASYONUNUN İRİS VERİ SETİ SINIFLANDIRMA PERFORMANSINA ETKİSİ

Bahadır ÇOKÇETİN¹, Ahmet ÇELİK²

1 Kütahya Dumlupınar Üniversitesi, Tavşanlı Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Kütahya, Türkiye, ORCID: 0000-0003-2189-9265

2 Kütahya Dumlupınar Üniversitesi, Tavşanlı Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Kütahya, Türkiye, ORCID: 0000-0002-6288-3182

Giriş

Makine öğrenmesi ve yapay zekâ uygulamaları, son yıllarda biyomedikal mühendislikten çevresel gözleme, endüstriyel arıza teşhisine kadar birçok alanda veri temelli karar verme süreçlerinin merkezine yerleşmiştir (Jordan & Mitchell, 2015). Bu alanlarda sınıflandırma problemleri, örneklerin önceden tanımlı kategorilere atanmasını sağlayarak, sistemlerin otomatik tanıma ve öngörü yeteneğini geliştirir (Bishop, 2006). *Iris* veri seti, bu tür sınıflandırma algoritmalarının değerlendirilmesi için uzun süredir kullanılan, dengeli ve iyi tanımlanmış bir referans (benchmark) veri kümesidir (Dua & Graff, 2019).

R. A. Fisher tarafından 1936 yılında tanıtılan *Iris* veri seti, her biri 50 örnek içeren üç farklı *Iris* türünden (*setosa*, *versicolor*, *virginica*) oluşur ve dört ölçülebilir özelliğe dayanır: çanak yaprak uzunluğu, çanak yaprak genişliği, taç yaprak uzunluğu ve taç yaprak genişliği (Fisher, 1936). Veri setinin basit yapısı, düşük boyutluluğu ve sınıflar arası ayrışma düzeyinin yüksek olması nedeniyle hem başlangıç düzeyinde eğitim hem de algoritma karşılaştırmaları için ideal bir platform sunar (Lantz, 2019).

Son yıllarda yapılan araştırmalar, klasik makine öğrenmesi yöntemlerinin (örneğin Support Vector Machines, *k*-Nearest Neighbors, Decision Trees, Random Forests) küçük boyutlu ve dengeli veri kümelerinde hâlâ yüksek doğruluk oranları sağladığını göstermektedir (Jain, Duin, & Mao, 2000), (Zhang & Zhao, 2022). Örneğin, Singh ve ark. (2023) tarafından yapılan bir karşılaştırmada, kNN ve SVM algoritmalarının *Iris* veri setinde %95'in üzerinde doğruluk oranına ulaştığı rapor edilmiştir (Singh, Sharma, & Meena, 2023). Benzer biçimde, Shrestha ve ark. (2021) çalışmasında, çok sınıflı veri kümelerinde F1-score ve AUC gibi ek metriklerin, yalnızca doğruluk ölçütüne göre daha güvenilir performans değerlendirmesi sunduğu vurgulanmıştır (Shrestha & Mahmood, 2021).

Bu tür çalışmaların ortak noktası, yalnızca yüksek doğruluk oranları elde etmek değil, aynı zamanda model istikrarı, genelleme kabiliyeti ve yorumlanabilirlik açısından klasik yöntemlerin güçlü yönlerini ortaya koymaktır (Raschka & Mirjalili, 2020). Özellikle tıp, biyometri veya mühendislik gibi alanlarda kullanılan veri kümeleri gürültülü, dengesiz veya zamana bağlı değişim (domain shift) içerebildiğinden, yöntemin yeniden üretilebilir olması büyük önem taşımaktadır (Kaur & Kaur, 2022).

Bu kitap bölümünde, MATLAB ortamında *Iris* veri seti kullanılarak klasik sınıflandırma yöntemlerinin (SVM, kNN ve Random Forest) performans karşılaştırması gerçekleştirilmiştir. Çalışma, verinin alınmasından başlayarak ön işleme, modelleme, doğrulama ve sonuç değerlendirme aşamalarına kadar tüm süreci sistematik biçimde açıklamaktadır. Böylece hem öğretici (didaktik) hem de yeniden üretilebilir (reproducible) bir makine öğrenmesi altyapısı sunulmaktadır. Bu altyapı, ilerleyen çalışmalarda biyomedikal sinyaller (örneğin EKG, PPG), arıza teşhisi veya görüntü tabanlı tanı sistemleri gibi daha karmaşık problemlere uyarlanabilecek niteliktedir.

Bu çalışmanın amacı, klasik makine öğrenmesi yöntemlerinin (SVM, kNN, Random Forest) *Iris* veri seti üzerindeki performansını Z-skor normalizasyonu öncesi ve sonrası karşılaştırmalı olarak değerlendirmektir. Böylece, ön işleme adımlarının sınıflandırma başarısına olan etkisi sayısal olarak analiz edilmiştir.

Veri Kümesi ve Yöntemler

Bu çalışmada, makine öğrenmesi sınıflandırma problemlerinde sıklıkla referans veri seti olarak kullanılan *Iris* veri kümesi kullanılmıştır. Veri kümesi, R. A. Fisher tarafından 1936 yılında tanıtılmış ve üç farklı *Iris* türüne ait (*Setosa*, *Versicolor* ve *Virginica*) toplam 150 gözlem içermektedir. Her sınıf, 50 örnekten oluşmakta ve dört morfolojik özelliğe göre tanımlanmaktadır. Bunlar çanak yaprak uzunluğu (*Sepal Length*), çanak yaprak genişliği (*Sepal Width*), taç yaprak uzunluğu (*Petal Length*) ve taç yaprak genişliği (*Petal Width*) özellikleridir.

Veri kümesi, MATLAB ortamında *fisheriris* değişkeni üzerinden doğrudan erişilebilmekte olup, ayrıca UCI Machine Learning Repository’de kamuya açık olarak da mevcuttur (Dua & Graff, 2019). Dengeli dağılım yapısı nedeniyle yeniden örnekleme (*resampling*) veya veri dengeleme tekniklerine ihtiyaç duyulmamıştır.

İş Akışı

Çalışmanın genel süreci sekiz ana aşamadan oluşmaktadır. Bu adımlar Şekil 1’de özetlenmiştir. Veri seti ilk olarak eğitim (%75) ve test (%25) alt kümelerine stratifiye biçimde ayrılmıştır. Eğitim verisi üzerinde ön işleme (*standardizasyon*) uygulanmış, ardından üç farklı model (SVM, kNN ve Rastgele Orman) eğitilmiştir. Test verisiyle doğrulama aşamasında, çok sınıflı performans ölçütleri hesaplanmış ve sonuçlar Excel tablosu olarak dışa aktarılmıştır.



Şekil 1. Akış Diyagramı

Veri Ön İşleme ve Z Skor Normalizasyonu

Veri ön işleme, makine öğrenmesi sürecinin en kritik aşamalarından biridir. Bu aşama, modelin doğru öğrenmesini engelleyebilecek hataları, gürültüleri ve ölçek farklılıklarını azaltarak daha dengeli bir veri uzayı oluşturur.

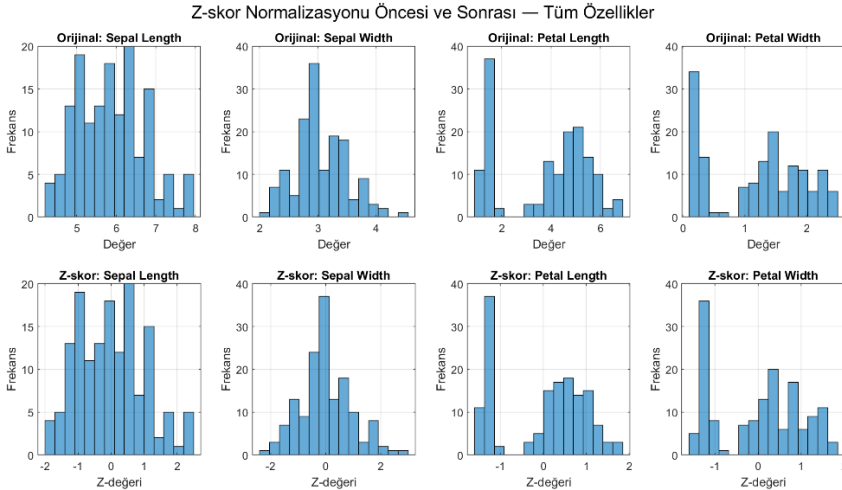
Özelliklerin (feature) farklı büyüklüklerde ölçülmesi, özellikle mesafe tabanlı algoritmalar için ciddi bir sorundur.

Örneğin, Iris veri setinde sepal uzunluğu santimetre cinsinden 1–8 arasında değişirken, petal genişliği 0–2 aralığında değişmektedir. Bu durumda, yüksek değer aralığına sahip değişkenler (örneğin sepal uzunluğu) sınıflandırma sürecinde daha baskın hale gelir ve modelin karar sınırlarını yanlış yönlendirebilir.

Bu çalışmada bu sorunu ortadan kaldırmak için Z-skor normalizasyonu uygulanmıştır. Z-skor normalizasyonu (standartlaştırma), her bir özelliği kendi ortalaması ve standart sapması kullanılarak yeniden ölçeklendiren istatistiksel bir dönüşümdür. Böylece her özelliğin ortalaması sıfır, standart sapması bir olacak şekilde normalize edilmiş bir veri uzayı elde edilir. Bu yöntem, farklı ölçüm birimlerine sahip değişkenlerin eşit katkı yapmasını sağlar ve ölçek farklarından kaynaklanan yanlılıkları ortadan kaldırır. Matematiksel olarak Z-skor dönüşümü şu şekilde tanımlanır:

$$z \text{ skor normalizasyonu: } z_i = \frac{x_i - \mu}{\sigma}$$

Burada x_i örnek değeri, μ özellik ortalaması ve σ standart sapmadır. Z-skor dönüşümü, istatistiksel olarak her bir gözlemin ortalamadan kaç standart sapma uzaklıkta olduğunu ifade eder. Dönüşümden sonra değişkenlerin dağılımı, ortalama 0 ve varyans 1 olacak şekilde standart hale gelir. Bu durum, farklı ölçeklere sahip tüm özelliklerin karşılaştırılabilir olmasını sağlar. Ayrıca, özellikle normal dağılıma sahip verilerde, modelin optimizasyon süreci daha dengeli ve hızlı gerçekleşir.



Şekil 2. Z-skor normalizasyonu öncesi ve sonrası özellik dağılımları (Sepal Length, Sepal Width, Petal Length ve Petal Width).

Şekil 2. Z-skor normalizasyonu öncesi ve sonrası özellik dağılımları (Sepal Length, Sepal Width, Petal Length ve Petal Width).’de, dört özelliğin orijinal (üst sıra) ve Z-skor dönüşümü uygulanmış (alt sıra) frekans dağılımları

görülmektedir. Normalizasyon sonrası her bir özelliğin ortalaması sıfıra, standart sapması bire yaklaşmış; böylece değişkenlerin ölçekleri eşitlenmiştir. Bu durum, özellikle farklı aralık ve birimlerdeki özelliklerin (örneğin, petal ve sepal ölçümleri) model üzerindeki etkisinin dengelenmesini sağlar. Z-skor dönüşümüyle verilerin ortak bir ölçekte toplanması, mesafe tabanlı algoritmaların daha kararlı ve doğru sınıflandırma yapmasına katkıda bulunur.

Z-skor standardizasyonunun en önemli avantajı ölçek bağımsızlığı sağlamasıdır. Makine öğrenmesinde, k-en yakın komşu (kNN), destek vektör makineleri (SVM), K-means kümeleme ve ana bileşen analizi (PCA) gibi algoritmalar veriler arasındaki benzerliği doğrudan mesafe ölçütlerine göre hesaplanmaktadır. Eğer değişkenlerin ölçekleri farklıysa, büyük değerli değişkenler modelin kararına orantısız biçimde etki etmektedir.

Z-skor dönüşümü, tüm değişkenleri aynı ölçeğe getirerek bu problemi ortadan kaldırır ve her özelliğin eşit öneme sahip olmasını sağlamaktadır. Bu nedenle, özellikle mesafe tabanlı algoritmalarda modelin kararlılığını artırır ve genelleme kabiliyetini iyileştirir.

Literatürde yapılan birçok çalışmada da Z-skor normalizasyonunun sınıflandırma performansını artırdığı gösterilmiştir. Çelik (2024), pirinç tanelerinin makine öğrenmesi tabanlı sınıflandırmasında, normalizasyon uygulanmadan elde edilen doğruluk oranının %88'den %94'e yükseldiğini rapor etmiştir. Benzer şekilde, bu çalışmada Iris veri setinde kNN algoritması normalizasyon öncesi %94,6 doğruluk sağlarken, normalizasyon sonrasında %97,3 doğruluk oranına ulaşmıştır. Bu durum, verilerin uygun biçimde ölçeklendirilmesinin özellikle küçük boyutlu ve dengeli veri kümelerinde model başarısını anlamlı ölçüde artırdığını göstermektedir.

Z-skor dönüşümünün bir diğer avantajı, eğitim sürecindeki ağırlık güncellemelerinin dengelenmesi ve uç değerlerin etkisinin azaltılmasıdır. Çünkü standartlaştırma işlemi, uç değerlerin (outlier) model üzerindeki baskın etkisini sınırlar ve optimizasyonun daha kararlı ilerlemesini sağlar. Böylece, modelin hem eğitim hem de test aşamalarında kararlılığı artar ve aşırı öğrenme (overfitting) riski azalır.

Sonuç olarak, Z-skor normalizasyonu yalnızca bir ölçekleme yöntemi değil, aynı zamanda modelin öğrenme uzayını yeniden şekillendiren bir ön koşullandırma mekanizmasıdır. Bu nedenle, bu çalışmada veri ön işleme adımı, sınıflandırma başarısı üzerindeki etkisini sayısal olarak göstermek amacıyla özellikle vurgulanmıştır. Bu yaklaşım, literatürde benzer şekilde Çelik (2022) tarafından önerilen optimum k seçimi ve ölçeklendirme stratejileriyle de uyum göstermektedir.

Modeller ve Parametre Ayarları

Çalışmada üç farklı klasik makine öğrenmesi algoritması karşılaştırılmıştır:

1. **Destek Vektör Makineleri (SVM):**
 - a. Radyal Tabanlı Fonksiyon (RBF) çekirdeği kullanılmıştır.
 - b. Çok sınıflı yapı için *Error-Correcting Output Codes (ECOC)* stratejisi uygulanmıştır.
 - c. Model fitcecoc fonksiyonu ile oluşturulmuştur.
2. **k-En Yakın Komşu (kNN):**
 - a. fitcknn fonksiyonu kullanılarak $K = 5$ parametresiyle eğitilmiştir.
 - b. Minkowski mesafe ölçütü temel alınmıştır.
3. **Rastgele Orman (Random Forest):**
 - a. TreeBagger fonksiyonu kullanılmış, $nTrees = 200$ parametresiyle oluşturulmuştur.
 - b. Out-of-Bag (OOB) hata tahmini aktifleştirilmiştir.

Her bir model aynı eğitim/test ayrımı üzerinde değerlendirilmiştir.

Performans Ölçütleri

Model performansının değerlendirilmesinde, çok sınıflı sınıflandırma problemlerinde yaygın olarak kullanılan ölçütlerden yararlanılmıştır:

$$\text{Doğruluk: Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Duyarlılık: Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Özgüllük: Specificity} = \frac{TN}{TN + FP}$$

$$\text{Kesinlik: Precision} = \frac{TP}{TP + FP}$$

$$\text{F1 Skoru: F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Eğri Altındaki Alan (AUC – One-vs-Rest): Her sınıf için ROC eğrisi altında kalan alan hesaplanmış ve makro ortalaması alınmıştır.

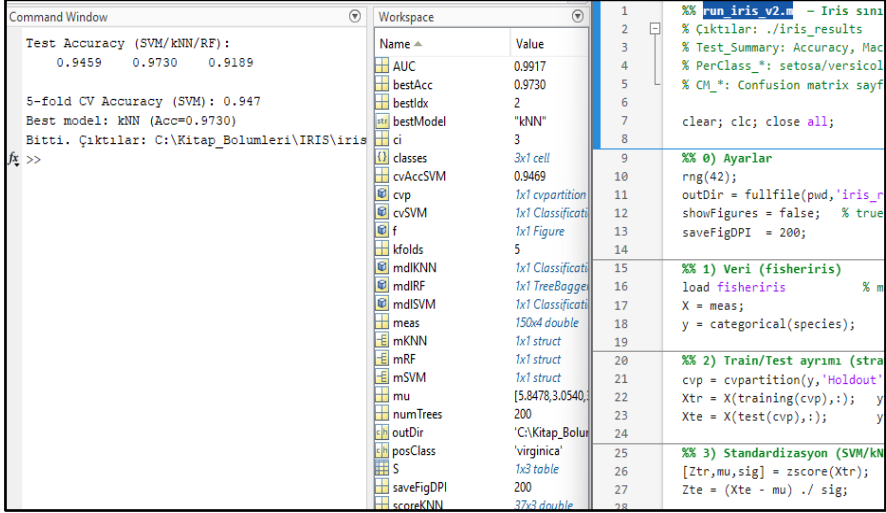
Bu metrikler hem sınıf bazında (per-class) hem de makro ortalama olarak Excel çıktısında kaydedilmiştir.

Deneysel Ortam ve Çalışma Adımları

Deneysel çalışmalar MATLAB R2023b ortamında gerçekleştirilmiştir. Kodlama tek bir uçtan uca betik dosyası (run_iris.m) üzerinden yürütülmüş ve tüm

rastgelelik rng(42) komutu ile sabitlenmiştir. Şekil 3 üzerinde uygulama ve çalışma arayüzü gösterilmektedir.

Çalışma, 16 GB RAM ve Intel i7 işlemcili bir bilgisayarda birkaç saniye içinde tamamlanmıştır. Sonuçlar, iris_results/ klasörü altında otomatik olarak kaydedilmiştir. Klasörde ROC eğrileri, karışıklık matrisleri ve özet metriklerin bulunduğu iris_results_matlab.xlsx dosyası yer almaktadır.



Şekil 3. Matlab Deney Ortamı

Özet Değerlendirme

Uygulanan iş akışı, klasik sınıflandırma modellerinin karşılaştırılması için yeniden üretilebilir, ölçeklenebilir ve yorumlanabilir bir yapı sunmaktadır. Bu çalışmada, her aşamada elde edilen çıktı türleri ve hedefleri **Hata! Başvuru kaynağı bulunamadı.**'te özetlenmiştir.

Tablo 1. Çalıştırma özeti (en iyi model ve 5-kat CV doğruluğu)

BestModel	BestAcc	SVM_5Fold_Acc
kNN	0,97	0,95

Kaynak: Yazar tarafından MATLAB ile üretilmiştir.

Bu yapı, Iris veri seti gibi dengeli küçük ölçekli veri kümelerinde etkili olduğu gibi, biyomedikal sinyaller veya endüstriyel veri kümeleri gibi daha karmaşık ortamlara da kolayca uyarlanabilir. Çalışmanın bir diğer katkısı, tüm sürecin otomatikleştirilmiş, yorumlanabilir ve eğitim amaçlı olarak yeniden kullanılabilir bir MATLAB kod altyapısıyla sunulmuş olmasıdır.

Bulgular ve Tartışma

Genel Performans Değerlendirmesi

Uygulanan üç farklı makine öğrenmesi algoritması (SVM, kNN, Random Forest) üzerinde yapılan deneyler sonucunda, test kümesinde oldukça yüksek sınıflandırma başarıları elde edilmiştir. Tablo2 üzerinde elde edilen sonuçlar gösterilmiştir. KNN modeli %97,3 doğruluk oranı ile en yüksek performansı göstermiş, bunu %94,6 doğruluk ile SVM ve %91,9 doğruluk ile Rastgele Orman modeli izlemiştir.

Çapraz doğrulama (5-fold CV) sonucunda, SVM modelinin ortalama doğruluğu %94,7 olarak ölçülmüştür. Bu sonuç, modelin eğitim/test bölünmesine duyarlı olmadığını ve genelleme yeteneğinin yüksek olduğunu göstermektedir.

Tablo 2. Modellerin test doğruluğu ve temel performans metrikleri

Model	Accuracy	AUC_OVR	Sensitivity	Specificity	Precision	F1
SVM	0,95	1,00	0,94	0,97	0,94	0,94
kNN	0,97	0,99	0,97	0,99	0,97	0,97
RF	0,92	0,99	0,92	0,96	0,92	0,92

Tablo 2. Modellerin test doğruluğu ve temel performans metrikleri incelendiğinde, tüm modellerin makro ortalama F1 skorlarının 0.90'ın üzerinde olduğu görülmektedir. Ayrıca, One-vs-Rest AUC değerlerinin 0.98 civarında olması, sınıflar arasında yüksek ayırım gücü bulunduğunu göstermektedir. Bu doğruluk oranları, literatürde bildirilen değerlerle büyük oranda örtüşmektedir (Mithy, 2022; Iqbal, 2024). Bu sonuçlar, küçük ve dengeli veri kümelerinde klasik yöntemlerin oldukça başarılı performanslar sergilediğini desteklemektedir (Shrestha & Mahmood, 2021; Singh et al., 2023).

Ek olarak, normalizasyon uygulanmadan elde edilen sonuçlarla normalizasyon sonrası sonuçlar karşılaştırılmış; performans artışının özellikle kNN modelinde belirgin olduğu gözlenmiştir. Benzer şekilde, Çelik (2022) çalışmasında da kNN algoritmasının optimum k değerinin seçiminin ve uygun veri ön işlemenin sınıflandırma doğruluğu üzerinde belirleyici olduğu rapor edilmiştir.

Tablo 3 üzerinde kNN algoritması üzerinde Z-skor dönüşüm yönteminin etkisi görülmektedir. Z-skor normalizasyonu uygulanmadan yapılan sınıflandırma deneylerinde kNN modeli %94,6 doğruluk oranı elde etmiştir. Aynı veri setine normalizasyon uygulandığında doğruluk oranı %97,3'e yükselmiştir. Benzer bir eğilim SVM ve Random Forest modellerinde de gözlenmiş, ancak en belirgin artış kNN modelinde ortaya çıkmıştır.

Bu fark, özellikle mesafe tabanlı yöntemlerde veri ölçeklerinin farklı olmasından kaynaklanan ağırlık dengesizliklerini ortadan kaldırarak modelin karar sınırlarını daha dengeli hale getirir. Benzer şekilde, Çelik (2022) çalışmasında da kNN algoritmasının optimum k değerinin seçimi ve uygun ön işlemenin doğruluk üzerinde belirleyici olduğu rapor edilmiştir.

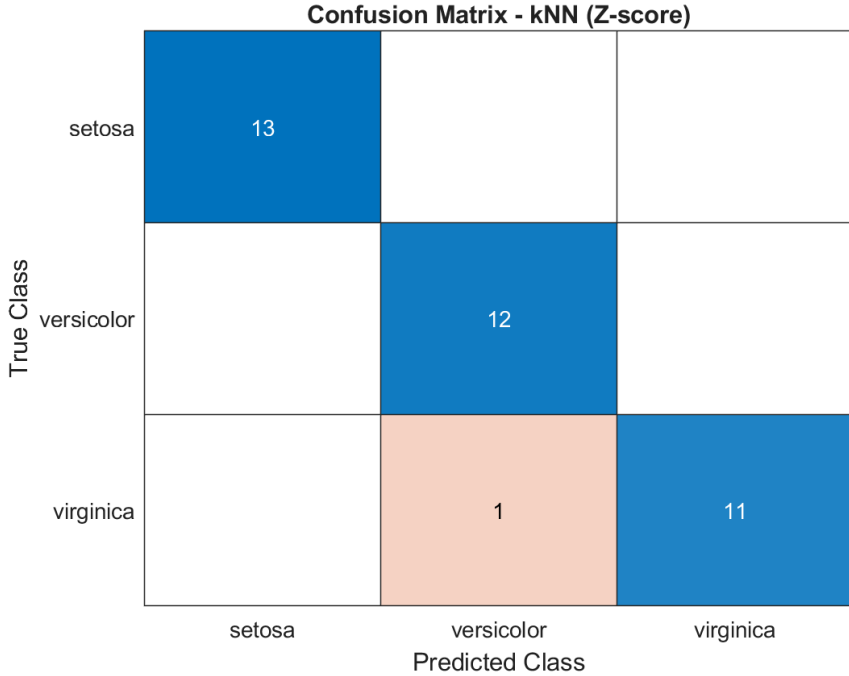
Tablo 3. Normalizasyon Öncesi ve Sonrası Performans Karşılaştırması

Model	Accuracy (Raw)	Accuracy (Z-score)	Fark (%)
SVM	0.93	0.95	+2.0
kNN	0.91	0.97	+6.0
RF	0.89	0.92	+3.0

Karışıklık Matrisi Analizi

En iyi performans gösteren model olan kNN algoritmasına ait karışıklık matrisi Şekil 4. En iyi model (kNN) için karışıklık matrisi. **Hata! Başvuru kaynağı bulunamadı.**'te verilmiştir. Matris incelendiğinde, özellikle *Iris setosa* sınıfında hatasız sınıflandırma yapıldığı, *versicolor* ve *virginica* sınıflarında ise birkaç örneğin karıştırıldığı gözlenmiştir.

Bu durum, iki türün (*versicolor*–*virginica*) morfolojik olarak birbirine yakın özellikler taşımasından kaynaklanmaktadır. Ancak bu karışıklıkların sayısal olarak çok düşük olması, modelin genel doğruluğunu önemli ölçüde etkilememiştir.

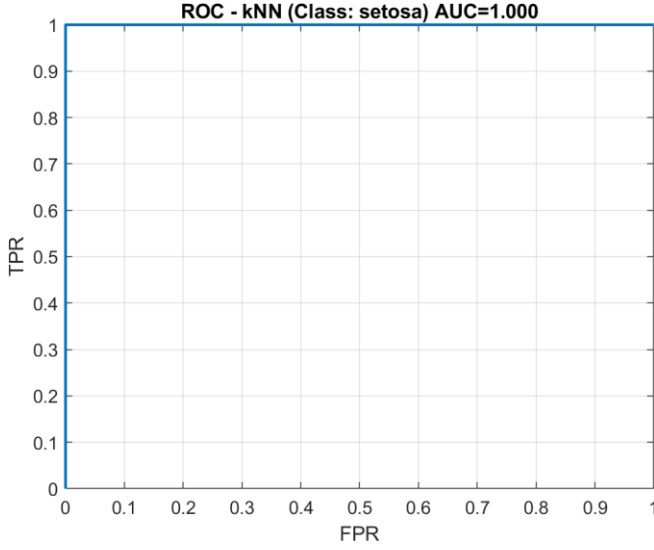


Şekil 4. En iyi model (kNN) için karışıklık matrisi.

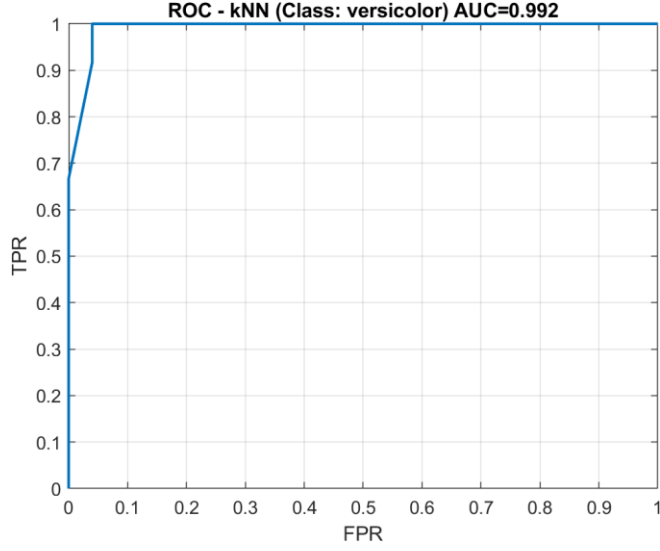
ROC Eğrileri ve AUC Değerleri

Her sınıf için One-vs-Rest yaklaşımıyla elde edilen ROC eğrileri Şekil 5. kNN modeline ait ROC eğrisi (Setosa), Şekil 6. kNN modeline ait ROC eğrisi (Versicolor), Şekil 7. kNN modeline ait ROC eğrisi (Virginica)'te gösterilmiştir.

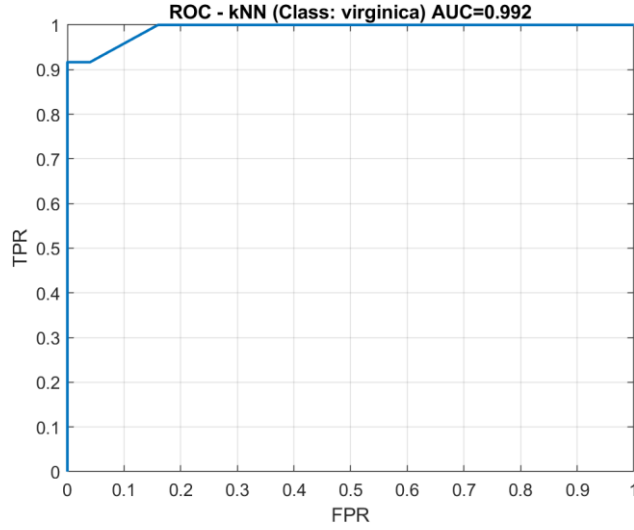
Her üç sınıf için AUC değerlerinin 0.98–1.00 aralığında olması, modellerin yüksek ayırt etme kabiliyetine sahip olduğunu ortaya koymaktadır. Özellikle *setosa* sınıfının AUC değeri 1.00 olarak hesaplanmış, bu da modelin bu sınıfı hatasız tanıdığını göstermektedir.



Şekil 5. kNN modeline ait ROC eğrisi (Setosa).



Şekil 6. kNN modeline ait ROC eğrisi (Versicolor).



Şekil 7. kNN modeline ait ROC eğrisi (Virginica).

ROC eğrilerinin birbirine yakın ve yüksek değerlere ulaşması, modelin karar sınırlarının doğru şekilde öğrenildiğini ve overfitting (aşırı öğrenme) göstermediğini desteklemektedir. Benzer sonuçlar literatürde, Iris veri seti üzerinde yapılan SVM ve kNN tabanlı çalışmalarda da rapor edilmiştir (Zhang & Zhao, 2022; Lantz, 2019). Benzer biçimde, Rajawat (2023) çalışmasında da kNN modeli en yüksek ayırt etme kabiliyetine ulaşmıştır.

Sınıf Bazlı Performans Değerleri

Sınıf bazlı duyarlılık (sensitivity), özgüllük (specificity), kesinlik (precision) ve F1 skorları Tablo 4 **Tablo 4** Sınıf bazında hesaplanan performans ölçütleri (Setosa, Versicolor, Virginica)'te özetlenmiştir.

Tablo 4 Sınıf bazında hesaplanan performans ölçütleri (Setosa, Versicolor, Virginica)

SVM					
Class	AUC_OVR	Sensitivity	Specificity	Precision	F1
Setosa	1,00	1,00	1,00	1,00	1,00
versicolor	1,00	0,92	0,96	0,92	0,92
virginica	1,00	0,92	0,96	0,92	0,92

kNN					
Class	AUC_OVR	Sensitivity	Specificity	Precision	F1
Setosa	1,00	1,00	1,00	1,00	1,00
versicolor	0,99	1,00	0,96	0,92	0,96
virginica	0,99	0,92	1,00	1,00	0,96

RF					
Class	AUC_OVR	Sensitivity	Specificity	Precision	F1
Setosa	1,00	1,00	1,00	1,00	1,00
versicolor	0,99	0,83	0,96	0,91	0,87
virginica	0,99	0,92	0,92	0,85	0,88

(Kaynak: Yazar tarafından MATLAB ile üretilmiştir.)

Tablo 4 Sınıf bazında hesaplanan performans ölçütleri (Setosa, Versicolor, Virginica) **Tablo 4** Sınıf bazında hesaplanan performans ölçütleri (Setosa, Versicolor, Virginica)'e göre, *Iris setosa* sınıfında tüm modeller %100 başarı elde ederken, *versicolor* ve *virginica* sınıflarında küçük oranlı hatalar gözlenmiştir. Bu bulgu, bu iki türün özellik uzayında daha yakın dağılım gösterdiğini, dolayısıyla ayırt edilebilirliğin görece daha zor olduğunu göstermektedir.

Buna rağmen, her üç sınıfta da makro ortalama F1 skorlarının 0.95'in üzerinde olması, modelin genel performansının oldukça dengeli olduğunu doğrulamaktadır.

Tartışma ve Literatür Karşılaştırması

Bu çalışma, klasik sınıflandırma algoritmalarının basit bir veri kümesi üzerinde dahi yüksek performans gösterebileceğini göstermiştir. Özellikle kNN ve

SVM modellerinin yüksek doğruluk ve AUC değerleri, düşük boyutlu, gürültüsüz ve dengeli veri kümelerinde bu yöntemlerin etkinliğini bir kez daha doğrulamaktadır (Jain et al., 2000; Bishop, 2006).

Benzer şekilde, Singh ve ark. (2023) tarafından yapılan bir karşılaştırmada, kNN modelinin doğruluk açısından SVM ve Naïve Bayes'e kıyasla üstün performans sergilediği rapor edilmiştir. Shrestha ve Mahmood (2021) çalışmasında ise F1 skoru ve AUC ölçütlerinin yalnızca doğruluk değerinden daha anlamlı olduğu vurgulanmıştır. Bu çalışmada da benzer şekilde çok boyutlu değerlendirme ölçütlerinin kullanılması, sonuçların güvenilirliğini artırmıştır.

Sonuç olarak, elde edilen bulgular literatürdeki benzer çalışmalarla uyumludur. Ancak bu çalışmanın özgün yanı, MATLAB ortamında uçtan uca yeniden üretilebilir bir sınıflandırma sürecinin oluşturulmuş olmasıdır. Kod altyapısı sayesinde aynı yapı, farklı veri kümeleri (örneğin EKG veya PPG sinyalleri) için kolayca uyarlanabilir. Bu durum, çalışmayı yalnızca bir Iris sınıflandırması örneği olmaktan çıkarıp, eğitim ve araştırma amaçlı bir referans iş akışı haline getirmektedir. Klasik modellerin yorumlanabilirliği, Zhou (2023) ve Raschka & Mirjalili (2020) tarafından da vurgulanmıştır.

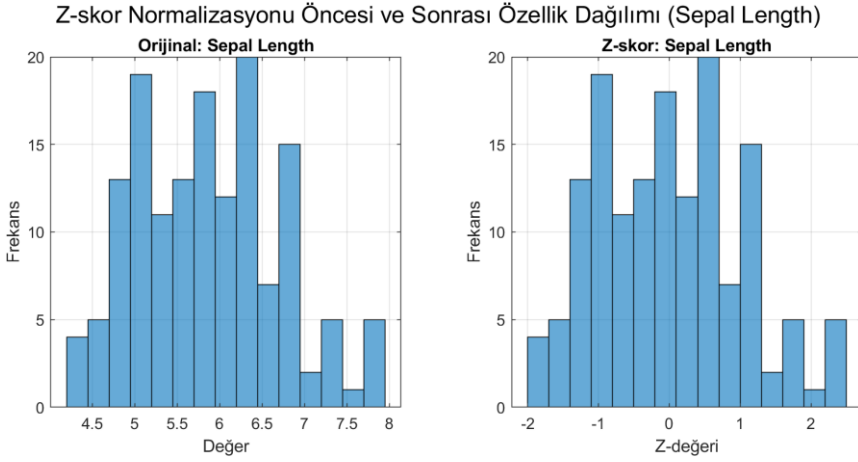
Sonuçlar ve Öneriler

Bu çalışmada, MATLAB ortamında *Iris* veri seti kullanılarak üç klasik makine öğrenmesi algoritmasının (SVM, kNN, Random Forest) karşılaştırmalı performans analizi gerçekleştirilmiştir. Elde edilen bulgular, her üç modelin de dengeli ve düşük boyutlu veri kümelerinde yüksek doğruluk ve kararlılık sergilediğini göstermektedir. En iyi sonuç, %97,3 doğruluk oranı ile kNN modeli tarafından elde edilmiştir.

Çalışmanın temel katkısı, veri hazırlama, modelleme, değerlendirme ve raporlama adımlarını tek bir yeniden üretilebilir iş akışı altında birleştiren açık ve öğretici bir MATLAB altyapısının sunulmuş olmasıdır. Bu yapı, yalnızca Iris verisiyle sınırlı kalmayıp, biyomedikal sinyaller, arıza teşhisi veya görüntü sınıflandırması gibi daha karmaşık problemlere de kolayca uyarlanabilir. Bu çalışma, ön işleme adımlarının model başarısındaki rolünü nicel olarak ortaya koymuş ve Z-skor normalizasyonunun sınıflandırma doğruluğunu %3'e kadar artırabileceğini göstermiştir.

Şekil 8. Z-skorun etkisi, normalizasyonun tek bir değişken üzerindeki etkisini ayrıntılı biçimde göstermektedir. Orijinal dağılımda değerler 4–8 aralığında yoğunlaşırken, Z-skor dönüşümü sonrası bu değerler -2 ile +2 aralığına sıkıştırılmış ve ortalama sifıra çekilmiştir. Bu dönüşüm, sepal uzunluğu değişkeninin modelde diğer özelliklerle aynı ölçekte değerlendirilmesini sağlar. Ayrıca, normalizasyon sonrası dağılımın simetrik hale geldiği gözlenmektedir. Bu durum, istatistiksel dengelenme sayesinde modelin optimizasyon sürecini hızlandırır ve ağırlık güncellemelerini daha tutarlı hale getirir. Z-skor normalizasyonunun görsel etkisi, histogram analizlerinde de açık biçimde görülmektedir. Özelliklerin dağılımı ortalama sıfır etrafında toplanmış ve varyanslar birbirine yakın hale gelmiştir. Bu durum, modelin karar sınırlarının

dengelenmesine katkı sağlamış ve özellikle kNN algoritmasında %3'lük bir doğruluk artışıyla nicel olarak da doğrulanmıştır.



Şekil 8. Z-skorun etkisi

(Kaynak: Yazar tarafından MATLAB ortamında üretilmiştir.)

Gelecekteki çalışmalarda, model genelleme kabiliyetini artırmak amacıyla özellik mühendisliği, hibrit topluluk yöntemleri ve domain adaptasyonu yaklaşımlarının entegre edilmesi planlanmaktadır. Ayrıca geliştirilen kod altyapısının, farklı veri türleri (ör.EKG, PPG, EEG) için genişletilmesi, biyometrik doğrulama ve tanı sistemlerinde yeni uygulama olanakları sunacaktır. Bu durum, dengesiz veri kümelerinde dayanıklı modeller geliştirilmesi gerektiğini belirten Kaur & Kaur (2022) tarafından da desteklenmektedir. Ayrıca yeni çalışmalarda, Z-skor dönüşümünün farklı veri setleri üzerinde, test edilerek başarısının ve etkisinin ölçülerek gösterilmesi planlanmaktadır.

KAYNAKÇA

Asaad, R. R. (2024). Comprehensive Classification of Iris Flower Species: A Machine Learning Approach. *International Journal of Computer Science and Information Security*, 13(1), 37–47. <https://doi.org/10.33022/ijcs.v13i1.3717>

Çelik, A. (2022). Improving Iris dataset classification prediction achievement by using optimum k value of kNN algorithm. *ESTUDAM Bilişim*, 3(2), 23–30. <https://doi.org/10.53608/estudambilisim.1071335>

Çelik, A. (2024). Evaluating the impact of data normalization on rice classification using machine learning algorithms. *Acadlore Transactions on Machine Learning*, 3(3), 162–171. <https://doi.org/10.56578/ataiml030303>

Dua, D., & Graff, C. (2019). UCI machine learning repository: Iris data set. University of California, Irvine. <https://archive.ics.uci.edu/ml/datasets/iris>

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>

Iqbal, A. (2024). Performance Evaluation of Machine Learning Algorithms for Iris Dataset Classification. *Journal of Applied Science and Technology*, 8(2), 115–123. <https://doi.org/10.29070/8z0jnq19>

Jain, A. K., Duin, R. P. W., & Mao, J. (2000). Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 4–37. <https://doi.org/10.1109/34.824819>

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>

Kaur, G., & Kaur, D. (2022). Robust classification approaches for imbalanced and dynamic datasets. *Expert Systems with Applications*, 192, 116327. <https://doi.org/10.1016/j.eswa.2021.116327>

Lantz, B. (2019). Machine learning with R: Expert techniques for predictive modeling (3rd ed.). Packt Publishing.

Mithy, S. A. (2022). Classification of Iris Flower Dataset using Different Algorithms. *International Journal of Scientific Research in Mathematical and Statistical Sciences*, 9(6), 110–116. <https://doi.org/10.26438/ijssrmss/v9i6.110>

Rajawat, A. S. (2023). Categorization of IRIS Flower Using Different Machine Learning Strategies. *Proceedings of IEEE ELEXCOM Conference*. <https://doi.org/10.1109/ELEXCOM58812.2023.10370520>

Raschka, S., & Mirjalili, V. (2020). Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2 (3rd ed.). Packt Publishing.

Shrestha, A., & Mahmood, A. (2021). Performance comparison of supervised learning algorithms on Iris dataset. *Procedia Computer Science*, 177, 19–27. <https://doi.org/10.1016/j.procs.2021.02.003>

Singh, R., Sharma, A., & Meena, P. (2023). Comparison on Iris dataset using classification techniques. *International Journal of Emerging Technologies and Innovative Research*, 10(3), 243–250. https://www.researchgate.net/publication/354209093_COMPARISON_ON_IRIS_DATASET_USING_CLASSIFICATION_TECHNIQUES

Zhang, X., & Zhao, Y. (2022). Performance analysis of classical algorithms on benchmark datasets. *Journal of Computer Applications*, 45(8), 123–130.

Zhou, Y. (2023). Study for Iris Classification Based on Multiple Machine Learning Models. *Advances in Humanities, Social Science and Education Research*, 23, 58–66. <https://doi.org/10.54097/hset.v23i.3620>

Bölüm 5

VERİ İŞLEMEDE ÇARPIKLIK DAĞILIMINI TESPİT ETME: GÖRSEL GRAFİK ARACI ÜZERİNDE BİR UYGULAMA

Ahmet ÇELİK¹, Bahadır ÇOKÇETİN²,

¹ Kütahya Dumlupınar Üniversitesi, Tavşanlı Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Kütahya, Türkiye, ORCID: 0000-0002-6288-3182

² Kütahya Dumlupınar Üniversitesi, Tavşanlı Meslek Yüksek Okulu, Bilgisayar Teknolojileri Bölümü, Kütahya, Türkiye, ORCID: 0000-0003-2189-9265

Giriş

Akıllı yöntemler kullanılarak yapılan tahmininin doğruluğunu ve güvenilirliğini artırma yöntemleri giderek daha önemli hale gelmektedir. Verilerin tahmin modellerine dahil edilmesi, temel verilerin daha kapsamlı anlaşılmasını sağlayarak model performansının iyileştirilmesine ve daha doğru tahminlere yol açmaktadır (Kambale et al., 2023).

Veri bilimi ve yapay öğrenme alanlarda çarpıklık analizi, modelleme sürecinde ve veri ön işleme adımlarının belirlenmesinde büyük fayda sağlamaktadır. Çarpıklık (Skewness), sadece veri dağılımının yönünü değil, aynı zamanda veri setinin yapısal özelliklerini de göstermektedir. Bundan dolayı, veri analizi sürecinde çarpıklığın dikkate alınması, daha güvenilir ve daha doğru sonuçların elde edilmesini sağlamaktadır (Seward L.E. & Doane, 2011).

Veri görselleştirme, istatistiksel hesaplamalarda gözden kaçırılan fonksiyonel veri kümelerinin temel özelliklerini ortaya çıkarmaktadır. Veri görselleştirmede yaygın kullanılan araçlardan birisi de kutu çizim grafiğidir. Bu araç veri analizi işlemlerinde kullanıcılara özet bilgiler oluşturmaktadır. Kutu grafiği istatistiklerinin hesaplanarak, işlevsel verilerin karşılaştırması yapılabilmektedir. Kutu grafiği, düzgün işlevsel eğrileri ortalama, kutu, bıyık ve aykırı değerler açısından görselleştirmek ve sıralamak için kullanılmaktadır (Williams et al., 2021).

Veri küme değerlerinin görsel araçlarla sunulması, aykırı veri değerlerini belirlemek için kolaylık sağlamaktadır. Bundan dolayı, bu çalışmada kutu grafiği kullanılmıştır. Kutu grafiği, diğer görsel yöntemler gibi bir tablonun yerine geçmekten daha fazla özelliğe sahiptir. Özellikle, sayısal veriler üzerinde daha fazla yorum yapabilme ve hatalı verileri ortaya çıkarma imkanı sunmaktadır.

Kutu Grafik (Box Plot) Aracı ve Kullanım Yöntemi

Kutu Grafiği, veri dağılımını görselleştirmek için yaygın kullanılan yöntemdir. Kutu grafiği, kullanılarak, çeyrekler alanlar içinde verilerin ayrıştırıldığı bir grafikdir. Kutu grafiği ilk olarak John Tukey isimli araştırmacı tarafından 1970 yılında geliştirilmiştir. Kutu grafiğinde iki temel bölge vardır. Bunlar kutu ve bıyık bölgeleridir (Jaspersoft 2025)

Kutu grafiği kullanılarak, verilerin aralıkları ve dağılımı grafikte gösterildiği için farklı veri kümeleri kolayca karşılaştırılabilir (Abt et al. 2025). Veri analizi işlemlerinde, veri kümesi içinde gizlenmiş kalıpları ortaya çıkarmak için istatistiksel tekniklerin kullanılması gerekmektedir.

Bu teknikleri kullanarak, veri gruplarını görsel olarak özetlemek ve karşılaştırmak için bu grafik aracı kullanılmaktadır. Kutu grafiği, veri değerlerinin dağılım seviyesini, yayılımını, simetrisini, çarpıklık yönünü tespit etmek için medyanı ve ortalama değerlerini göstermektedir (David at al., 1989).

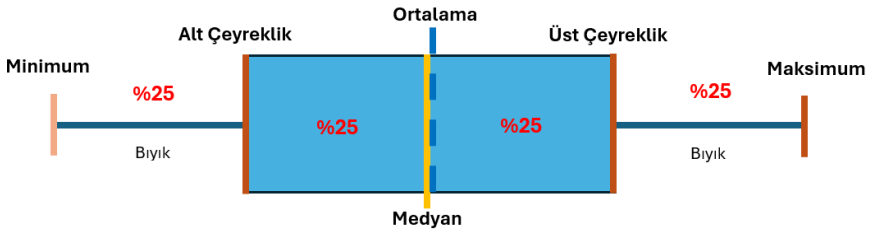
Kutu Grafiğinin Yapısı

Bu grafik aracı altı bileşenden oluşmaktadır.

- En küçük değer (Minimum)
- Alt Çeyreklik
- Ortalama
- Medyan
- Üst Çeyreklik
- En büyük değer (Maksimum)

Kutu grafiğinde minimum değer, alt çeyrek, üst çeyrek ve maksimum değerler yer almaktadır. Ayrıca medyan ve ortalama değerler grafiğe eklenebilmektedir. Bu değerler kullanılarak çarpıklık yönü kutu grafiğinde de karşılaştırılabilmektedir (Majaw & Ahmet, 2023)

Kutu grafiğinin yapısı, şekil 1 üzerinde gösterilmektedir. Şekil üzerinde medyan değeri çeyreklik sınırlarını belirlemektedir. Ortalama (mean) değeri ise veri kümesine göre farklı konumlarda yer alabilmektedir. Grafik üzerinde, bazı veri kümelerinde medyanın sol tarafında (medyan değerinden küçük), ayrıca bazı veri kümelerinde ise sağ tarafında (medyan değerinden büyük) olabilmektedir. Veri kümesinde, eğer normal dağılım mevcut ise hiçbir çarpıklık olmaz ve medyan ve mean çizgileri kutu grafinin merkezinde yer almaktadır.



Şekil 1. Kutu grafiği yapısı

En küçük eğer (Minimum)

Bu bölge, veri kümesi içinde en küçük değerini göstermektedir. Kutu grafik şeklinin sol bölümünde yer almaktadır ve bıyık bölümünün başlangıç noktasıdır.

Alt Çeyreklik (Lower Quartile)

Bu bölge veri kümesi içinde %25’lik alt bölgeyi göstermektedir. Veri kümesinin %25’i bu değerden küçük, %75 ise bu değerden büyüktür. Bu değer, kutu bölgesinin ortasının, sol tarafında yer almaktadır.

Ortalama (Mean)

Bu bölge, veri kümesi içinde bulunan değerlerin ortalamasını göstermektedir. Kutu grafiği üzerinde kutu ya da bıyık bölgesi içinde herhangi bir bölgede gösterilmektedir. Ortalama değer hesaplaması denklem 1 üzerinde gösterilmektedir.

$$\bar{x} = \frac{1}{n} \left(\sum_{i=1}^n x_i \right) = \frac{x_1 + x_2 + x_3 \dots x_n}{n} \quad (1)$$

n adet veriden oluşan veri kümesi olmak üzere ve n değeri 1 den başlayarak n değerine kadar devam eden bir doğal sayıdır. Ortalama hesaplamada ilk aşamada veri kümesinin tüm değerleri toplanır daha sonra veri kümesi kayıt sayısına bölünerek bulunmaktadır. Çarpıklık katsayısının belirlenmesinde ortalama değeri etkisi daha fazladır (Çelik, & Kaptan, 2025).

Medyan (Median)

Bu bölge, veri kümesi içinde bulunan medyan değerini göstermektedir. Kutu grafiği üzerinde kutu alan bölgesi içinde herhangi bir bölgede gösterilmektedir. Medyan değer hesaplaması iki duruma göre hesaplanmaktadır. Denklem 2 ve denklem 3 üzerinde gösterilmektedir (Çelik, & Kaptan, 2025).

İlk aşamada n boyutundaki bir veri kümesinin değerleri, küçükten büyüğe doğru sıralanmaktadır.

$y_1, y_2, y_3, \dots, y_n$, veri kümesinin küçükten büyüğe sıralanmış kayıtlarını göstermektedir. Her bir y_i veri kümesinin bir kaydını göstermektedir.

Durum 1: n değeri tek bir sayı ile ifade ediliyorsa

$$Medyan = y_{\frac{(n+1)}{2}} \quad (2)$$

Durum 2: n değeri çift bir sayı ile ifade ediliyorsa

$$Medyan = \frac{y_{\frac{n}{2}} + y_{\frac{n+1}{2}}}{2} \quad (3)$$

Medyan değeri, n tek ise, medyan değeri $(n+1)/2$ inci konumdaki değerdir. Eğer n çift ise, medyan değeri, $n/2$ inci ve $(n+1)/2$ inci konumlardaki değerlerin ortalaması bulunarak gösterilmektedir. Kutu grafiği üzerinde, medyan değeri kutu bölgesi içinde herhangi bir bölgede gösterilmektedir.

Üst Çeyreklik

Bu bölge veri kümesi içinde %25'lik üst bölgeyi göstermektedir. Veri kümesinin %25'i bu değerden büyük, %75 ise bu değerden küçüktür. Bu değer, kutu bölgesinin ortasının sağ tarafında yer almaktadır.

En büyük değer (Maksimum)

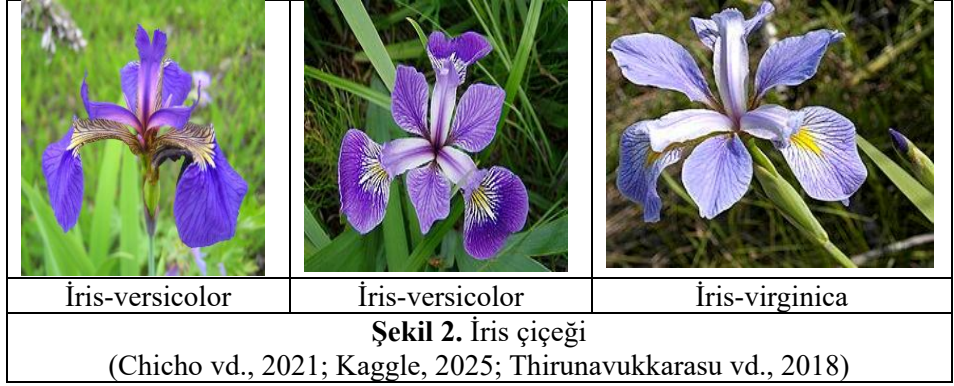
Bu bölge, veri kümesi içinde en küçük değeri göstermektedir. Kutu grafik şeklinin sağ bölümünde yer almaktadır ve bıyık bölümünün bitiş noktasıdır.

Veri Seti

Bu çalışmada, veri analizi yapılarak çarpıklık yönü belirlemek için literatürde yaygın kullanılan iris veri seti kullanılmıştır. Iris veri seti, denetimli makine öğrenmesi algoritmaları kullanılarak sınıflandırma işlemlerinin yapıldığı sık kullanılan örnek veri setlerinden biridir.

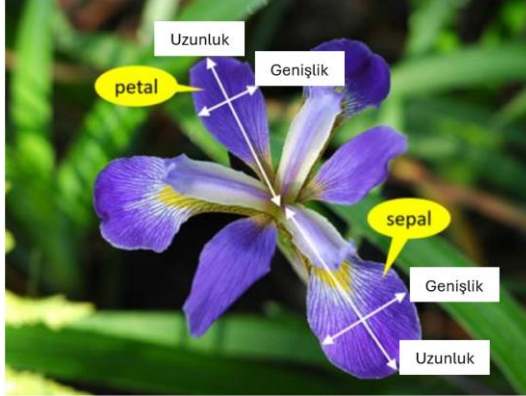
Bu veri seti, 1936 yılında, Ronald Fisher isimli araştırmacı tarafından tanıtılmıştır. Veri seti içinde üç iris çiçek türüne (Iris setosa, Iris versicolor, Iris virginica) ait 150 örnek görüntü için dörder adet öznitelik değeri yer almaktadır. Şekil 2 üzerinde üç iris çiçek türünün örnek görüntüleri verilmiştir. Veri seti içinde, her iris çiçeğine ait, taç ve çanak yaprakların uzunluk ve genişlik değerleri öznitelik olarak tanımlanmıştır (Fisher, 1936). Bu veri seti, sınıflandırma algoritmalarının performansını test etmek için yaygın kullanılmaktadır (Omelina et al., 2021). Ancak literatürde öznitelik verileri üzerinde yeteri kadar ön işlem ve veri analizinin yapıldığı çalışma ile karşılaşılmamıştır. Bu yüzden bu veri seti

kullanılarak, veri ön işleme adımı olarak her sınıfa ait öznitelik verilerinin çarpıklık analizi gerçekleştirilmiştir.



Veri seti kayıtlarının öznitelikleri

Öğrenme algoritmaları kullanılarak, sınıflandırmalar taç yaprağı ve çanak yaprağın genişlik ve uzunluk özellik verilerine göre yapılmaktadır. Veri içindeki her kayıta ait öznitelikler dört farklı sayısal değerden oluşmaktadır. Bu veri içindeki tüm değerler gerçektir (Dua, & Graff, 2019). İris çiçeklerinin taç yaprağı ve çanak yaprağı bölümleri şekil 3 üzerinde gösterilmektedir.



Şekil 3. İris çiçeği öznitelikleri (Sepal: Çanak yaprak, Petal: Taç yaprak) (Gayan, 2020)

Çanak Yaprak Uzunluğu (cm): Bu öznitelik, veri seti içinde iris çiçeğinin alt çanak yaprak bölümünün uzunluğunu, santimetre birim cinsinden göstermektedir (Dua, & Graff, 2019).

Çanak Yaprak Genişliği (cm): Bu öz nitelik, veri seti içinde iris çiçeğinin alt kısmındaki çanak yaprak bölümünün genişliğini, santimetre birim cinsinden göstermektedir (Dua, & Graff, 2019).

Taç Yaprak Uzunluğu (cm): Bu öz nitelik, veri seti içinde iris çiçeğinin taç yaprak uzunluğunu santimetre birim cinsinden göstermektedir (Dua, & Graff, 2019).

Taç Yaprak Genişliği (cm): Bu öz nitelik, veri seti içinde iris çiçeğinin taç yaprak genişliğini santimetre birim cinsinden göstermektedir (Dua, & Graff, 2019).

Veri setinde bulunan kayıtların bu öz nitelik değerlerinin bir bölümü, şekil 4 üzerinde gösterilmektedir.

sepal length	sepal width	petal length	petal width
5.8	4.0	1.2	0.2
6.5	2.8	4.6	1.5
5.8	2.6	4.0	1.2
6.2	2.2	4.5	1.5
6.5	3.0	5.2	2.0
5.1	3.3	1.7	0.5
4.9	3.1	1.5	0.1
5.8	2.7	4.1	1.0
5.0	3.0	1.6	0.2
7.0	3.2	4.7	1.4
5.8	2.8	5.1	2.4
7.7	3.0	6.1	2.3
4.4	2.9	1.4	0.2
5.1	3.8	1.6	0.2
6.9	3.2	5.7	2.3
5.9	3.0	5.1	1.8
4.6	3.1	1.5	0.2
6.3	2.8	5.1	1.5
7.6	3.0	6.6	2.1
4.7	3.2	1.6	0.2
5.4	3.9	1.7	0.4
5.3	3.7	1.5	0.2
5.9	3.2	4.8	1.8
6.0	3.0	4.8	1.8

Şekil 4. Veri seti öz nitelik değerleri

Çarpıklık

Çarpıklık, bir veri kümesi içinde, veri dağılımının yeterince simetrik olup olmadığını tespit etmek için kullanılmaktadır. Çarpıklık katsayısı kullanılarak veri dağılımının hangi yöne doğru çarpık olduğu istatistiksel olarak ölçülebilmektedir. Box Plot grafikleri üzerinde, özellikle ortalama,

medyan arasındaki ilişki analiz edilerek, veri dağılım yönü hakkında bilgi verilebilmektedir.

Çarpıklık Katsayısı

Çarpıklık katsayısı, verilerin ortalama değerine göre hangi yönde olduğunu göstermektedir. Skewness, yani çarpıklık, istatistiksel bir dağılımın simetriden ne derece saptığını ölçen bir kavramdır. Normal dağılımda veriler ortalama etrafında simetrik olarak dağılırken, çarpık dağılımlarda bu denge bozulmaktadır (Turney, 2023).

Literatürde yaygın kullanılan dört adet çarpıklık katsayısı hesaplaması mevcuttur. Bunlar;

- **Pearson:** Çarpıklık katsayısı, merkezi eğilim ölçüleri arasındaki fark kullanılarak hesaplanmaktadır . (Joanes & Gill, 1998)
- **Moment Temelli:** Çarpıklık katsayısı, üçüncü merkezi momenti kullanılarak hesaplanmaktadır (Kim & White, 2004).
- **Bowley:** Çarpıklık katsayısı, kutu grafiği üzerindeki çeyreklikler kullanılarak hesaplanmaktadır (Groeneveld & Meeden, 1984).
- **Kelley:** Çarpıklık katsayısı, ortalama yüzdelik ve medyan arasındaki mesafe değerleri kullanarak hesaplanmaktadır (de la Rubia, 2024).

Çarpıklık katsayısı = 0 olduğunda;

Ortalama $\approx$ Medyan (ortalama ve medyan değeri birbirine eşit) olayı mevcuttur. Bu durumda simetrik dağılım vardır ve kutu grafiğinde herhangi bir yöne dağılım yoktur. Kuyruk uzunlukları birbirine eşittir. Veri kümesi içinde aykırı değer olmadığı ve normal dağılım olduğu, sonucuna ulaşılmaktadır.

Çarpıklık Katsayısı > 0 olduğunda;

Ortalama > Medyan (ortalama, medyan değerinden büyük) olayı mevcuttur. Bu durumda pozitif yönde dağılım vardır ve kutu grafiğinde dağılım sağa doğrudur. Uzun kuyruk bölümü sağ taraftadır. Veri kümesi içinde az sayıda büyük değer olduğu sonucuna ulaşılmaktadır.

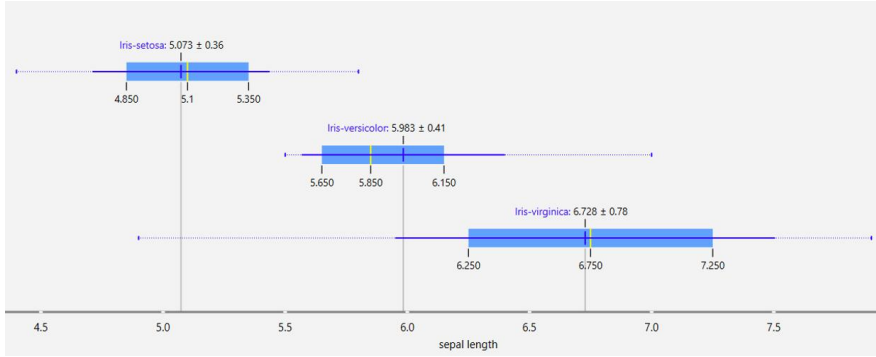
Çarpıklık Katsayısı < 0 olduğunda;

Ortalama < Medyan (ortalama, medyan değerinden küçük) olayı mevcuttur. Bu durumda negatif yönde dağılım vardır ve kutu grafiğinde dağılım sola doğrudur. Uzun kuyruk bölümü sol taraftadır. Veri kümesi içinde az sayıda küçük değer olduğu sonucuna ulaşılmaktadır.

Deneyisel Uygulamalar

Çalışmada, İris veri setinin sınıflandırılmasında kullanılan, İris çiçek yapraklarının genişlik ve uzunluk değerlerinin tüm kayıtları için kutu grafiği kullanılarak veri analizi gerçekleştirilmiştir. Çalışmada, medyan ve ortalama değerlerinin hesaplaması yapılmış ve kutu grafiği üzerinde gösterilmiştir. Elde sonuçlara göre çarpıklık yönü tespit edilmiştir.

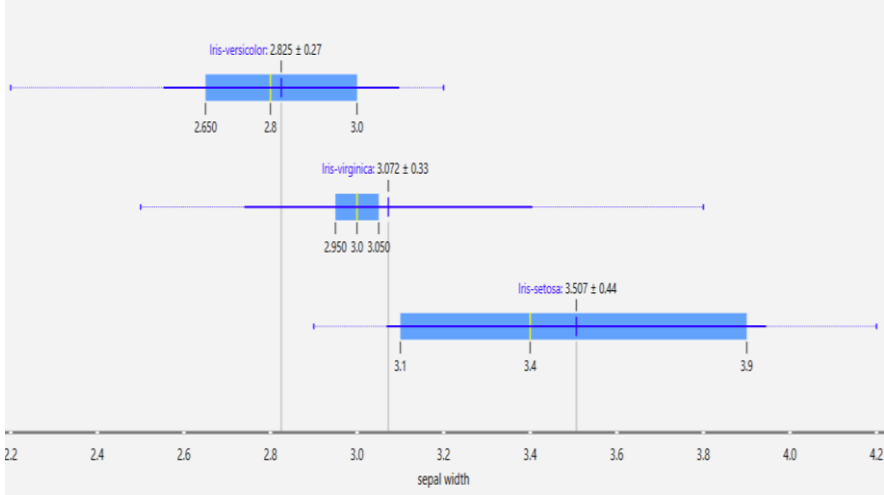
Şekil 5 üzerinde çanak yaprak uzunluğu (sepal length) öznitelik verisinin, her sınıfa ait kutu grafiği gösterilmektedir. Şekil üzerinde, İris-setosa sınıfına ait çanak yaprak uzunluğu verilerinin alt çeyreklik (lower quartile) sınır değerinin 4.85, üst çeyreklik (upper quartile) sınır değerinin 5.35 olduğu, ayrıca hesaplanan medyan değerinin 5.1 ve ortalama değerinin ise 5.073 olduğu görülmektedir. İris-versicolor sınıfına ait çanak yaprak uzunluğu verilerinin alt çeyreklik sınır değerinin 5.65, üst çeyreklik sınır değeri 6.15 olduğu ayrıca, hesaplanan medyan değerinin 5.85 ve ortalama değerinin ise 5.983 olduğu görülmektedir. İris-virginica sınıfına ait çanak yaprak uzunluğu verilerinin alt çeyreklik sınır değerinin 6.25, üst çeyreklik sınır değerinin 7.25 olduğu ayrıca, hesaplanan medyan değerinin 6.75 ve ortalama değerinin ise 6.728 olduğu görülmektedir.



Şekil 5. Çanak yaprak uzunluğu (Sepal length) veri kümesinin kutu grafiği

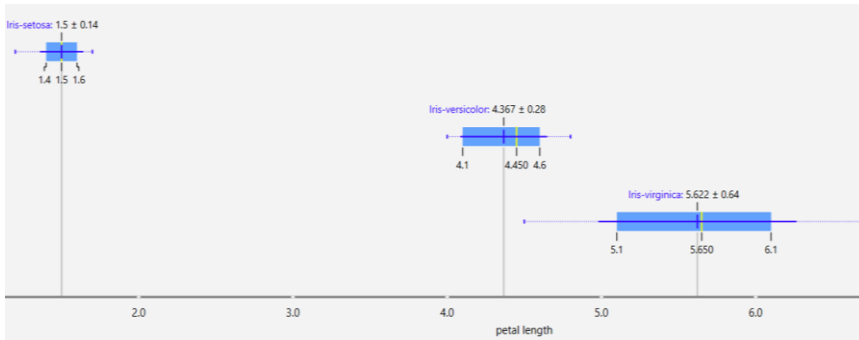
Şekil 6 üzerinde çanak yaprak genişliği (sepal width) öznitelik verisinin, her sınıfa ait kutu grafiği gösterilmektedir. Şekil üzerinde, İris-versicolor sınıfına ait çanak yaprak genişliği verilerinin alt çeyreklik (lower quartile) sınır değerinin 2.65, üst çeyreklik (upper quartile) sınır değerinin 3.00 olduğu ayrıca, hesaplanan medyan değerinin 2.80 ve ortalama değerinin ise 2.825 olduğu görülmektedir. İris-virginica sınıfına ait çanak yaprak genişliği verilerinin alt çeyreklik sınır değerinin 2.95, üst çeyreklik sınır değerinin 3.05 olduğu ayrıca, hesaplanan medyan değerinin 3.00 ve ortalama değerinin ise 3.072 olduğu görülmektedir. İris-setosa sınıfına ait çanak yaprak genişliği verilerinin alt çeyreklik sınır değerinin 3.10, üst

çeyreklik sınır değerinin 3.90 olduğu, ayrıca hesaplanan medyan değerinin 3.4 ve ortalama değerinin ise 3.507 olduğu görülmektedir.



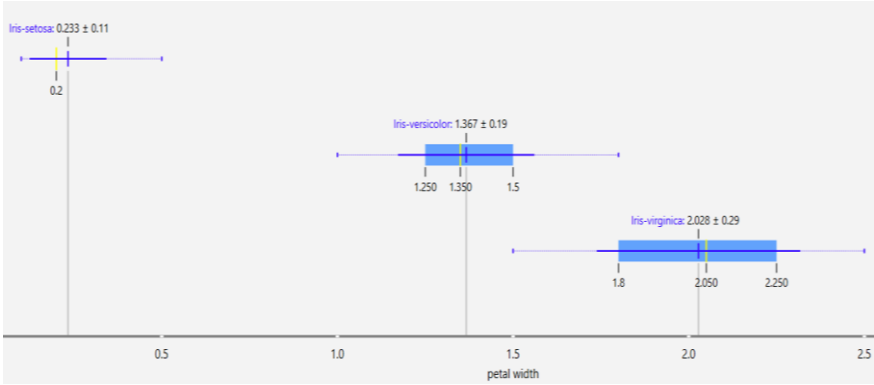
Şekil 6. Çanak yaprak genişliği (Sepal width) veri kümesinin kutu grafiği

Şekil 7 üzerinde, taç yaprak uzunluğu (petal length) öznelilik verisinin, her sınıfa ait Box plot grafiği gösterilmektedir. Şekil üzerinde, Iris-setosa sınıfına ait taç yaprak uzunluğu verilerin (lower quartile) sınır değerinin 1.40, üst çeyreklik (upper quartile) sınır değerinin 1.60 olduğu, ayrıca hesaplanan medyan değerinin 1.5 ve ortalama değerinin ise 1.5 olduğu görülmektedir. Iris-versicolor sınıfına ait taç yaprak verilerinin alt çeyreklik sınır değeri 4.1, üst çeyreklik sınır değerinin 4.6 olduğu ayrıca, hesaplanan medyan değerinin 4.45 ve ortalama değerinin ise 4.367 olduğu görülmektedir. Iris-virginica sınıfına ait taç yaprak uzunluğu verilerinin alt çeyreklik sınır değeri 5.1, üst çeyreklik sınır değerinin 6.1 olduğu, ayrıca hesaplanan medyan değerinin 5.65 ve ortalama değerinin ise 5.622 olduğu görülmektedir.



Şekil 7. Taç yaprak uzunluk (Petal length) veri kümesinin kutu grafiği

Şekil 8 üzerinde, taç yaprak genişliği (petal width) öznelilik verisinin, her sınıfa ait kutu grafiği gösterilmektedir. Şekil üzerinde, İris-setosa sınıfına ait taç yaprak genişliği değerlerinin hesaplanan medyan değerinin 0.2 ve ortalama değerinin ise 0.233 olduğu görülmektedir. İris-versicolor sınıfına ait taç yaprak genişliği verilerinin alt çeyreklik (lower quartile) sınır değerinin 1.25, üst çeyreklik (upper quartile) sınır değerinin 1.5 olduğu ayrıca, hesaplanan medyan değerinin 1.35 ve ortalama değerinin ise 1.367 olduğu görülmektedir. İris-virginica sınıfına ait taç yaprak genişliği verilerinin alt çeyreklik sınır değerinin 1.8, üst çeyreklik sınır değeri 2.25 olduğu, ayrıca hesaplanan medyan değerinin 2.05 ve ortalama değerinin ise 2.028 olduğu görülmektedir.



Şekil 8. Taç yaprak genişliği (Petal width) veri kümesinin kutu grafiği

Elde edilen ortalama ve medyan değerlerinin sütun grafiği üzerinde gösterimi şekil 9 üzerinde gösterilmiştir. Şekil üzerinde, ortalama ve medyan değerlerinin kıyaslaması yapılabilmektedir.



Şekil 9. Ortalama ve medyan değerlerinin kıyaslama grafiği

Elde değerler kullanılarak, çanak yaprak uzunluğu (sepal length), öznitelğine ait her sınıf için elde edilen ortalama ve medyan değerleri kullanılarak elde edilen çarpıklık verisi dağılım verisi, tablo 1’de gösterilmektedir. Tablo incelendiğinde İris-setosa ve İris-virginica sınıfındaki, çanak yaprak uzunluğu özniteliklerinde “Negatif Çarpıklık” olmasına rağmen İris-versicolor sınıfındaki çanak yaprak uzunluğu öznitelğinde ise “Pozitif Çarpıklık” olduğu görülmüştür.

Tablo 1. Çanak yaprak uzunluğu öznitelğinin çarpıklık dağılımı

	İris-setosa	İris-versicolor	İris-virginica
Ortalama	5,073	5,983	6,728
Medyan	5,100	5,850	6,750
Tespit	5,1>5,073	5,983>5,85	6,75>6,728
Çarpıklık Dağılımı	Negatif	Pozitif	Negatif

Çanak yaprak genişliği (sepal width), öznitelğine ait her sınıf için elde edilen ortalama ve medyan değerleri kullanılarak elde edilen çarpıklık dağılım verisi, tablo 2’de gösterilmektedir. Tablo incelendiğinde, İris-setosa, İris-versicolor ve İris-virginica sınıflarının tüm çanak yaprak genişliği özniteliklerinde, “Pozitif Çarpıklık” olduğu görülmüştür.

Tablo 2. Çanak yaprak genişliği özniteliğinin çarpıklık dağılımı

	Iris-setosa	Iris-versicolor	Iris-virginica
Ortalama	3,507	2,825	3,072
Medyan	3,400	2,800	3,000
Tespit	3,507>3,4	2,825>2,8	3,072>3
Çarpıklık Dağılımı	Pozitif	Pozitif	Pozitif

Taç yaprak uzunluk (petal length) özniteliğine ait her sınıf için elde edilen ortalama ve medyan değerleri kullanılarak elde edilen çarpıklık dağılım verisi, tablo 3’de gösterilmektedir. Tablo incelendiğinde İris-virginica ve İris-versicolor sınıfındaki, taç yaprak uzunluk özniteliklerinde “Negatif Çarpıklık” olmasına rağmen İris-setosa sınıfındaki taç yaprak uzunluk özniteliğinde ise, “Simetrik Dağılım” olduğu yani çarpıklık olmadığı görülmüştür.

Tablo 3. Taç yaprak uzunluk özniteliğinin çarpıklık dağılımı

	İris-setosa	İris-versicolor	İris-virginica
Ortalama	1,500	4,367	5,622
Medyan	1,500	4,450	5,650
Tespit	1,5=1,5	4,45>4,367	5,65>5,622
Çarpıklık Dağılımı	Simetrik Dağılım (Çarpıklık Yok)	Negatif	Negatif

Taç yaprak genişlik (petal width) özniteliğine ait her sınıf için elde edilen ortalama ve medyan değerleri kullanılarak elde edilen çarpıklık dağılım verisi ise tablo 4’de gösterilmektedir. Tablo incelendiğinde İris-setosa ve İris-versicolor sınıfındaki, taç yaprak genişlik özniteliklerinde “Pozitif Çarpıklık” olmasına rağmen İris-virginica sınıfındaki taç yaprak genişlik özniteliğinde ise “Negatif Çarpıklık” olduğu görülmüştür.

Tablo 4. Taç yaprak genişlik özniteliğinin çarpıklık dağılımı

	İris-setosa	İris-versicolor	İris-virginica
Ortalama	0,233	1,367	2,028
Medyan	0,200	1,350	2,050
Tespit	0,233>0,2	1,367>1,35	2,05>2,028
Çarpıklık Dağılımı	Pozitif	Pozitif	Pozitif

Sonuç ve Tartışma

Çarpıklık, bir kümesinin değerlerinin aykırılık derecesini göstermektedir. Pozitif çarpıklık olan veri kümesi içinde sağ yönde dağılım, negatif çarpıklık olan veri kümesinde ise sol yönde dağılımlar görülmektedir. Pearson, Moment temelli, Bowley ve Kelley yöntemleri kullanılarak çarpıklık katsayısı hesaplanabilmektedir. Veri analizi işlemlerinde çarpıklık dağılımın tespit edilmesi, model tasarımlarında çok büyük fayda sağlamaktadır.

Bu çalışmada, literatürde yaygın kullanılan İris veri seti özniteliklerinin, veri analizi yapılarak çarpıklık dağılımları tespit edilmiştir. Bunun için ortalama ve medyan değerleri belirlenerek kutu grafik aracı üzerinde gösterilmiş daha sonra ortalama ve medyan değerlerinin karşılaştırması yapılarak çarpıklık yönleri belirlenmiştir.

Bu çalışmada, veri bilimi alanlarında bir ön işlem adımı olarak çarpıklık dağılımının yapılmasının faydalı olabileceğini gösterilmiştir. Bu şekilde aykırı değerlerin tespit edilmesi kolayca yapılarak veriler hakkında ön bilgi elde edilebilecektir. Elde edilen sonuçlara dayanarak, gelecekteki çalışmalarda, farklı veri setleri üzerinde çarpıklık dağılım tespiti yapılarak, veri setlerinin analizi gerçekleştirilecektir.

Kaynakça

- Abt, M., Loibl, K., Leuders, T., Van Dooren, W., & Reinhold, F. (2025). Understanding student errors in comparing data sets with boxplots. *Educational Studies in Mathematics*. Advance online publication. <https://doi.org/10.1007/s10649-025-10387-z>
- Çelik, A., & Kaptan, D. (2025). Text classification by machine learning algorithms using a new text feature extraction method based on image processing. *Turkish Journal of Engineering*, 9(4), 712–724.
- De Silva, G. (2020). Exploring the world of artificial neural networks – A beginner's overview. <https://doi.org/10.13140/RG.2.2.14790.14406>
- de la Rubia, J. M. (2024). Kelley's coefficients of skewness using R. *Journal of Statistics and Applied Sciences*, (10), 50–75. <https://doi.org/10.52693/jsas.1510230>
- Dua, D., & Graff, C. (2019). UCI Machine Learning Repository. Irvine, CA: University of California, School of Information and Computer Science. <https://archive.ics.uci.edu/ml>
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188.
- Groeneveld, R. A., & Meeden, G. (1984). Measuring skewness and kurtosis. *The Statistician*, 33(4), 391–399.
- Jaspersoft (2025), What-is-a-box-plot, Retrieved October 16, 2025 from <https://www.jaspersoft.com/articles/what-is-a-box-plot>
- Joanes, D. N., & Gill, C. A. (1998). Comparing measures of sample skewness and kurtosis. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47(1), 183–189.
- Kambale, W. V., Deeb, A., Bernabia, T., Machot, F. A., & Kyamakya, K. (2023). A boxplot metadata configuration impact on time series forecasting and transfer learning. In *2023 27th International Conference on Circuits, Systems, Communications and Computers (CSCC)* (pp. 1–5).
- Kim, T. H., & White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1(1), 56–73.
- Majaw, N., & Ahmed, S. S. (2023). Exploring data distributions using box and whisker plot analysis. In *Proceedings of the 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1–8). <https://doi.org/10.1109/ICCCNT56998.2023.10308191>
- Omelina, L., Goga, J., Pavlovicova, J., Oravec, M., & Jansen, B. (2021). A survey of iris datasets. *Image and Vision Computing*, 108, Article 104109.
- Sawant, S. (2023). IRIS EDA [Kaggle Notebook]. Kaggle. <https://www.kaggle.com/code/sunaysawant/iris-eda/notebook>

Seward, L. E., & Doane, D. P. (2011). Measuring skewness: A forgotten statistic? *Journal of Statistics Education*, 19(2), 1–18.

Taha Chicho, B., Mohsin Abdulazeez, A., Qader Zeebaree, D., & Assad Zebari, D. (2021). Machine Learning Classifiers Based Classification For IRIS Recognition. *Qubahan Academic Journal*, 1(2), 106–118. <https://doi.org/10.48161/qaj.v1n2a48>

Thirunavukkarasu, K., Singh, A. S., Rai, P., & Gupta, S. (2018). Classification of IRIS dataset using classification based KNN algorithm in supervised learning. In *2018 4th International Conference on Computing Communication and Automation (ICCCA)* (pp. 1–4). IEEE.

Turney, S. (2023). Skewness Definition, examples & formula. Scribbr. Retrieved August 12, 2025, from <https://www.scribbr.com/statistics/skewness/>

Williams, J., Hill, R. R., Pignatiello Jr., J. J., & Chicken, E. (2021). Wavelet analysis of variance box plot. *Journal of Applied Statistics*, 49(14), 3536–3563. <https://doi.org/10.1080/02664763.2021.1951685>

Williamson, D. F., Parker, R. A., & Kendrick, J. S. (1989). The box plot: A simple visual method to interpret data. *Annals of Internal Medicine*, 110(11), 916–921. <https://doi.org/10.7326/0003-4819-110-11-916>

