VERİ MAHREMİYETİNİN VE İSTATİSTİKİ ÖZELLİKLERİN KORUNMASINDA MAKİNE ÖĞRENİMİ İLE META SEZGİSEL YÖNTEMLERİN ENTEGRASYONU

T.C. KOCAELİ SAĞLIK VE TEKNOLOJİ ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

ÖZGÜR SAĞIR

YÜKSEK LİSANS TEZİ
BİLİŞİM SİSTEMLERİ MÜHENDİSLİĞİ YÜKSEK LİSANS PROGRAMI

KOCAELİ 2025

ONAY SAYFASI

Bu tez Kocaeli Sağlık ve Teknoloji Üniversitesi Lisansüstü Eğitim Enstitüsü tarafından onaylanmıştır.

Prof.Dr.Yurdanur DİKMEN
Enstitü Müdürü

Bu tezin **Kocaeli Sağlık ve Teknoloji Üniversitesi, Lisansüstü Eğitim Enstitüsü Bilişim Sistemleri Mühendisliği Bölümünde Yüksek Lisans** derecesinin tüm gerekliliklerini karşıladığını onaylıyorum.

Dr.Öğr.Üyesi Mehmet KARA
Bölüm Başkanı

ÖZGÜR SAĞIR tarafından teslim edilen VERİ MAHREMİYETİNİN VE İSTATİSTİKİ ÖZELLİKLERİN KORUNMASINDA MAKİNE ÖĞRENİMİ İLE META SEZGİSEL YÖNTEMLERİN ENTEGRASYONU başlıklı bu tezin kapsam ve kalite bakımından Yüksek Lisans derecesi için yeterli olduğunu değerlendirmektedir.

Dr.Öğr.Üyesi Ulaş VURAL
Danışman

Tez Jürisi Üyeleri: (1. isim: Jüri başkanı; 2. isim: Danışman)

Dr.Öğr.Üyesi Hikmetcan ÖZCAN
Bilgisayar Mühendisliği Bölümü, Kocaeli Üniversitesi _____
Sağlık ve Teknoloji Üniversitesi
Dr.Öğr.Üyesi Ulaş VURAL
Bilişim Sistemleri Mühendisliği Bölümü, Kocaeli _____
Sağlık ve Teknoloji Üniversitesi
Prof.Dr.Nevcihan DURU
Bilişim Sistemleri Mühendisliği Bölümü, Kocaeli _____

Tarih: 20/01/2025

ORİJİNALLİK SAYFASI

İşbu belge ile çalışmamda yer alan tüm bilgilerin akademik ve etik kurallara uygun olarak elde edildiğini ve sunulduğunu beyan ederim. Ayrıca, kurallar gereği bu çalışmada özgün olmayan tüm materyal ve sonuçlar için ilgili kaynakların verildiğini beyan ederim.

Ad, Soyad: Özgür SAĞIR

İmza:

ABSTRACT

INTEGRATING MACHINE LEARNING AND META HEURISTICS IN PROTECTING DATA PRIVACY AND STATISTICAL PROPERTIES

SAĞIR, Özgür

Master's Thesis, Graduate Education Institute Information Systems Engineering

Supervisor: Dr.Ulaş VURAL

January 2025, 64 pages

This study offers an innovative approach to preserving the privacy of sensitive data types, such as health data. Considering the tendency of traditional methods to reduce data analysis accuracy, the integration of machine learning and metaheuristic methods is aimed at achieving both data privacy and the preservation of statistical properties without compromise.

In this study, experiments were conducted on synthetic patient data using genetic algorithms and differential privacy mechanisms. Initially, privacy levels were optimized through different epsilon values, followed by data analysis using classification algorithms. Additionally, Principal Component Analysis (PCA) was employed to reduce data dimensions, and the results were evaluated.

The findings demonstrate that the Laplace mechanism effectively ensures privacy for both continuous and categorical data, and genetic algorithm-optimized parameters play a significant role in balancing accuracy and privacy. The XGBoost algorithm achieved the highest accuracy scores in the integrated datasets.

This study provides significant contributions to the literature by developing more reliable and efficient analysis methods for sensitive data types through the combination of differential privacy methods and metaheuristic algorithms.

Keywords: Differential privacy, machine learning, meta-heuristics, genetic algorithms, data privacy.

ÖZ

VERİ MAHREMİYETİNİN VE İSTATİSTİKİ ÖZELLİKLERİN KORUNMASINDA MAKİNE ÖĞRENİMİ İLE META SEZGİSEL YÖNTEMLERİN ENTEGRASYONU

SAĞIR, Özgür

Yüksek Lisans, Lisansüstü Eğitim Enstitüsü Bilişim Sistemleri Mühendisliği

Tez Yöneticisi: Dr.Öğr.Üyesi Ulaş VURAL

Ocak 2025, 64 sayfa

Bu çalışma, sağlık verileri gibi hassas veri türlerinin gizliliğini koruma konusuna yenilikçi bir yaklaşım sunmaktadır. Geleneksel yöntemlerin veri analiz doğruluğunu düşürme eğilimi dikkate alınarak, makine öğrenimi ve meta sezgisel yöntemlerin entegrasyonu ile hem veri mahremiyetinin sağlanması hem de istatistiki özelliklerin bozulmadan korunması hedeflenmiştir.

Çalışmada, genetik algoritmalar ve diferansiyel mahremiyet mekanizmaları temel alınarak sentetik hasta verileri üzerinde deneyler gerçekleştirilmiştir. Öncelikle, farklı epsilon değerleriyle gizlilik seviyeleri optimize edilmiş, ardından sınıflandırma algoritmaları ile veri analizi yapılmıştır. Ayrıca, Principal Component Analysis (PCA) kullanılarak veri boyutu düşürülmüş ve sonuçlar değerlendirilmiştir.

Elde edilen sonuçlar, özellikle Laplace mekanizmasının sürekli ve kategorik verilere etkili bir şekilde gizlilik sağlayabildiğini ve genetik algoritmalarla optimize edilmiş parametrelerin doğruluk ve mahremiyeti dengelemede önemli bir rol oynadığını göstermiştir. XGBoost algoritması, entegre edilen veri setlerinde en yüksek doğruluk skorlarını elde etmiştir.

Bu çalışma, literatüre, diferansiyel mahremiyet yöntemlerinin meta sezgisel algoritmalarla birleştirilmesi yoluyla, hassas veri türleri üzerinde daha güvenilir ve etkili analiz yöntemleri geliştirilmesine yönelik önemli katkılar sunmaktadır.

Anahtar Kelimeler: Diferansiyel mahremiyet, makine öğrenimi, meta sezgisel yöntemler, genetik algoritmalar, veri gizliliği.

İTHAF SAYFASI

Canımdan kıymetli, biricik oğlum Mithat Kutay'a ithaf edilmiştir.

TEŞEKKÜR

Tez çalışmam boyunca değerli fikirleri ve önerileriyle her zaman yanımda olan ve yardımını hiçbir zaman esirgemeyen kıymetli danışman hocam **Dr. Öğr. Üyesi Ulaş VURAL**'a**,** konu ile ilgili sürekli akıl danıştığım sevgili dostum Candide ÖZTÜRK'e ve bütün kıymetli hocalarıma çok teşekkür ederim.

Tüm bu süreçlerde bana her türlü kolaylığı sağlayan değerli çalışma arkadaşlarım Fatih SARIALTIN, Davut İSPEKTER, Hüseyin SARIKAYA, M. Onur ARSLANOĞLU, Sinan Cem KOZAN, Ümmühan KARATAŞ ve Havva ALKUL'a sonsuz teşekkürler.

Son olarak, her zaman yanımda olduklarını bildiğim, düştüğüm zaman beni kaldıran ve en başından beri beni cesaretlendiren değerli eşim Güzide ŞENDAĞ SAĞIR'a, varlığıyla bana huzur ve cesaret veren biricik oğlum Mithat Kutay'a ve sevgili geniş aileme; destekleri, varlıkları ve yanımda olmaları nedeniyle minnettarım.

Özgür SAĞIR
Ocak-2025, KOCAELİ

İÇİNDEKİLER

ŞEKİLLER LİSTESİ

TABLOLAR LİSTESİ

SEMBOL/KISALTMA LİSTESİ

ABAC	Attribute-Based Access Control (Öznitelik Tabanlı Erişim Kontrolü)
AES	Advanced Encryption Standard (Gelişmiş Şifreleme Standardı)
CT	Computed Tomography (Bilgisayarlı Tomografi)
CXR	Chest X-Ray (Göğüs Röntgeni)
DEAP	Distributed Evolutionary Algorithm in Python – (Python'da Dağıtılmış Evrimsel Algoritmalar)
FN	False Negative (Yanlış Negatif)
FP	False Positive (Yanlış Pozitif)
GDPR	General Data Protection Regulation (Genel Veri Koruma Tüzüğü)
KNN	K-Nearest Neighbors (K-En Yakın Komşu)
KVKK	Kişisel Verilerin Korunması Kanunu
MSE	Mean Squared Error (Ortalama Kare Hata)
PCA	Principal Component Analysis (Temel Bileşen Analizi)
PSO	Parçacık Sürü Optimizasyonu
RBAC	Role-Based Access Control (Rol Tabanlı Erişim Kontrolü)
RKHS	Reproducing Kernel Hilbert Space (Üretici Çekirdek Hilbert Uzayı)
RSA	Rivest-Shamir-Adleman
SMPC	Secure Multi-Party Computation (Güvenli Çok Partili Hesaplama)
SVM	Support Vector Machine (Destek Vektör Makinesi)
TN	True Negative (Gerçek Negatif)
TP	True Positive (Gerçek Pozitif)

1. GİRİŞ

Veri gizliliği, özellikle sağlık ve finans gibi hassas verilerin yaygın kullanımı nedeniyle günümüzün en önemli konularından biri haline gelmiştir. Bu tür verilerin gizliliğini korumak için uygulanan geleneksel yöntemler, sıklıkla veri analizinin doğruluğunu olumsuz etkileyebilmektedir (Atalay, 2021). Son yıllarda, makine öğrenimi ve meta sezgisel yöntemlerin entegrasyonu, hem veri mahremiyetini koruyarak hem de veri setlerinin istatistiki özelliklerini bozmadan çeşitli analizler yapabilmenin yeni bir yolunu sunmaktadır. Bu çalışma, veri gizliliğini sağlarken aynı zamanda analitik performanstan ödün vermeyen yenilikçi yöntemlerin geliştirilmesine odaklanmaktadır.

Veri mahremiyetinin korunması ve veri etiğinin sağlanması, özellikle Genel Veri Koruma Tüzüğü (General Data Protection Regulation – GDPR) ve Kişisel Verilerin Korunması Kanunu (KVKK) gibi yasal düzenlemelerin etkisiyle daha fazla önem kazanmıştır (Sağır ve Vural, 2024(I)). Bu düzenlemeler, bireylerin kişisel bilgilerinin korunması için katı kurallar getirirken, aynı zamanda veri analizinin ve makine öğrenimi uygulamalarının da zorluklarla karşılaşmasına yol açmaktadır (Dwork ve Roth, 2014). Geleneksel veri anonimleştirme yöntemleri, veri gizliliğini sağlarken analiz performansını düşürme eğilimindedir. Özellikle veri doğruluğunun kaybolması ve istatistiksel eğilimlerin bozulması, analiz sonuçlarının güvenilirliğini tehdit etmektedir (Narayanan ve Shmatikov, 2008).

Bu bağlamda, makine öğrenimi ve meta sezgisel yöntemler, veri gizliliği ve istatistiksel tutarlılığı korumak için alternatif bir yaklaşım sunmaktadır. Meta sezgisel yöntemler, küresel optimizasyon yetenekleri sayesinde büyük veri setleri üzerinde gizlilik kaybını en aza indiren ve istatistiki özellikleri bozmayan çözümler üretebilir (Yang, 2014). Örneğin, genetik algoritmalar ve parçacık sürü optimizasyonu gibi teknikler, veri bozulmasını minimum düzeyde tutarak gizlilik koruma mekanizmalarını iyileştirebilir (Eiben ve Smith, 2015).

Diferansiyel mahremiyet, son yıllarda en çok tartışılan veri gizliliği yöntemlerinden biri olarak öne çıkmaktadır. Bu yaklaşım, belirli bir bireyin veri setine dahil olup olmamasının analiz sonuçlarını önemli ölçüde etkilememesini garanti eder (Dwork, 2006). Ancak, diferansiyel mahremiyet uygulamaları sırasında eklenen gürültü

miktarı, veri kalitesini önemli ölçüde düşürebilir. Bu nedenle, diferansiyel mahremiyetin etkinliğini artırmak için makine öğrenimi ile meta sezgisel yöntemlerin entegrasyonu, daha verimli gürültü ekleme stratejileri geliştirilmesine olanak tanımaktadır (Balle ve Wang, 2018).

Bir diğer çalışmamızda (Sağır ve Vural, 2024(II)), diferansiyel mahremiyet konusunda Laplace ve Gaussian mekanizmaları sağlık verileri üzerinde kullanılmıştır. Gaussian mekanizması, verilerin gizliliğini koruma açısından daha fazla varyans ekleyerek daha geniş bir dağılım yaratmıştır. Bu mekanizma özellikle Random Forest algoritması ile en iyi sonuçları vermiştir. Ancak genel olarak Gaussian veriseti diğer algoritmalar için doğrulukta düşüşe neden olmuştur. Laplace mekanizması ise, daha stabil sonuçlar üretmiş ve orijinal veriye daha yakın değerler sunmuştur. XGBoost algoritması, Laplace mekanizmasında en yüksek performansı göstermiştir. Bu durum, Laplace mekanizmasının verilerde daha az varyansa yol açarak daha yüksek doğruluk sunduğunu göstermektedir.

Bu çalışmada ise, Laplace ve Gaussian mekanizmalarının ötesinde, derin öğrenme ve meta sezgisel algoritmalar kullanılarak geliştirilen yöntemlerin, veri mahremiyetini koruma ve istatistiksel özellikleri muhafaza etme üzerindeki etkileri incelenecektir. Makine öğreniminin veri dağılımlarındaki karmaşık ilişkileri öğrenme yeteneği ile meta sezgisel yöntemlerin optimizasyon becerileri birleştirilerek, daha esnek ve etkili gizlilik koruma yöntemleri sunulması hedeflenmektedir.

2. VERİ MAHREMİYETİ SAĞLAMA YÖNTEMLERİ

Veri mahremiyeti, günümüzün dijital çağında kişisel ve kurumsal bilgilerin gizliliğini korumak adına kritik bir konudur. Kişisel verilerin güvenli bir şekilde saklanması, yetkisiz erişimlerden korunması ve veri ihlallerinin önlenmesi için çeşitli yöntemler ve teknolojiler geliştirilmiştir. Bu yöntemler arasında şifreleme, anonimleştirme, erişim kontrolü ve diferansiyel mahremiyet gibi teknikler öne çıkar. Her bir yöntem, belirli bir güvenlik seviyesini sağlamak amacıyla tasarlanmıştır ve veri sahiplerinin gizliliğini korumayı hedefler. Bu bağlamda, veri mahremiyeti sağlama yöntemleri hem bireylerin hem de kurumların dijital güvenliğini artırarak, modern dünyanın dijital dönüşüm sürecinde önemli bir rol oynar.

2.1. Şifreleme

Şifreleme, verilerin matematiksel algoritmalar kullanılarak gizli bir forma dönüştürülmesi işlemidir. Bu yöntem, yetkisiz kişilerin verilere erişmesini engellemek için kullanılır ve veri gizliliğini sağlamak için yaygın olarak uygulanır. Modern şifreleme teknikleri, simetrik ve asimetrik şifreleme olarak ikiye ayrılır (Beşkirli ve ark., 2019). Simetrik şifreleme, tek bir anahtarın hem şifreleme hem de deşifreleme için kullanıldığı yöntemdir (örneğin, AES (Advanced Encryption Standard - Gelişmiş Şifreleme Standardı). Asimetrik şifreleme ise iki farklı anahtar (açık ve özel anahtar) kullanır ve daha güvenli kabul edilir (örneğin, RSA (Rivest-Shamir-Adleman) algoritması).

Simetrik şifreleme hızlı ancak anahtar paylaşımı sorunu varken, asimetrik şifreleme güvenli ancak yavaştır. Bu nedenle, çoğu sistem her iki algoritmanın güçlü yönlerini birleştirerek daha güvenli ve verimli bir şifreleme sağlar (Kodaz ve Botsalı, 2010).

2.2. Anonimleştirme

Anonimleştirme, kişisel verilerin kimliğini belirlemeyi imkânsız hale getiren bir süreçtir. Bu yöntem, özellikle sağlık ve finans gibi hassas verilerin korunmasında önemlidir. Anonimleştirilen veriler, bireylerin kimliklerini ortaya çıkarmadan analiz edilebilir ve paylaşılabilir. K-anonimlik, l-çeşitlilik ve t-yakınlık gibi yöntemler bu teknik kapsamında kullanılır (Vural ve Aydos, 2018).

2.3. Pseudonymization

Pseudonymization, verilerin kimliğini belirleyici unsurlarını gizleyerek ve bu unsurları sahte verilerle değiştirerek koruma sağlar. Ancak, bu sahte veriler belirli koşullar altında geri dönüştürülebilir. Bu yöntem, Avrupa Birliği'nin GDPR gibi veri koruma yasalarında önemli bir rol oynar ve kimlik bilgisi gizli tutularak veri analizine olanak tanır (Neubauer ve Heurix, 2011).

2.4. Erişim Kontrolü

Erişim kontrolü, yalnızca yetkilendirilmiş kişilerin verilere erişimini sağlamak için kullanılan bir yöntemdir. Erişim politikaları, kullanıcıların belirli verilere hangi düzeyde erişebileceğini tanımlar (Aygün, 2015). Örnek olarak, Rol Tabanlı Erişim Kontrolü (Role-Based Access Control - RBAC) ve Öznitelik Tabanlı Erişim Kontrolü (Attribute-Based Access Control - ABAC) gibi yöntemler kullanılabilir (Bailey ve ark., 2011).

2.5. Diferansiyel Mahremiyet

Diferansiyel mahremiyet yaklaşımı, kişisel verilerin gizliliğini koruyarak veri analizi yapılmasına imkân tanır (Dwork, 2006). Bu yöntem, veri setine rastgele gürültülerle kişisel bilgilerin ortaya çıkmasını engeller. Böylece, analizlerden elde edilen veriler, kimliklerini ifşa etmeden güvenli bir şekilde kullanılabilir (Abowd ve Schmutte, 2019). Diferansiyel mahremiyetin sunduğu genişletilmiş garanti, veri setinde bir bireyin bulunup bulunabildiğinin tespit edilip korunmasını sağlar (Mironov, 2017). Bu özellik, özellikle hassas sağlık verilerinin korunmasında önemli bir rol oynamaktadır.

Laplace mekanizması, gürültülerin Laplace verilerine göre belirlendiği daha basit bir yöntemken (Mironov, 2017), Gaussian mekanizması ise gürültülerin Gaussian koşullarının temel olarak eklendiği daha karmaşık bir ortamlara sahiptir (Balle ve ark., 2018). Her iki mekanizma da gizlilik sağlamak amacıyla gürültü ekleme teknisyenlerini kullanır, ancak birbirlerine göre farklı avantajlar ve dezavantajlar sunarlar.

Diferansiyel mahremiyetin önemli bir kişiden biri, hassas bilgilerin korunmasında etkili bir çözüm sunmasıdır; Kimlik bilgileri sayesinde kişisel verilerin etkisi azaltılır ve anonimlik sağlanır. Bu mekanizmanın özellikle büyük veri kümeleri üzerinde

güvenli analiz yapılması mümkündür (Liu ve ark., 2024). Ayrıca yöntemin bileşimsel (bileşimsel) yapısı, farklı uygulamaların bir araya getirilmesine imkân tanıyarak karmaşık veri analizlerinde bile gizliliğin gizliliğinin korunmasını sağlar (Kairouz ve Viswanath, 2015). Bununla birlikte, bu avantaj, çok sayıda uygulamanın bir araya getirilmesi durumunda gizlilik iznini açabilir ve Gizlilik seviyesi düşebilir.

Diğer bir güçlük ise diferansiyel mahremiyetin etkili bir şekilde uygulanabilmesi için epsilon ve deltanın silinebilmesinin doğru ayarını gerektirmesidir [(Kairouz ve Viswanath, 2015) ve (Murtagh ve Vadhan, 2015)]. Küçük epsilon değerleri daha yüksek gizlilik sağlarken, büyük epsilon değerleri daha düşük gizlilik sunar. Bu izin seçimi, gizlilik ve veri aralığı arasındaki bir denge parçalarının karmaşıklığı mümkündür. Ayrıca gürültü ekleme verisi olumsuzluk yaratabilir ve bu da analiz sonuçlarının doğruluğunu düşürebilir.

2.6. Veri Maskeleme

Veri maskeleme, hassas verilerin gerçek değerlerinin yerini sahte veya değiştirilmiş verilerle doldurarak korunmasını sağlar. Bu yöntem, özellikle test ve geliştirme ortamlarında veri sızıntılarını önlemek için kullanılır. Maskelenen veriler, orijinal verilerin özelliklerini taşır, ancak hassasiyetini yitirmiş olur (Yang ve ark., 2015).

2.7. Kapsülleme

Veri kapsülleme, verilerin farklı katmanlarda korunarak güvenliğinin sağlanması sürecidir. Bu yöntem, izinsiz erişimlerden korunmak için veri yapılarını ve kontrol mekanizmalarını kullanır. Özellikle ağ güvenliği ve yazılım geliştirme süreçlerinde veri kapsülleme, verilerin güvenli bir şekilde aktarılmasını sağlar (Lau ve Taweel, 2007).

2.8. Meta Sezgisel Yöntemler

Meta sezgisel yöntemler, karmaşık optimizasyon problemlerini çözmek için kullanılan heuristik algoritmaları ifade eder. Bu yöntemler, geleneksel eniyileme yöntemlerinin kabul edilebilir bir çözüm süresinde etkin ve uygun çözümler üretebileceğini hedefler (Var ve İnan, 2018). Başlıca meta sezgisel yöntemler arasında genetik algoritma, tavlama benzetimi, tabu arama, karınca kolonisi algoritması, yapay sinir ağları ile yapay zeka ve makine öğrenimi gösterilmektedir (Sümer, 2019).

2.8.1. Genetik Algoritma

Genetik algoritmalar, biyolojik evrim süreçlerini taklit ederek en iyi çözümleri bulmayı amaçlar. Bu yöntem, popülasyonlar, çaprazlama, mutasyon ve seçim gibi genetik operatörleri kullanır ve karmaşık optimizasyon problemlerini çözmede etkilidir (Holland, 1975).

2.8.2. Tavlama Benzetimi

Bu yöntem, fiziksel tavlama sürecini benzetir ve arama alanında global optimuma ulaşmayı hedefler. Yüksek sıcaklıkta başlayan ve yavaşça soğuyan bir sistem gibi, çözüm uzayında daha düşük enerjili durumlara geçiş yapar (Kirkpatrick ve ark., 1983).

2.8.3. Tabu Arama

Tabu arama, yerel arama algoritmalarının bir varyantıdır ve belirli bir süre boyunca ziyaret edilen çözümleri tabu listesinde tutarak tekrar ziyaret etmeyi engeller. Bu yöntem, yerel minimumlardan kaçınarak daha iyi çözümler bulmayı hedefler (Glover, 1986).

2.8.4. Karınca Kolonisi Algoritması

Bu algoritma, karıncaların yemek bulma davranışlarını taklit eder. Karıncalar, feromon izleri bırakarak en kısa yolları bulur ve bu bilgi, algoritmanın daha etkili çözümler bulmasına yardımcı olur (Dorigo ve ark., 1997).

2.8.5. Yapay Sinir Ağları

Yapay sinir ağları, insan beyninin işleyişine benzer bir şekilde bilgiyi işler. Bu ağlar, çok sayıda bağlantılı düğüm ve katmandan oluşur ve verileri öğrenerek çıkarımlarda bulunur (LeCun ve ark., 2015).

2.8.6. Yapay Zeka ve Makine Öğrenimi

Bu yöntemler, veri analizi ve öğrenme süreçlerinde kullanılır ve çok çeşitli algoritmaları içerir. Makine öğrenimi, verilerden öğrenerek tahmin ve karar verme süreçlerini otomatikleştirir (Mitchell, 1997).

Makine öğrenmesi, bilgisayar algoritmalarının verilerdeki kalıpları öğrenerek tahminler yapmasını sağlayan bir alandır. KNN, Naive Bayes, Decision Trees, Logistic Regression, K-means, SVM ve yapay sinir ağları gibi çeşitli algoritmalar, farklı türdeki problemler için kullanılır. Bu algoritmalar, denetimli, denetimsiz ve pekiştirmeli olmak üzere üç ana öğrenme stratejisiyle çalışır. Denetimli öğrenmede, algoritma doğru cevaplarla eğitilerek yeni verilere dair tahminler yapar. Denetimsiz öğrenmede ise verilerdeki doğal grupları keşfeder. Pekiştirilmiş öğrenmede ise algoritma, yaptığı eylemlerin sonuçlarına göre öğrenir ve en iyi sonucu veren eylemleri seçmeye çalışır (Atalay ve Çelik, 201).

Yapay zeka, günümüzde veriye dayalı karar verme süreçlerinin ayrılmaz bir parçası haline gelmiştir. Büyük veri setlerinden anlamlı bilgiler çıkarmak için kullanılan yapay zeka algoritmaları, işletmelerin daha iyi tahminlerde bulunmasına, müşteri davranışlarını daha iyi anlamasına ve yeni iş fırsatları yaratmasına yardımcı olmaktadır. Makine öğrenmesi, derin öğrenme ve doğal dil işleme gibi yapay zeka alt alanları, veri analizi süreçlerini otomatikleştirerek insan hatasını minimize eder ve daha hızlı sonuçlar elde edilmesini sağlar. Sağlık, finans, pazarlama gibi birçok sektörde yapay zeka ve veri analitiği birlikte kullanılmakta, bu sayede daha etkili ve verimli çözümler üretilmektedir. Örneğin, sağlık sektöründe yapay zeka, tıbbi görüntüleme verilerini analiz ederek hastalıkların erken teşhisine yardımcı olurken, finans sektöründe ise dolandırıcılık faaliyetlerinin tespitinde kullanılmaktadır (Himeur ve ark., 2023).

2.9. Federatif Öğrenme

Federatif öğrenme, merkezi bir veri seti kullanmadan dağıtık veri kaynaklarında makine öğrenimi modelleri eğitmeyi mümkün kılan bir yöntemdir. Bu yaklaşım, veri gizliliğini koruyarak farklı cihazların veya kurumların verilerini bir araya getirmeden ortak bir model geliştirmeyi hedefler (McMahan ve ark., 2017). Her bir veri kaynağı, kendi yerel verisi üzerinde modelini günceller ve yalnızca bu güncellemeler merkezi bir sunucuya gönderilir (Konecny ve ark., 2016). Böylece, veri sahiplerinin hassas bilgilerinin gizliliği korunurken, tüm katılımcıların katkısıyla daha güçlü ve genelleştirilebilir bir model elde edilir (Yang ve ark., 2019). Federatif öğrenme, sağlık, finans ve kullanıcı gizliliği gerektiren diğer alanlarda geniş uygulama potansiyeline sahiptir.

3. MAKİNE ÖĞRENİMİ İLE META SEZGİSEL YÖNTEMLERİN ENTEGRASYONU

Makine öğrenimi, büyük veri setlerinden öğrenme ve tahmin yapma yeteneğine sahip algoritmalardan oluşan bir disiplindir. Özellikle sınıflandırma, regresyon, kümeleme gibi problemler için yaygın olarak kullanılmaktadır. Ancak, makine öğrenimi algoritmalarının performansı, modelin eğitimi sırasında belirlenen hiperparametrelere, veri ön işleme süreçlerine ve modelin karmaşıklığına bağlı olarak değişebilir. Bu nedenle, makine öğrenimi algoritmalarının etkinliğini artırmak için hiperparametre optimizasyonu, özellik seçimi ve model birleştirme gibi stratejilere ihtiyaç duyulmaktadır.

Meta sezgisel yöntemler, karmaşık optimizasyon problemlerini çözmek için kullanılan doğaya dayalı arama algoritmalarından oluşur. Bu yöntemler, özellikle arama uzayında global çözümleri keşfetme ve lokal minimumlardan kaçınma yetenekleri ile öne çıkar. Genetik algoritmalar, parçacık sürü optimizasyonu, karınca kolonisi optimizasyonu ve tabu arama gibi birçok farklı meta sezgisel yöntem bulunmaktadır. Bu yöntemler, arama uzayında geniş çaplı bir keşif sağlamak ve yüksek kaliteli çözümler elde etmek için kullanılabilir.

Makine öğrenimi ve meta sezgisel yöntemlerin entegrasyonu, her iki yaklaşımın güçlü yanlarından faydalanmayı amaçlamakta ve veri mahremiyeti ile istatistiki özellikleri korumayı hedefleyen çalışmalarda önemli bir rol oynamaktadır. Bu yöntemlerin etkin bir şekilde bir araya getirilmesi, veri gizliliği sağlanırken aynı zamanda veri analizi ve model doğruluğunun iyileştirilmesine olanak tanır.

3.1. Hiperparametre Optimizasyonu

Meta sezgisel yöntemler, makine öğrenimi modellerinin hiperparametrelerini optimize etmek için etkili araçlar sunar. Örneğin, Genetik Algoritma veya Parçacık Sürü Optimizasyonu, modelin en iyi performansını sağlamak için hiperparametrelerin optimum kombinasyonlarını arayabilir. Bu, özellikle modelin öğrenme hızı, ağaç sayısı, düğüm derinliği gibi hiperparametrelerin ayarlanmasında faydalıdır.

3.1.1. Genetik Algoritmalar

Genetik Algoritmalar, doğal oluşum ve genetik çaprazlamadan esinlenen, köklü bir şekilde birleştirilme yöntemidir. Genetik algoritmalar, makine öğrenimi modellerinin hiperparametrelerini optimize etmek için yaygın olarak kullanılır. Algoritma, günün çözümlerinden oluşan bir baştanu başlatıyor ve bu bölümde her bireyin belirli bir çözümünü temsil ediyor. Uygunluk fonksiyonu sayesinde çözümlerin kalitesi kayıtlı ve en iyi çözümler, çaprazlama ve değiştirilebilen gibi işlemlerle yeni nesil çözümler oluşturmak için seçilir. Bu süreçte, hiperparametrelerin optimum düzeydekilerinin ayrı ayrı olmasını sağlar ve model miktarının artırılmasını sağlar (Vanneschi, 2023).

Örneğin, SVM algoritması için "C" ve "gamma" hiperparametrelerinin oluşumunda Genetik Algoritmalar mevcuttur. Bu sayede SVM algoritmasının performansı, optimize edilmiş hiperparametrelerle daha üst düzeyde taşınabilir.

3.1.2. Parçacık Sürü Optimizasyonu

Parçacık Sürü Optimizasyonu, kuş sürüleri ve balık sürülerinin toplu hareketlerinden ilham alınarak, merkezi tabanlı bir görünüm geliştirilmesidir. Parçacık Sürü Optimizasyonu'nda, her bireyin (parçacık) potansiyel bir çözüm temsil eder ve parçaları, en iyi bilinen çözüm (küresel en iyi) ve kendi en iyi çözümleri (yerel en iyi) çerçevesinde hareket ederler. Parçacıklar, kendi hızlarını ve pozisyonlarını güncelleyerek, daha iyi çözümler ararlar. PSO, hiperparametre ayarları için özellikle hızlı ve etkin bir yöntem sunar (Kennedy ve Eberhart, 1995).

Örneğin, Random Forest modelinin ağaç sayısı ve maksimum derinlik gibi hiperparametrelerini optimize etmek için PSO kullanılabilir. Bu, miktarın arttırılmasına yardımcı olabilir.

3.2. Özellik Seçimi

Meta sezgisel algoritmalar, yüksek boyutlu veri setlerindeki önemli özelliklerin seçilmesinde etkili olabilir. Özellik seçimi süreci, makine öğrenimi modellerinin performansını artırabilir ve daha iyi genelleştirme sağlanabilir. Özellik seçimi için kullanılan yöntemler arasında Genetik Algoritmalar ve Karınca Kolonisi Optimizasyonu gibi yaklaşımlar yer alır (Vanneschi, 2023).

3.2.1. Genetik Algoritmalar ile Özellik Seçimi

Genetik Algoritmalar, yüksek boyutlu veri setlerinde önemli özelliklerin seçilmesi için etkili bir yöntem. Özel seçim, her bireyin, belirli bir özelliği kümesini temsil eder. Genetik algoritmaların evrimsel süreci sayesinde, gereksiz veya etkisiz özellikler elenirken, modelin arttırılmasının en önemli özellikleri belirlenmiştir. Bu, makine öğrenimi modellerinin daha hızlı çalışmaları ve daha iyi genelleştirme yapılmasına olanak tanır.

3.2.2. Karınca Kolonisi Optimizasyonu

Karınca Kolonisi Optimizasyonu, karıncaların yiyecek bulma ve iz bırakma davranışlarından kaynaklanan bir programdır. Bu optimizasyon, özel seçim problemlerinde uygulanabilir. Her yönetici, potansiyel bir özellik alt ayarını seçer ve bu seçimin etkinliği, ilgili modelin performansıyla birlikte kullanılır. Karıncalar, daha başarılı yollar üzerinde daha fazla feromon (kimyasal iz) bırakarak diğer karıncaların bu yolları izlemesini sağlar. Bu sayede önemli özellikler daha belirgin hale gelir ve model performansı artırılır (Dorigo ve Stützle, 2019).

3.3. Veri Üretimi ve Önyükleme

Meta parçalanma kısıtlamaları, sınırlı veri setleriyle çalıştırılıp veri çeşitliliğini artırmak için kullanılabilir. Bu engelleme, yeniden düzenleme ve sentetik veri oluşturma işlemleri ile veri ayarlamanın hızını artırma ve modelin aşırı öğrenmesini tüketmeye yardımcı olur. Örneğin, PSO ile veri noktalarının belirli bir dağılıma göre yeniden örneklenmesi sağlanarak veri seti zenginleştirilebilir. Bu durum da, makine öğrenimi modelinin öğrenmesini artırarak daha doğru tahminler yapılmasını sağlar.

3.4. Hibrit Modelleme

Makine öğrenimi ve meta sezgisel yöntemlerin birleştirilmesi, hibrit modelleme stratejileri oluşturulmasına olanak tanır. Örneğin, bir sınıflandırma problemi için makine öğrenimi modeli oluşturulurken, meta sezgisel yöntemler modelin ağırlıklarını veya parametrelerini optimize edebilir. Bu yaklaşım, modelin performansını artırmak ve daha tutarlı sonuçlar elde etmek için kullanılabilir.

4. YÖNTEM

Bu çalışmada, Python kütüphaneleri kullanılarak oluşturulan sentetik hasta verileri veri seti olarak tercih edilmiştir (Yare, 2024). İlk olarak, veri setinde bulunan dengesizlik, Smote (Fernández ve ark., 2018) algoritması yardımıyla azınlık sınıfa ait verilerin sentetik olarak çoğaltılması ile giderilmiştir.

Özniteliklere ait açıklamalar aşağıda olduğu gibidir:

Name (İsim): Rastgele oluşturulmuş birey isimleri.

Gender (Cinsiyet): Bireylerin cinsiyeti (Erkek/Kadın).

Age (Yaş): Bireylerin yaşları (18-80 yaş).

Systolic BP: Sistolik kan basıncı.

Diastolic BP: Diyastolik kan basıncı.

Cholesterol (Kolesterol): Kolesterol seviyeleri.

Height (Boy (cm)): Bireylerin boyları santimetre cinsindendir.

Weight (Ağırlık (kg)): Bireylerin kilogram cinsinden ağırlığı.

BMI: Boy ve kilo üzerinden hesaplanan Vücut/Beden Kitle İndeksi.

Smoker: Sigara içme durumu (Doğru/Yanlış).

Diabetes: Diyabet durumu (Doğru/Yanlış).

Health: Birleştirilmiş ölçütlere (İyi/Orta/Kötü) dayalı genel sağlık değerlendirmesi.

Öznitelikler, üçü kategorik ve yedisi sayısal olmak üzere sınıflandırılmıştır. Kategorik öznitelikler (smoker, diabetes, health) için Gaussian mekanizmasında epsilon değeri 0,0075 olarak belirlenmişken, Laplace mekanizması için bu uygulanmamıştır. Sayısal öznitelikler (age, systolic BP, diastolic BP, cholesterol, height, weight, BMI) için ise epsilon değeri 1,00 olarak atanmıştır. Sonrasında, öznitelikler için uygun sınırlar tanımlanarak Gauss ve Laplace mekanizmaları kullanılarak diferansiyel mahremiyet uygulanmıştır.

Daha sonra, her özniteliğe meta sezgisel yöntemlerden olan DEAP (Distributed Evolutionary Algorithm in Python - Python'da Dağıtılmış Evrimsel Algoritmalar)

genetik algoritması, sürekli olan özniteliklere *"epsilon_cont"* ve kategorik olan özniteliklere ise *"epsilon_cat"* algoritmaları uygulanmıştır.

Diferansiyel gizlilik bağlamında uygulanan bu epsilon parametleri, sırasıyla sürekli ve kategorik verilere eklenen gürültüyü kontrol ederek gizlilik düzeyini belirler. Epsilon, daha düşük değerlerle daha yüksek gizlilik sağlar ancak veri doğruluğunu azaltır. Laplace ve Gaussian gibi çeşitli mekanizmalarda epsilon için optimum değerleri seçmek, hassas veri işlemede gizlilik ve doğruluk arasında bir denge sağlamada çok önemlidir.

DEAP algoritması, sistemi optimize etmeyi ve evrimsel algoritmalar oluşturulmayı sağlar. Yöntemin modüler tasarımı, kullanıcıların algoritmaları uyarlamalarına; seçim, çaprazlama ve mutasyon gibi genetik yöntemler arasında kolayca geçiş yapmalarına olanak sağlar. Bu uyarlanabilirlik, özellikle çok amaçlı optimizasyon senaryolarında, hızlı deney ve prototipleme için faydalıdır. DEAP algoritması mimarisi nedeniyle kendisini diğer kara kutu uygulamalarından ayırır çünkü şeffaflığı ve esnekliği teşvik eder. Bu durum da, DEAP algoritmasını eğitim ve araştırma amaçları için özellikle uygun hale getirir (Fortin ve ark., 2012).

Tüm algoritmalar uygulandıktan sonra alınan çıktıda; sürekli (1,1858581064705151) ve kategorik (2,4068144689152127) veriler için ayrı olacak şekilde en ideal epsilon değerleri, nesiller ve her nesilde yapılan değerlendirme sayıları tespit edilerek diferansiyel mahremiyet ve genetik algoritmaların uygulandığı yeni bir veri seti elde edilmiştir. Ayrıca, elde edilen yeni veri seti ve orijinal veri seti arasında Ortalama Kare Hata (Mean Squared Error - MSE) yöntemi ile ortalama ve varyans değerlerinin karşılaştırma sonuçlarına ulaşılmıştır. Bu yöntem, bir modelin tahmin yeteneğini değerlendirmek için kullanılan bir performans metriğidir. Modelin tahmin ettiği değer ile gerçek değer arasındaki farkların karelerinin ortalamasını hesaplar (Shakeel ve ark., 2020). Değer sıfıra yaklaştıkça, modelin tahminlerinin gerçek değerlere o kadar yakın olduğu, yani modelin performansının daha iyi olduğu söylenebilir (Dursun ve Toraman, 2021).

Elde edilen yeni veri setine, sınıflandırma algoritmaları kullanılarak analiz gerçekleştirilmiştir. Bu veri seti üzerinde Logistic Regression, Random Forest, Gradient Boosting, XGBoost, SVM, KNN, Neural Network ve Adaboost algoritmaları uygulanarak, doğruluk (accuracy), F1 skor ve karışıklık matrisi (confusion matrix)

değerleri hesaplanmıştır. Logistic Regression algoritması, doğrusal bir karar sınırı oluştururken; Random Forest ve Gradient Boosting, birden fazla karar ağacını bir araya getirerek daha karmaşık ilişkileri modelleyebilir. SVM, verileri en iyi ayıracak hiper düzlemi bulmaya çalışırken, KNN algoritması ise gözlemleri en yakın komşularının sınıflarına göre tahmin eder. Neural Network, çok katmanlı yapılar aracılığıyla insan beyninin nöronlarının çalışma prensibinden esinlenerek öğrenme sağlar. Adaboost, zayıf öğrenicileri birleştirerek daha güçlü bir sınıflandırıcı oluştururken, XGBoost ise Gradient Boosting'in daha hızlı ve ölçeklenebilir bir versiyonunu sunar. Bu algoritmaların seçiminde veri setinin yapısı, problemin karmaşıklığı ve mevcut hesaplama kaynakları gibi faktörler etkili olmuştur (Hastie ve ark., 2009).

Sonrasında entegre sistemin, verinin %70'i ile kendini eğitmesi ve %30'u ile de test etmesi sağlanmıştır. Bu oranlar ile aşırı öğrenme ya da eksik öğrenme durumlarının önüne geçilmiştir. Daha sonra Principal Component Analysis - PCA (Kurita, 2019) algoritması uygulanarak verinin %95'ini açıklayan öznitelikler seçilerek sonuç 7 öznitelik olarak bulunmuştur.

Sonuç değerlendirmesinin yapılabilmesi maksadıyla doğruluk, F1 skorları ve karşıklık matrisi hesaplaması (Huang ve Ling, 2005) yapılmış, ayrıca karışıklık matrisi görselleştirilmiştir.

5. BULGULAR

Diferansiyel mahremiyet ve meta sezgisel yöntemlerin entegrasyonu sonucu elde edilen yeni veriseti üzerinde uygulanan sınıflandırma algoritmalarına ait sonuçlar Tablo 5.1'de gösterilmiştir. Buna göre elde edilen yeni verisetinde en iyi çalışan algoritma XGBoost (0,9463 / 0,9495) olmuştur. Değerlendirme yapılırken doğruluk (accuracy) ve F1 skorları baz alınmıştır.

Tablo 5.1 Sınıflandırma algoritmalarina ait sonuçlar

Model	**Orijinal Veri Seti**	**Gaussian Veri Seti**	**Laplace Veri Seti**	**Entegre Veri Seti**
	Doğruluk / F1 Skor	***Doğruluk / F1 Skor***	***Doğruluk / F1 Skor***	***Doğruluk / F1 Skor***
Logistic Regression	0,7634 / 0,7626	0,6628 / 0,6564	0,7617 / 0,7617	0,7483 / 0,7486
Random Forest	0,9430 / 0,9435	**0,8003 / 0,7988**	0,9513 / 0,9519	0,9430 / 0,9434
Gradient Boosting	0,9480 / 0,9482	0,7987 / 0,7973	0,9312 / 0,9320	0,9362 / 0,9367
XGBoost	**0,9631 / 0,9632**	0,7953 / 0,7920	**0,9614 / 0,9615**	**0,9463 / 0,9465**
SVM	0,8893 / 0,8886	0,7617 / 0,7567	0,8859 / 0,8863	0,8926 / 0,8925
KNN	0,7416 / 0,7171	0,6560 / 0,6394	0,6879 / 0,6656	0,6779 / 0,6507
Neural Network	0,9128 / 0,9124	0,7668 / 0,7615	0,9211 / 0,9218	0,9295 / 0,9297
Adaboost	0,8389 / 0,8402	0,7483 / 0,7496	0,8087 / 0,8103	0,8339 / 0,8337

Aynı zamanda veri setinin ortalama ve varyans değerleri hesaplanmış ve çıkan ortalama sonuçları Tablo 5.2'de, varyans sonuçları ise Tablo 5.3'te gösterilmiştir.

Tablo 5.2 Veri setlerinin ortalama sonuçları

Öznitelikler	Orijinal Veri Seti	Gaussian Veri Seti	Laplace Veri Seti	Entegre Veri Seti
Age	46,5353	46,5392	46,5350	46,5362
Systolic BP	127,8479	127,8409	127,8501	127,8472
Diastolic BP	81,7261	81,7252	81,7265	81,7261
Cholesterol	187,1783	187,1657	187,1754	187,1789
Height (cm)	175,6017	175,6028	175,6025	175,6022
Weight (kg)	75,2119	75,2077	75,2115	75,2116
BMI	24,8005	24,7991	24,7999	24,8011
Smoker	0,4657	0,4663	0,4658	0,4658
Diabetes	0,5141	0,5096	0,5141	0,5141
Health	1,0000	0,9980	0,9995	1,0000

Tablo 5.3 Veri setlerinin varyans sonuçları

Öznitelikler	Orijinal Veri Seti	Gaussian Veri Seti	Laplace Veri Seti	Entegre Veri Seti
Age	311,4942	311,5444	311,4980	311,5005
Systolic BP	94,7114	94,8573	94,7759	94,7179
Diastolic BP	39,5280	39,5084	39,5030	39,5258
Cholesterol	425,9884	426,5123	426,1431	425,9864
Height (cm)	186,2249	186,2018	186,2598	186,2229
Weight (kg)	185,2827	185,6613	185,3117	185,2988
BMI	33,2238	33,2304	33,2177	33,2229
Smoker	0,2489	0,2490	0,2490	0,2490
Diabetes	0,2499	0,2500	0,2499	0,2499
Health	0,6670	0,6730	0,6665	0,6670

Sürekli değişkenler (Age, Systolic BP, Diastolic BP, Cholesterol, Height, Weight, BMI) üzerinde yapılan analizlerde, Gaussian mekanizmasının, verilerin doğal dağılımından daha geniş bir yayılıma neden olduğu gözlemlenmiştir. Bu durum, Gaussian mekanizmasının, veriye ek bir gürültü katmasıyla ilişkilendirilebilir.

Özellikle, delta parametresinin eklenmesiyle bu durum daha da belirginleşir. Buna karşın, Laplace mekanizması, bu değişkenlerde daha kararlı bir dağılım sağlayarak, orijinal veriye daha yakın sonuçlar üretmiştir. Gaussian mekanizmasının, uç değerlerde daha fazla gözlem oluşturduğu ve bu nedenle verinin genel yapısını bozduğu söylenebilir. Laplace mekanizması ise, bu tür uç değerleri daha az vurgulayarak, verinin orijinal yapısını daha iyi korur.

Sigara kullanımı, diyabet gibi kategorik değişkenlerde ise, Laplace mekanizmasının, orijinal verinin sınıflandırma yapısını daha iyi koruduğu görülmüştür. Bu durum, Laplace mekanizmasının, kategorik veriler üzerindeki etkilerinin daha az olduğu şeklinde yorumlanabilir. Gaussian mekanizması ise, bu tür verilerde de belirgin bir gürültüye neden olmuş ve sınıflandırma hatalarına yol açmıştır. Sonuç olarak, sürekli değişkenlerde Laplace mekanizmasının, veri gizliliğini korurken orijinal verinin yapısını daha iyi koruduğu söylenebilir. Kategorik değişkenlerde ise, her iki mekanizmanın da farklı performanslar gösterdiği, ancak Laplace mekanizmasının genellikle daha başarılı olduğu görülmüştür.

Karışıklık Matrisi (Confusion Matrix), özellikle sınıflandırma problemlerinde bir modelin performansını değerlendirmek için kullanılır. Bu matris, bir modelin yaptığı tahminlerin gerçek sınıflara göre nasıl dağıldığını gösterir.

True Positive (TP): Modelin pozitif olarak sınıflandırdığı ve gerçekte de pozitif olan örneklerdir. Örneğin, bir kanser teşhis modelinde, gerçekten kanseri olan bir hastayı kanserli olarak teşhis etmek TP durumudur.

True Negative (TN): Modelin negatif olarak sınıflandırdığı ve gerçekte de negatif olan örneklerdir. Aynı kanser örneğinde, sağlıklı bir kişiyi sağlıklı olarak teşhis etmek TN durumudur.

False Positive (FP): Modelin pozitif olarak sınıflandırdığı ancak gerçekte negatif olan örneklerdir. Yani, sağlıklı bir kişiyi yanlışlıkla kanserli olarak teşhis etmek FP durumudur. Bu duruma, "yanlış alarm" da denebilir.

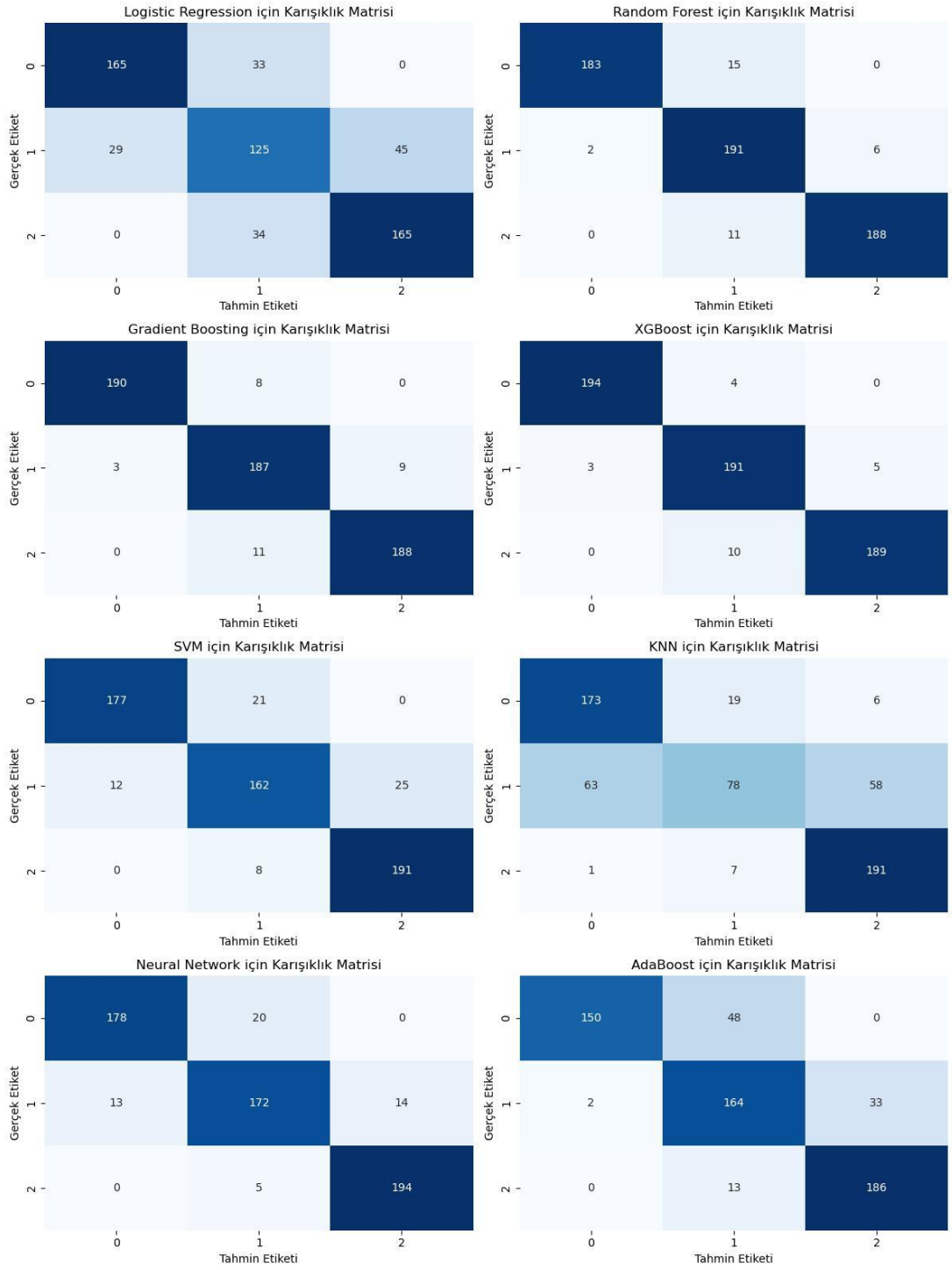
False Negative (FN): Modelin negatif olarak sınıflandırdığı ancak gerçekte pozitif olan örneklerdir. Yani, kanserli bir hastayı yanlışlıkla sağlıklı olarak teşhis etmek FN durumudur. Bu durum, genellikle daha ciddi sonuçlara yol açabilir (Lisa ve ark., 2011; Liang, 2022).

Sınıflandırma algoritmalarının oluşturduğu 3x3 karışıklık matrisleri Şekil 5.1'de, karışıklık matrisi açıklama tablosu ise Tablo 5.4'te gösterilmiştir.

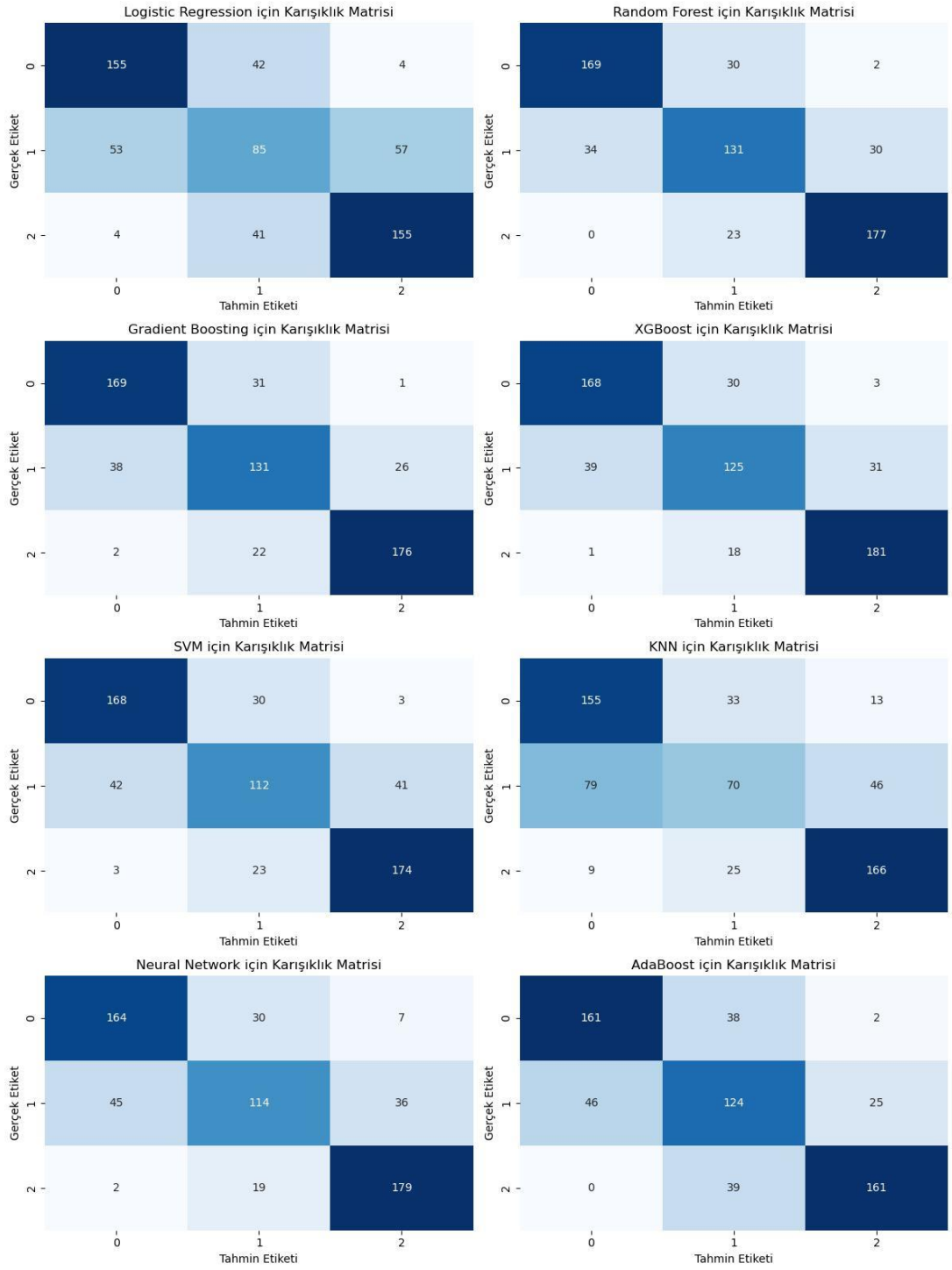
Tablo 5.4 Karışıklık Matrisi Açıklama Tablosu

Gerçek Etiket	0	TP	FN	FN
	1	FP	TP	FN
	2	FP	FN	TP
		0	1	2
		Tahmin Etiket		

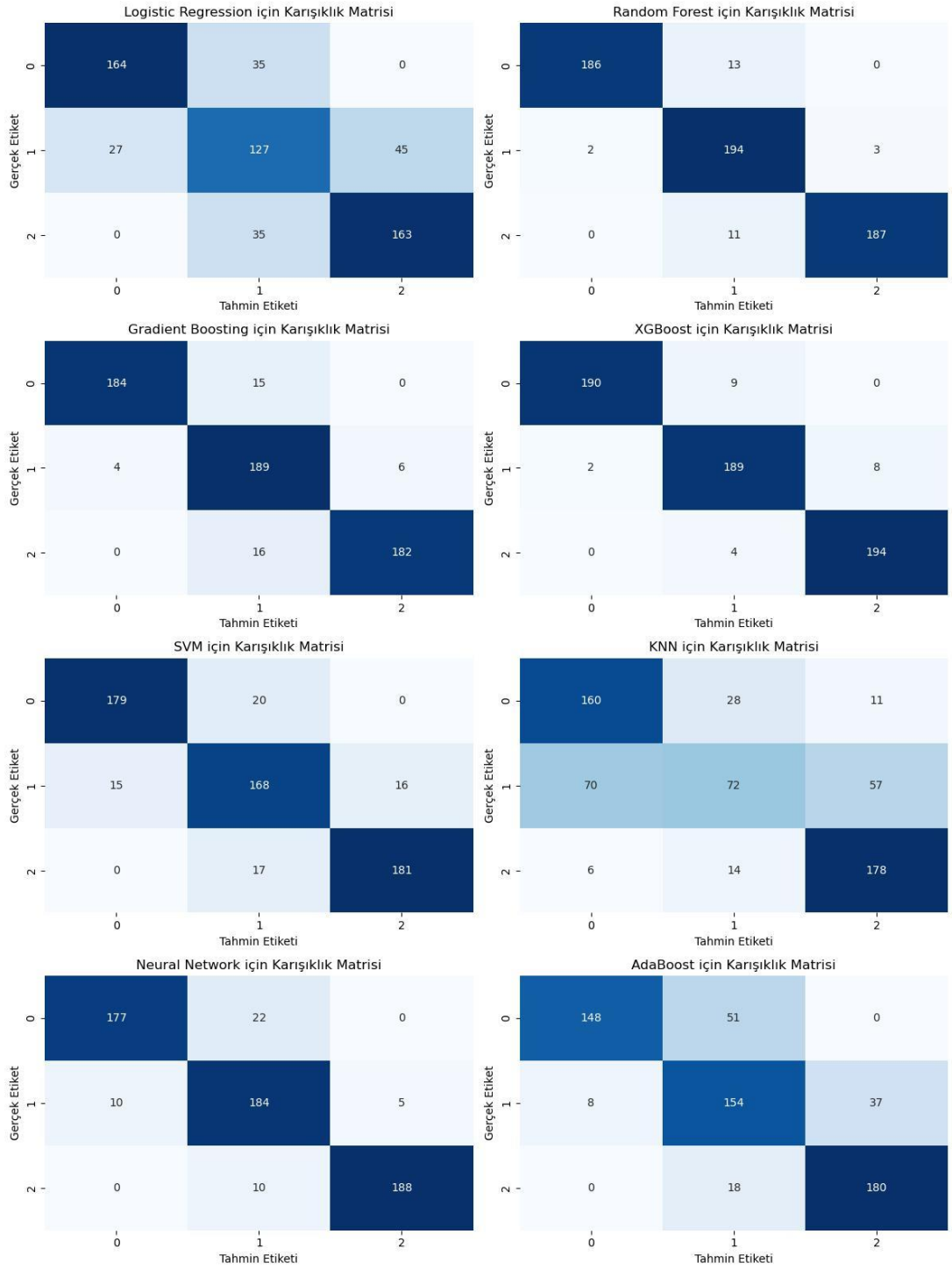
Şekil 5.1'de original veri seti için, Şekil 5.2'de Gaussian veri seti için, Şekil 5.3'te Laplace veri seti için ve Şekil 5.4'te de Entegre veri seti için gösterilen karışıklık matrisleri (confusion matrices) incelendiğinde, sınıflandırma algoritmalarının başarısı görsel olarak kıyaslanabilmektedir. Buna göre en başarılı algoritmalar XGBoost, Random Forest ve Gradient Boosting olmuştur.



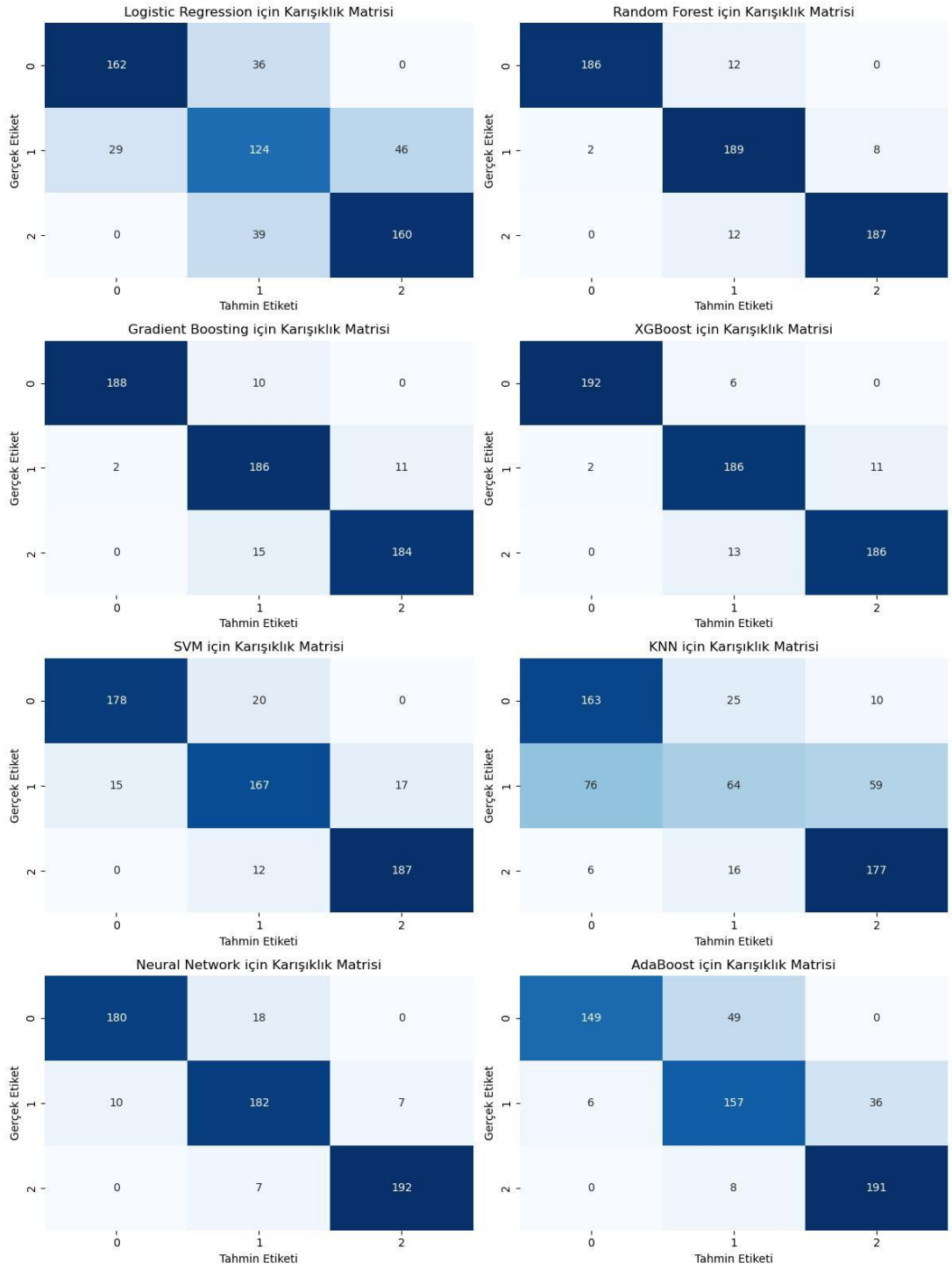
Şekil 5.1 Orijinal Veri Seti Karışıklık Matrisi



Şekil 5.2 Gaussian Veri Seti Karışıklık Matrisi



Şekil 5.3 Laplace Veri Seti Karışıklık Matrisi



Şekil 5.4 Entegre Veri Seti Karışıklık Matrisi

Orijinal veri setinin özniteliklerine ait değerler; en düşük, en yüksek, ortalama ve standart sapma olarak Tablo 5.5'te verilmiştir. Alt limit, üst limit ve kabul edilebilir hata payları bu değerler ışığında dikkate alınmıştır.

Tablo 5.5 Orijinal veri seti öznitelik değerleri

Öznitelik	En Düşük Değer	En Yüksek Değer	Ortalama Değer	Standart Sapma
Age	18	80	46,5353	17,6492
Systolic BP	105	150	127,8479	9,73197
Diastolic BP	65	95	81,7261	6,2871
Cholesterol	140	239	187,1783	20,6395
Height	150,0876	199,9663	175,6016	13,6464
Weight	50,0119	99,9549	75,2119	13,6119
BMI	13,0947	43,8067	24,8005	5,7640

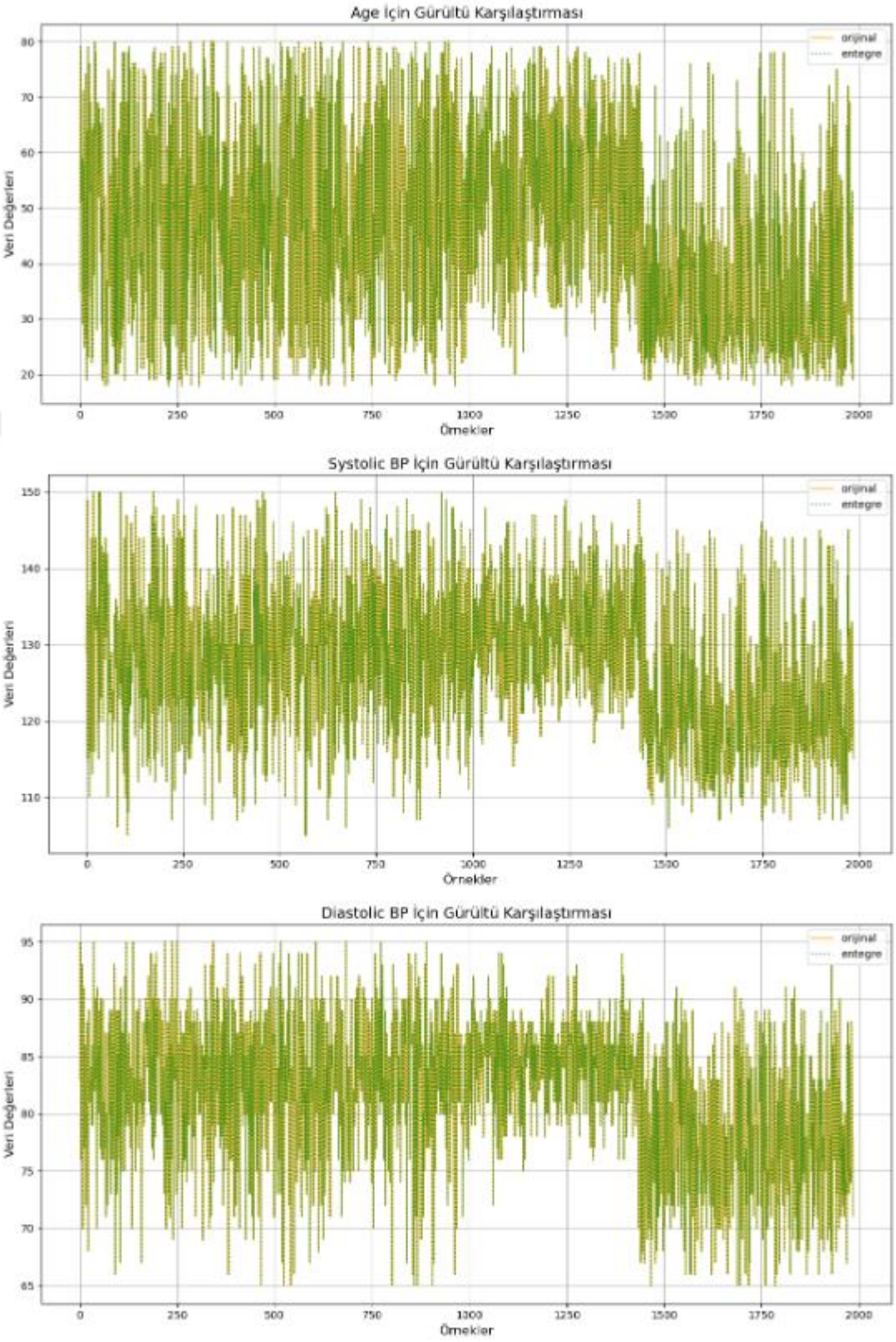
Aynı şekilde entegre veri setinin özniteliklerine ait değerler; en düşük, en yüksek, ortalama ve standart sapma olarak Tablo 5.6'da verilmiştir.

Tablo 5.6 Entegre veri seti öznitelik değerleri

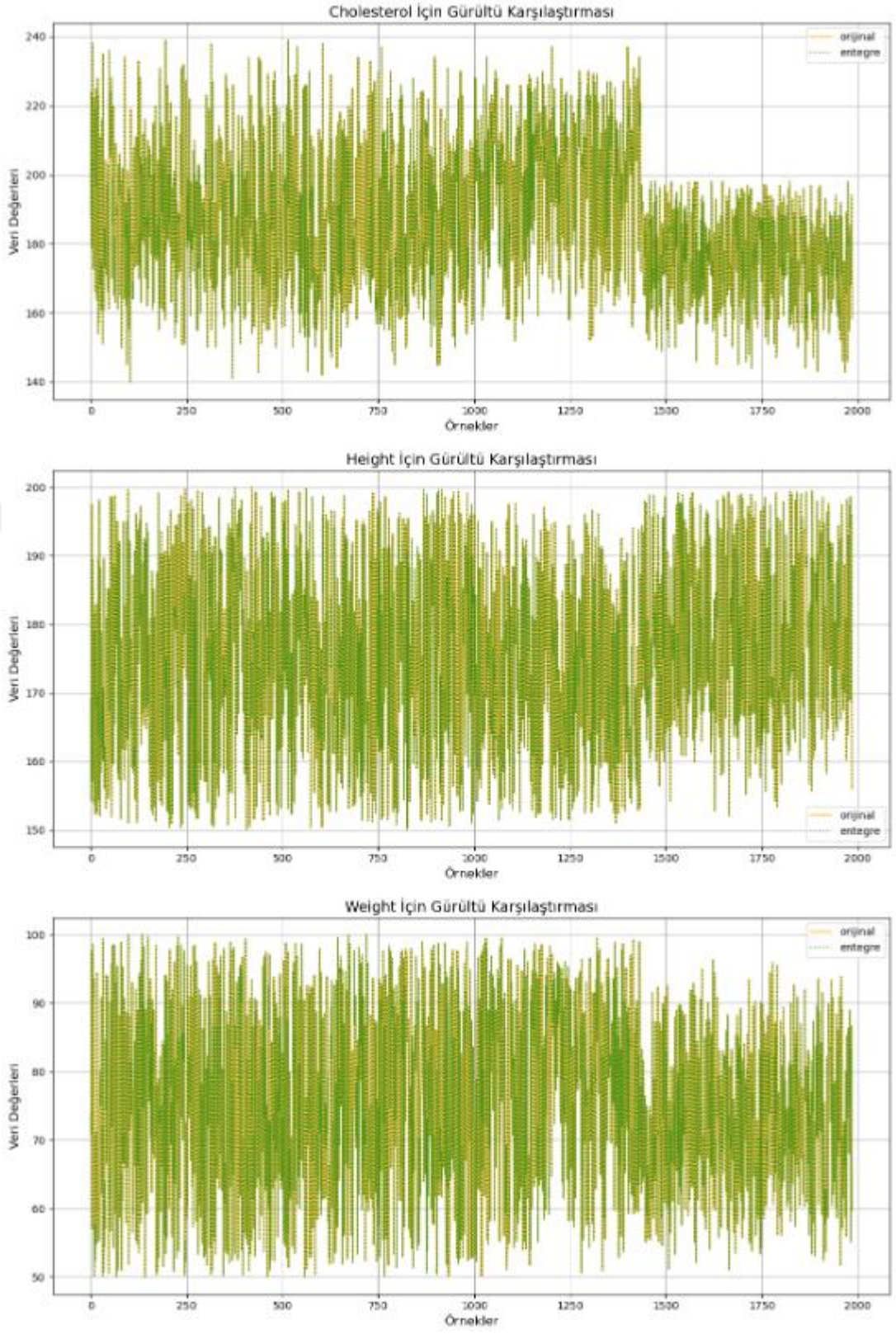
Öznitelik	En Düşük Değer	En Yüksek Değer	Ortalama Değer	Standart Sapma
Age	17,9346	80,0707	46,5362	17,6494
Systolic BP	104,9943	150,0288	127,8472	9,7323
Diastolic BP	64,9613	95,0347	81,7261	6,2870
Cholesterol	140,0456	238,9684	187,1785	20,6394
Height	150,0682	199,9811	175,6023	13,6463
Weight	50,0123	99,9417	75,2116	13,6125
BMI	13,0925	43,8053	5,7639	24,8011

Şekil 5.5'te "Age", "Systolic BP" ve "Diastolic BP" öznitelikleri için, Şekil 5.6'da "Cholosterol", "Height" ve "Weight" öznitelikleri için, Şekil 5.7'de ise "BMI"

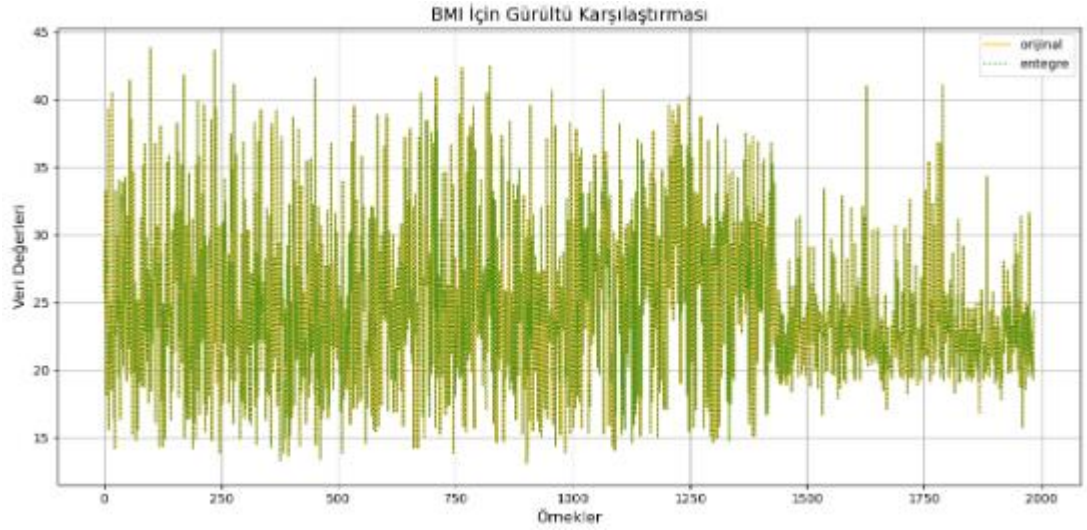
özniteliği için gürültü karşılaştırılması yapılmıştır. Buna göre; Tablo 5.5'te yer alan öznitelik değerlerinin kontrolü görsel olarak da sağlanmıştır.



Şekil 5.5 Age, Systolic BP ve Diastolic BP İçin Gürültü Karşılaştırması



Şekil 5.6 Cholosterol, Height ve Weight İçin Gürültü Karşılaştırması

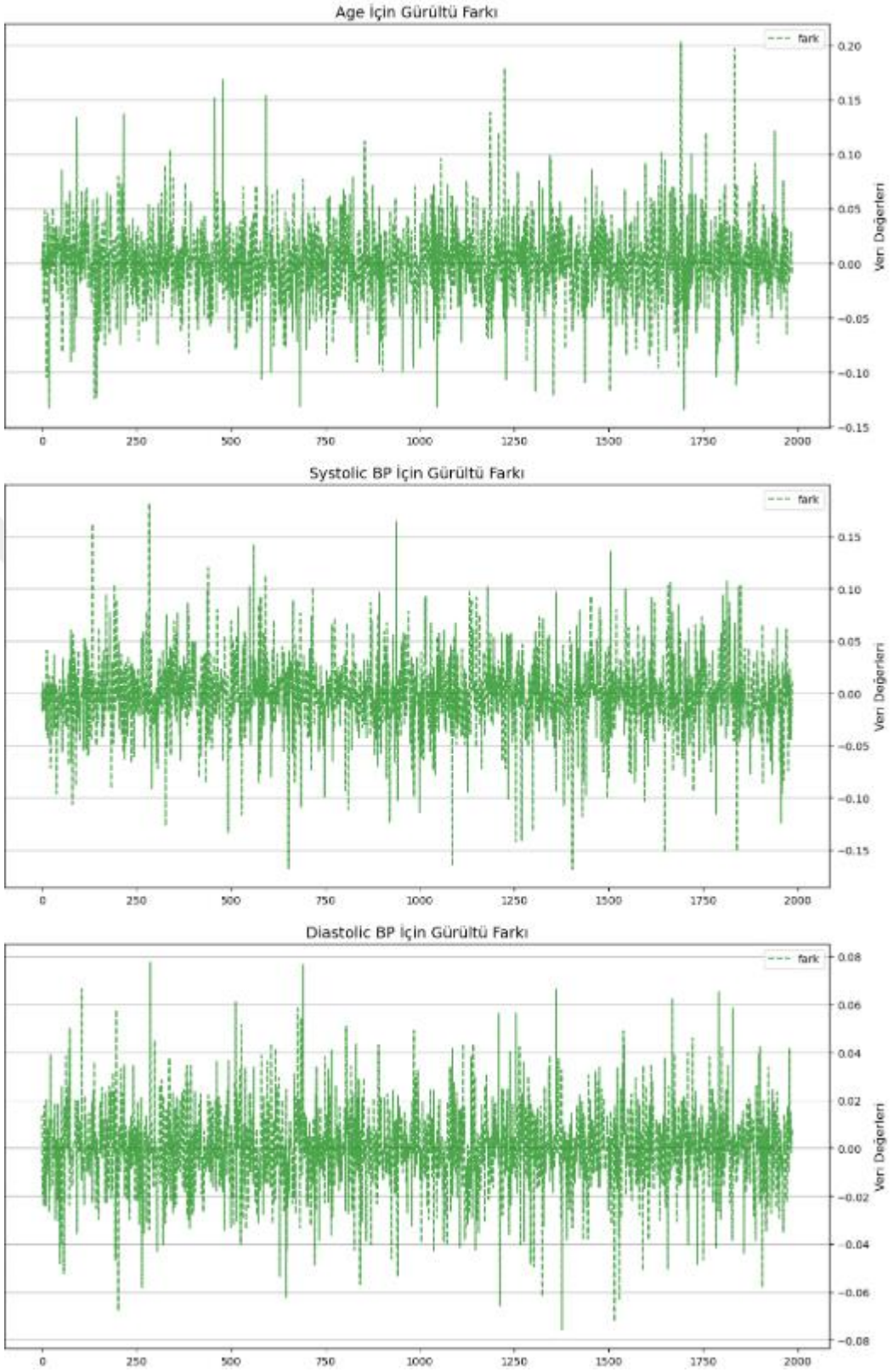


Şekil 5.7 BMI İçin Gürültü Karşılaştırması

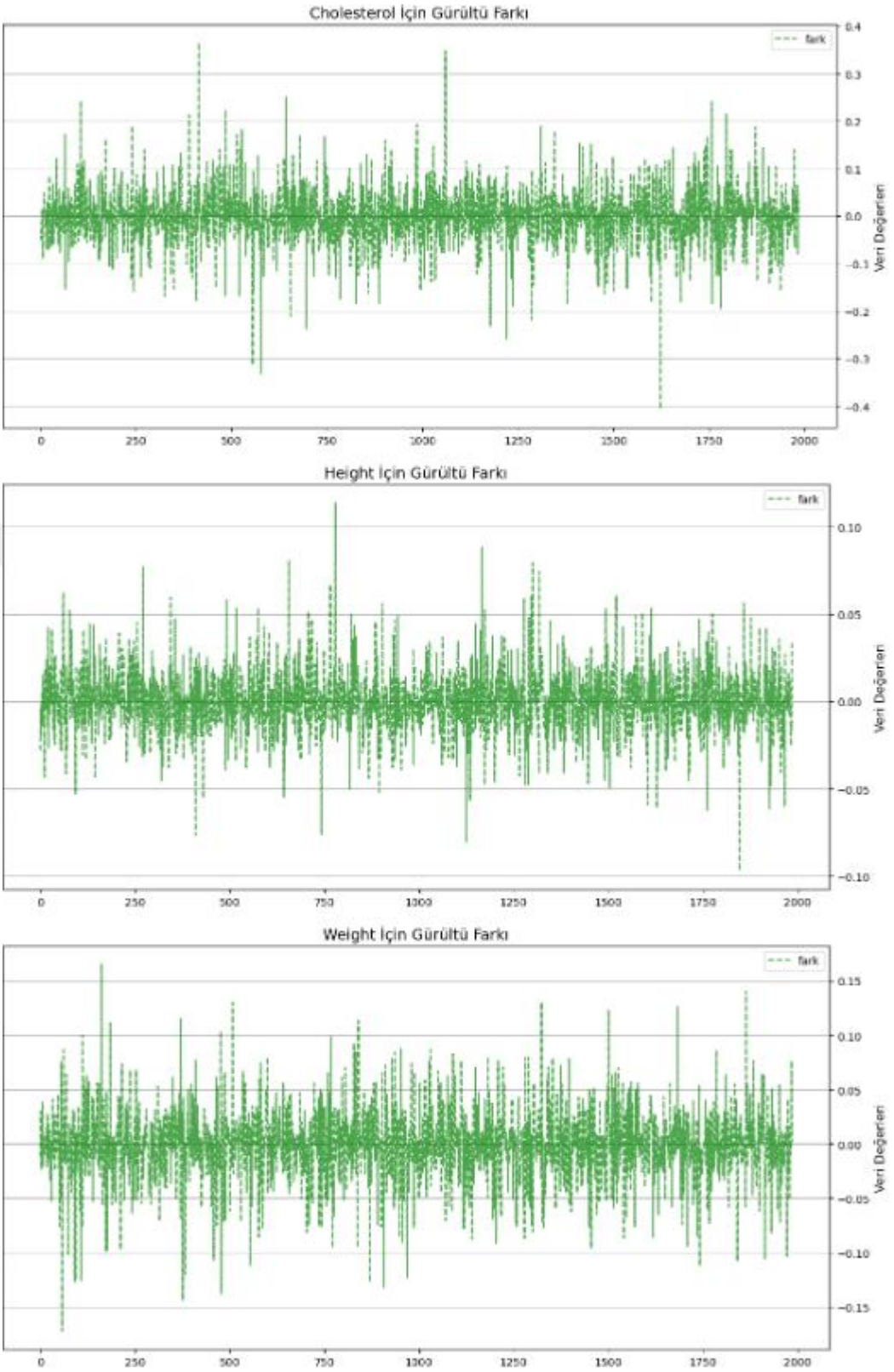
Öznitelikler için gürültü farklarına bakıldığında ise Şekil 5.8'de "Age", "Systolic BP" ve "Diastolic BP" öznitelikleri için, Şekil 5.9'da "Cholosterol", "Height" ve "Weight" öznitelikleri için, Şekil 5.10'da ise "BMI" özniteliği için gürültü farkları görülmektedir. "Age" özniteliğinde düşük düzeyde gürültü farkının, diğer özniteliklere kıyasla daha farklı bir şekilde dağıldığı, benzer şekilde, "Systolic BP" ve "Diastolic BP" özniteliklerinde gürültünün veri üzerindeki etkisinin daha geniş bir dağılım gösterdiği ve bazı örneklerde daha belirgin hale geldiği gözlemlenmiştir.

"Cholosterol" ve "Height" özniteliklerinde gürültünün veri üzerindeki yayılma dağılımı daha homojen olduğu, özellikle ortalama değerlerin okulda yoğunlaştığı dikkat çekmektedir. "Weight" özniteliği ise diğer özniteliklere kıyasla daha değişken bir yapıya sahip olup, bazı bireylerde gürültü farklılıklarının önemli ölçüde görüldüğü tespit edilmiştir.

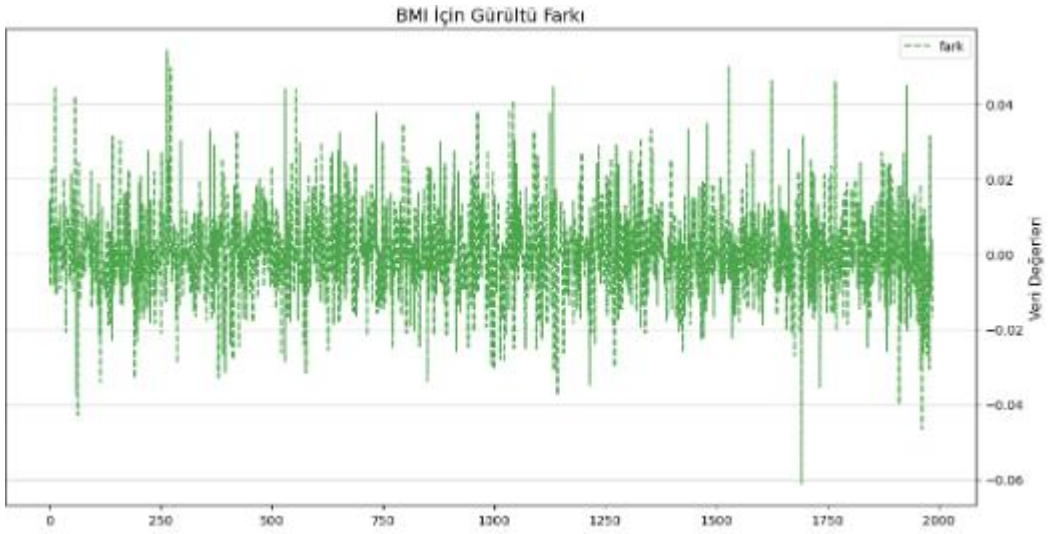
Son olarak, Şekil 5.10'da yer alan "BMI" özniteliği, özellikle diğer özniteliklere kıyasla daha sınırlı bir gürültü farkı aralığına sahiptir. Bu durum, söz konusu özniteliğin daha az değişken bir yapıya sahip olduğu ve gürültünün istatistiki özelliğinin bozulmadan tutulduğunu göstermektedir. Genel olarak, veri setindeki gürültü varyasyonları hem genel popülasyon düzeyinde hem de örnek bazında detaylı analizlere imkân tanıyarak, farklı öznitelikler üzerinde gürültünün etkisini kapsamlı bir şekilde değerlendirmeyi mümkün hale getirmektedir.



Şekil 5.8 Age, Systolic BP ve Diastolic BP İçin Gürültü Farkı



Şekil 5.9 Cholosterol, Height ve Weight İçin Gürültü Farkı



Şekil 5.10 BMI İçin Gürültü Farkı

Gürültü miktarlarının kıyaslanabilmesi amacıyla, orijinal veri setinden baştan, ortadan ve sondan seçilen toplam 30 örneğe ait değerler Tablo 5.7'de sunulmuştur. Aynı örneklere ait değerler, entegre veri seti üzerinde de incelenmiş olup, bu değerler Tablo 5.8'de yer almaktadır. Ayrıca, gürültü kıyaslamasının yapılabilmesi için, aynı örnekler üzerinde uygulanan gürültü miktarları Tablo 5.9'da detaylandırılmıştır. Bu kapsamda, farklı veri kümelerindeki örnekler üzerinde yapılan analizler, eklenen gürültü miktarlarının etkilerini daha ayrıntılı bir şekilde değerlendirme imkânı sağlamaktadır.

Tablo 5.7 Orijinal veri setine ait değerler

Age	Systolic BP	Diastolic BP	Cholesterol	Height (cm)	Weight (kg)	BMI
35	129	83	188	169,0721	73,7962	25,8161
70	133	95	224	154,1886	70,6715	29,7262
79	149	88	179	197,5149	97,5903	25,0154
52	115	76	204	157,2238	82,2028	33,2545
51	135	82	201	177,3703	57,0387	18,1304
59	134	93	238	189,5467	98,6473	27,4570
54	125	89	173	183,3315	93,7442	27,8914
29	110	70	174	182,1100	90,6918	27,3464
71	134	81	222	154,0144	93,3892	39,3708
35	131	84	206	179,0512	50,0612	15,6152
44	121	80	207	178,8273	77,2632	24,1605
23	111	82	190	179,0723	94,8199	29,5694
34	140	82	203	167,1566	86,7857	31,0600
59	133	87	210	196,0797	66,5948	17,3211
45	118	84	184	171,7193	52,7593	17,8921
44	133	77	185	197,4735	76,8996	19,7200
65	145	83	187	150,9623	87,4043	38,3527
33	119	79	157	189,9077	55,1996	15,3056
57	127	88	208	183,0415	52,7753	15,7518
38	124	84	158	185,2739	65,6736	19,1321
69	132	84	182	190,6384	76,6389	21,0212
35	126	74	178	168,9209	68,0748	23,9377
44	118	75	155	194,1066	76,2751	20,4386
30	118	74	179	197,4237	89,0341	22,8429
22	116	76	185	194,2933	74,9084	19,8504
49	133	88	168	198,6854	83,3433	21,1219
47	126	83	176	168,6987	55,1661	19,3665
53	118	78	188	196,1395	86,6539	22,5229
19	124	78	194	167,6434	68,2287	24,2683
27	115	71	158	155,9944	59,7053	24,3923

Tablo 5.8 Entegre veri setine ait değerler

Age	Systolic BP	Diastolic BP	Cholesterol	Height (cm)	Weight (kg)	BMI
35,0050	129,0000	83,0138	187,9586	169,0441	73,7845	25,8178
69,9937	132,9826	94,9884	223,9513	154,1849	70,7016	29,7403
79,0183	148,9931	87,9807	178,9850	197,4956	97,5818	25,0092
52,0145	115,0100	75,9856	203,9900	157,2238	82,1798	33,2460
51,0186	134,9822	81,9899	201,0148	177,3664	57,0425	18,1389
58,9620	133,9891	92,9760	237,9415	189,5522	98,6280	27,4793
54,0027	125,0001	89,0176	172,9116	183,3178	93,7823	27,9001
28,9881	110,0030	69,9990	174,0059	182,1251	90,6714	27,3451
71,0481	133,9835	80,9811	222,0075	154,0095	93,3970	39,3778
34,9882	131,0070	83,9949	206,0485	179,0490	50,0598	15,6170
43,9967	120,9768	80,0050	206,9475	178,8298	77,3296	24,1462
22,9957	110,9942	82,0295	190,0647	179,0927	94,8029	29,5692
34,0000	140,0134	81,9952	202,9982	167,1541	86,7842	31,0680
59,0066	133,0344	86,9962	209,9404	196,0726	66,5948	17,3203
44,9531	117,9996	83,9932	184,0116	171,7171	52,7433	17,8892
43,9581	133,0339	77,0027	184,9338	197,4641	76,9018	19,7169
64,9835	144,9871	83,0324	186,9357	150,9649	87,4108	38,3295
32,9991	118,9976	78,9982	156,8423	189,9103	55,1739	15,2762
56,9787	126,9882	88,0223	207,9114	183,0292	52,7554	15,7506
37,9835	123,9557	84,0087	158,0558	185,2739	65,6764	19,1227
68,9961	132,0132	83,9903	181,9311	190,6552	76,6109	21,0164
34,9819	125,9753	74,0066	178,0412	168,9298	68,0626	23,9292
43,9903	117,9562	75,0416	155,0163	194,0994	76,2911	20,4385
30,0096	118,0233	74,0107	179,0123	197,4205	88,9824	22,8122
22,0041	116,0038	76,0023	185,0676	194,3046	74,9034	19,8818
49,0173	133,0068	88,0146	168,0074	198,6682	83,3338	21,1111
47,0152	125,9559	83,0069	176,0635	168,6733	55,1678	19,3588
53,0296	117,9886	77,9981	187,9203	196,1276	86,6531	22,5163
18,9894	123,9778	78,0081	193,9843	167,6273	68,3049	24,2721
26,9989	115,0111	71,0054	158,0343	156,0280	59,7070	24,3753

Tablo 5.9 Örneklere eklenen gürültü miktarları

Age	Systolic BP	Diastolic BP	Cholesterol	Height (cm)	Weight (kg)	BMI
0,0050	0,0000	0,0138	0,0414	0,0280	0,0116	0,0017
0,0063	0,0174	0,0116	0,0487	0,0037	0,0301	0,0141
0,0183	0,0069	0,0193	0,0150	0,0193	0,0085	0,0062
0,0145	0,0100	0,0144	0,0100	0,0000	0,0230	0,0085
0,0186	0,0178	0,0101	0,0148	0,0039	0,0038	0,0084
0,0380	0,0109	0,0240	0,0585	0,0055	0,0192	0,0223
0,0027	0,0001	0,0176	0,0884	0,0137	0,0381	0,0087
0,0119	0,0030	0,0010	0,0059	0,0151	0,0204	0,0013
0,0481	0,0165	0,0189	0,0075	0,0049	0,0078	0,0069
0,0118	0,0070	0,0051	0,0485	0,0022	0,0015	0,0019
0,0033	0,0232	0,0050	0,0525	0,0025	0,0664	0,0143
0,0043	0,0058	0,0295	0,0647	0,0205	0,0170	0,0003
0,0000	0,0134	0,0048	0,0018	0,0025	0,0015	0,0081
0,0066	0,0344	0,0038	0,0596	0,0072	0,0000	0,0007
0,0469	0,0004	0,0068	0,0116	0,0022	0,0161	0,0029
0,0419	0,0339	0,0027	0,0662	0,0094	0,0022	0,0031
0,0165	0,0129	0,0324	0,0643	0,0026	0,0065	0,0232
0,0009	0,0024	0,0018	0,1577	0,0026	0,0257	0,0294
0,0213	0,0118	0,0223	0,0886	0,0122	0,0198	0,0013
0,0165	0,0443	0,0087	0,0558	0,0001	0,0027	0,0094
0,0039	0,0132	0,0097	0,0689	0,0168	0,0280	0,0049
0,0181	0,0247	0,0066	0,0412	0,0089	0,0122	0,0085
0,0097	0,0438	0,0416	0,0163	0,0073	0,0160	0,0001
0,0096	0,0233	0,0107	0,0123	0,0033	0,0517	0,0307
0,0041	0,0038	0,0023	0,0676	0,0113	0,0050	0,0314
0,0173	0,0068	0,0146	0,0074	0,0172	0,0095	0,0107
0,0152	0,0441	0,0069	0,0635	0,0254	0,0017	0,0078
0,0296	0,0114	0,0019	0,0797	0,0119	0,0009	0,0066
0,0106	0,0222	0,0081	0,0157	0,0161	0,0762	0,0038
0,0011	0,0111	0,0054	0,0343	0,0335	0,0017	0,0169

6. TARTIŞMA

Bu çalışmada, diferansiyel mahremiyet ve meta sezgisel yöntemlerin entegre edilmesiyle elde edilen sonuçlar, literatürdeki mevcut yaklaşımlara kıyasla önemli avantajlar sunmaktadır. Özellikle, sürekli ve kategorik veri türleri arasında farklı gizlilik oranları ve veri özelliklerine özgü bir yaklaşım benimsenmesi, veri mahremiyeti ve analiz doğruluğu arasındaki dengeyi optimize etmiştir. Bu bulgu, literatürdeki çoğu çalışmanın tek bir mekanizmayı ayrım gözetmeden uyguladığı göz önüne alındığında, önemli bir yenilik olarak değerlendirilebilir (İnan ve Var, 2018).

Genetik algoritma kullanılarak Epsilon parametrelerinin optimize edilmesi, diferansiyel mahremiyet izninin daha etkin bir şekilde sağlanmasına olanak tanımıştır. Bu yaklaşım, geleneksel yöntemlere kıyasla hem veri mahremiyetini maksimum seviyede korumuş hem de veri analizinin doğruluğunu muhafaza etmiştir (Bastı, 2012). Bu sonuçlar, genetik algoritmaların optimizasyon problemlerinde ne kadar etkili olduğunu bir kez daha göstermektedir (Özdemir, 2017).

Çalışmanın sağlık verileri üzerinde uygulanabilirliği, bu yöntemin pratikteki değerini artırmaktadır. Sağlık verileri gibi hassas veri türleri üzerinde mahremiyet korumasının etkin bir şekilde sağlanabileceği gösterilmiştir. Bu durum, sağlık verilerinin güvenli bir şekilde işlenmesi ve analiz edilmesi için güvenilir ve kullanışlı bir çözüm sunmaktadır (Canbay ve Sağıroğlu, 2020). Ayrıca, önerilen yöntemin diğer veri türlerine de genişletilebilir esnek bir çerçeve sağlaması, bu çalışmanın genel uygulanabilirliğini ve değerini artırmaktadır (Akpusugh ve Dekpoghol, 2024).

Kuş (2022) çalışmasında, blok zinciri teknolojisi üzerine kurulu finansal uygulamalarda veri mahremiyetini korumak için diferansiyel mahremiyet yöntemlerinin kullanımını incelemektedir. Özellikle Bitcoin gibi kripto para birimlerinin kullandığı blok zinciri yapısında, işlem miktarlarına yapay gürültü ekleme veya kullanıcı grafiğini değiştirme gibi yöntemlerle bireysel verilerin gizlenebileceği gösterilmektedir. Ayrıca, akıllı sayaçlarda kullanılan blok zincirlerinde de diferansiyel mahremiyetin uygulanabilirliğini araştırarak, enerji tüketimi verilerinin gizliliğini korurken adil bir ödeme sistemi tasarlanabileceğini öne sürmektedir. Çalışma, diferansiyel mahremiyet yöntemlerinin blok zinciri tabanlı finansal uygulamalarda veri mahremiyeti sağlamada etkili olabileceğini göstermektedir.

Sonuç olarak çalışma, blok zinciri teknolojisinde veri mahremiyeti sağlamak için diferansiyel mahremiyet yöntemlerinin önemini vurgulamaktadır.

Kara ve Eyüpoğlu (2020) araştırmalarında, sağlık hizmetleri 4.0 bağlamında veri gizliliği ve güvenliği sorunlarına odaklanmışlardır. Bu durum, diferansiyel mahremiyetin sadece finansal uygulamalarda değil, aynı zamanda hassas kişisel verilerin olduğu diğer alanlarda da uygulanabileceğini göstermektedir. Makale, sağlık hizmetlerindeki veri gizliliği konusunda önemli bir boşluğu doldurmakta ve bu alanda daha fazla araştırmaya ihtiyaç olduğunu vurgulamaktadır. Ayrıca, sağlık verilerinin karmaşık yapısı ve farklı veri türleri göz önüne alındığında, diferansiyel mahremiyetin uygulanmasının zorluklarına da değinmektedir.

Atalay (2021) çalışmasında, kişisel sağlık verilerinin korunması ve depolanması konusunu ele almaktadır. Sağlık verilerinin mahremiyetinin sağlanması, hem hukuki hem de teknik açıdan incelenmiştir. Özellikle 6698 sayılı Kişisel Verileri Koruma Kanunu ve ilgili yönetmeliklerin uygulanması üzerinde durulmuştur. Çalışma, sağlık verilerinin kriptolu şekilde korunması ve mahremiyet ihlallerinin en aza indirilmesi için öneriler sunmaktadır. Ayrıca makale, sağlık verilerinin korunması için kriptografi gibi geleneksel yöntemlere odaklanmıştır.

Canbay ve arkadaşları (2021) çalışmalarında, derin öğrenme yaklaşımları kullanılarak Bilgisayarlı Tomografi (Computed Tomography - CT) ve Göğüs Röntgeni (Chest X-Ray - CXR) görüntülerinden Covid-19 hastalığının teşhisinin yaygınlaştığını ve bu alandaki çalışmaların arttığını vurgulamaktadır. Derin öğrenme teknikleri, özellikle derin öğrenme yaklaşımları mimarisi, yüksek başarı oranları ile hastalık teşhisinde etkili sonuçlar üretmektedir. Bununla birlikte, hastalara ait verilerin mahremiyetinin korunması, Kişisel Verileri Koruma Kanunu (KVKK) ve Genel Veri Koruma Tüzüğü (General Data Protection Rule - GDPR) gibi yasal düzenlemeler açısından büyük önem taşımaktadır. Veri mahremiyetinin korunması için k-Anonimlik, l-Çeşitlilik ve t-Yakınlık modelleri ile diferansiyel mahremiyet tabanlı çözümler kullanıldığını aktarmışlardır.

Çalışmada, Covid-19 teşhisinde derin öğrenmeyi kullanan ve veri mahremiyetini dikkate alan güncel çalışmalar incelenmiş ve bu çalışmaların başarı oranları karşılaştırılmıştır. Değerlendirmelerin, Covid-19 ve diğer hastalıkların tespitinde veri mahremiyetine önem verilmesi gerektiğini, diferansiyel mahremiyetin farklı ε

değerleri ile uygulanmasının sonuçlarının incelenmesi gerektiğini ve daha fazla açık Bilgisayarlı Tomografi ve X-Ray veri kümesinin yayınlanmasının başarı oranlarını artırabileceğini gösterdiğini belirtmişlerdir. Ayrıca, ülkemizde açık veri kümelerinin eksikliği nedeniyle yeterli çalışmalar yapılamadığı ve Federe Öğrenme modelinin uygulanabileceği belirtilmiştir.

Yiğit ve Kale (2021) araştırmalarında, son dönemde derin öğrenme, konuşma tanıma, doğal dil işleme ve görüntü işleme gibi çeşitli alanlarda devrim niteliğinde gelişmelerin sağlandığını ancak, hastane ve banka veri tabanları gibi hassas verilerin derin öğrenme modellerinde kullanılmasının, güvenlik ve gizlilik riskleri doğurduğunu belirtmişlerdir. Çalışmalarında, verilerin gizliliğini ve güvenliğini artırmak için homomorfik şifreleme, garbled devreler, güvenli çok partili hesaplama ve diferansiyel mahremiyet gibi kriptografik araçları incelemişlerdir. Literatürde bu araçların derin öğrenme temelli mimarilerde nasıl uygulandığı ve çeşitli saldırı tekniklerine karşı nasıl koruma sağladığına dair bilgiler vermişlerdir. Gelecekteki araştırmalar için, derin öğrenme modellerinin gizlilik ve güvenliğini artırmaya yönelik yaklaşımların karmaşıklığının azaltılması ve modellerin güvenilirliğini değerlendirmek için kriterlerin geliştirilmesini önermişlerdir.

Chaudhuri ve ekibi (2011), hassas bilgileri içeren eğitim verileriyle çalışan bir derin sinir ağı modelinde sınıflandırma problemini ele almıştır. Deneysel risk minimizasyonu kullanarak sınıflandırıcıların gizliliğini korumayı hedefleyen bir yöntem geliştirmişlerdir. Çalışmalarında, sınıflandırıcı her bir eğitim örneği için tahmin değeri belirlerken, eğitim verisi üzerindeki ortalama tahmin hatasını en aza indirme yaklaşımı benimsenmiştir. Ayrıca, hassas verilerin gizliliğini sağlamak amacıyla diferansiyel mahremiyet yöntemini kullanmışlardır.

Chase ve ekibi (2017), yapay sinir ağlarının eğitimi sırasında kullanılan verilerin gizliliğini korumak amacıyla diferansiyel mahremiyet ve Güvenli Çok Partili Hesaplama (Secure Multi-Party Computation - SMPC) yöntemlerini birleştiren yeni bir protokol sunmaktadır. Özellikle tıp ve finans gibi alanlarda, veri gizliliği yasal, etik ve rekabetçi nedenlerle büyük önem taşımaktadır. Bu çalışma, farklı kurumların veri paylaşımı yapmadan işbirliği içinde sinir ağı modelleri eğitmelerine olanak tanıyan bir yöntem geliştirmiştir. Bu yöntem, her bir veri kaydının gizliliğini koruyarak daha büyük veri setleri oluşturmayı ve bu veri setleri üzerinde makine öğrenimi algoritmalarını uygulamayı mümkün kılmaktadır.

Çalışmanın sonuçları, diferansiyel mahremiyet ve SMPC yöntemlerinin birleştirilmesinin, sinir ağı modellerinin gizlilik korumalı bir şekilde eğitilmesinde etkili olduğunu göstermektedir. Çalışma, bu yöntemlerin kullanımıyla, veri gizliliği ihlallerinin önlenebileceğini ve aynı zamanda yüksek doğrulukta modellerin elde edilebileceğini ortaya koymuştur. Ayrıca, bu protokolün uygulanabilirliği ve verimliliği, çeşitli deneysel sonuçlarla desteklenmiştir. Bu bulgular, gizlilik korumalı makine öğrenimi alanında önemli bir ilerleme olarak değerlendirilmektedir.

Abadi ve arkadaşları (2016), derin öğrenme modellerinin eğitimi sırasında kullanılan büyük ve temsili veri setlerinin gizliliğini korumak amacıyla diferansiyel mahremiyet tekniklerini incelemektedir. Derin öğrenme modellerinin, genellikle hassas bilgileri içeren veri setleriyle eğitildiğini ve bu modellerin eğitim bilgilerini ifşa etmemesi gerekmektiğini belirtmiştir. Bu amaçla, diferansiyel mahremiyet çerçevesinde yeni algoritmik teknikler geliştirilmiş ve gizlilik maliyetlerinin detaylı bir analizi yapılmıştır. Çalışma, konveks olmayan hedefler altında, makul bir gizlilik bütçesi ile derin sinir ağlarının eğitilebileceğini ve yazılım karmaşıklığı, eğitim verimliliği ve model kalitesi açısından yönetilebilir bir maliyetle gerçekleştirilebileceğini göstermektedir.

Çalışmanın sonuçları, diferansiyel mahremiyetin derin öğrenme modellerine uygulanabilirliğini ve etkinliğini göstermektedir. Çalışma, yüksek parametre sayısına sahip, çok katmanlı ve konveks olmayan hedeflere sahip modellerin, tek haneli bir gizlilik bütçesi ile eğitilebileceğini ortaya koymuştur. Ayrıca, gizlilik kaybının daha sıkı tahminlerinin elde edilebileceği ve bu tahminlerin hem asimptotik hem de ampirik olarak daha doğru sonuçlar verdiği belirtilmiştir. Bu bulgular, derin öğrenme modellerinin gizlilik koruma mekanizmaları ile birlikte kullanılmasının mümkün ve etkili olduğunu göstermektedir.

Mulder ve Humbert (2023), diferansiyel mahremiyetin veri setleri hakkında bilgi paylaşırken bireysel gizliliği koruma teknolojisi olarak nasıl kullanıldığını incelemektedir. Diferansiyel mahremiyetin, sonuçlara gürültü ekleyerek veri tabanındaki bireylerin bilgilerinin ifşa edilmesini engellediğini ve bu teknolojinin, sayım, toplama veya ortalama gibi büyük veri tabloları üzerindeki sorgulamalarda etkili bir şekilde kullanılabildiğini belirtmişlerdir. Diferansiyel mahremiyetin, güçlü bir saldırganın sahip olabileceği ek bilgilere karşı dayanıklı, genel ve sağlam gizlilik garantileri sağladığını; nüfus sayımı, sağlık kayıtları, güvenlik, eğitim kalitesinin

izlenmesi ve iş stratejilerinin analizi gibi birçok alanda uygulanabildiğini aktarmışlardır.

Makalenin sonuçları, diferansiyel mahremiyetin bireysel gizliliği korurken veri paylaşımını mümkün kılan etkili bir araç olduğunu göstermektedir. Ancak, bu yöntemin en büyük dezavantajının, genellikle veri doğruluğunu gizlilik karşılığında feda etmesi olduğunu belirtmişlerdir. Diferansiyel mahremiyetin, hükümetler ve büyük şirketler için veri gizliliği taleplerine daha iyi uyum sağlamada önemli bir araç olabileceğini, ayrıca, diferansiyel mahremiyetin uygulanabilirliği ve kullanılabilirliğinin, açık kaynak sistemlerin geliştirilmesi ve erişilebilir araçların yaygınlaşması ile artmakta olduğunu aktarmışlardır.

Wood ve arkadaşları (2018), diferansiyel mahremiyetin temel kavramlarını ve uygulamalarını teknik olmayan bir kitleye tanıtmayı amaçlamaktadır. Diferansiyel mahremiyetin, veri analizleri ve istatistiksel sonuçların paylaşımı sırasında bireylerin gizliliğini korumak için kullanılan matematiksel bir çerçeve olduğunu, bu yöntemin, potansiyel saldırılara karşı güçlü bir koruma sağladığını ve kişisel bilgilerin ifşa edilmesini engellediğini aktarmışlardır. Makale, diferansiyel mahremiyetin tanımını, gizlilik risklerini nasıl ele aldığını, farklı analizlerin nasıl oluşturulduğunu ve pratikte nasıl kullanılabileceğini basit ve anlaşılır bir şekilde açıklamaktadır. Ayrıca, bu kavramların sosyal bilimler ve hukuk alanlarındaki uygulamalarına dair örnekler sunmaktadır.

Makalenin sonuçları, diferansiyel mahremiyetin bireylerin gizliliğini korurken veri paylaşımını mümkün kılan etkili bir araç olduğunu göstermektedir. Diferansiyel mahremiyetin, veri analizlerinde gizlilik kaybını minimize ederken, aynı zamanda yasal ve etik gerekliliklere uyum sağlamada önemli bir rol oynadığını belirtmişlerdir. Çalışma, diferansiyel mahremiyetin uygulanabilirliğini ve etkinliğini vurgulamakta ve bu yöntemin daha geniş bir kullanıcı kitlesi tarafından benimsenmesi gerektiğini önermektedir. Ayrıca, diferansiyel mahremiyetin sağladığı garantilerin, veri yöneticileri ve politika yapıcılar için önemli bir rehber olabileceği belirtilmiştir.

Hall ve arkadaşları (2013), diferansiyel mahremiyetin fonksiyonlar ve fonksiyonel veriler üzerindeki uygulamalarını incelemektedir. Geleneksel diferansiyel mahremiyet yöntemlerinin, genellikle sonlu boyutlu vektörler veya belirli bir ayrık kümenin elemanları üzerinde çalıştığını, kendi çalışmalarının ise fonksiyonların gizliliğini

korumaya odaklanmakta olduğunu belirtmişlerdir. Özellikle, ilgilenilen fonksiyona uygun bir Gauss süreci ekleyerek diferansiyel mahremiyet sağlanabileceği gösterilmiştir. Fonksiyonlar, Gauss sürecinin kovaryans çekirdeği tarafından oluşturulan yeniden Üretici Çekirdek Hilbert Uzayı'nda (Reproducing Kernel Hilbert Space - RKHS) yer aldığında, doğru gürültü seviyesinin Üretici Çekirdek Hilbert Uzayı normunda fonksiyonun "duyarlılığı" ölçülerek belirlendiği belirtilmiştir. Çalışmada, çekirdek yoğunluk tahmini, çekirdek destek vektör makineleri ve Üretici Çekirdek Hilbert Uzayı'ndaki fonksiyonlar gibi örnekler ele alınmıştır.

Makalenin sonuçları, diferansiyel mahremiyetin fonksiyonel veriler üzerinde uygulanabilirliğini ve etkinliğini göstermektedir. Çalışma, fonksiyonların gizliliğini korumak için uygun bir Gauss süreci eklemenin yeterli olduğunu ve bu yaklaşımın çeşitli makine öğrenimi görevlerinde kullanılabileceğini ortaya koymuştur. Özellikle, çekirdek yoğunluk tahmini ve destek vektör makineleri gibi yöntemlerde diferansiyel mahremiyetin sağlanabileceği ve bu yöntemlerin gizlilik koruma mekanizmaları ile birlikte etkili bir şekilde çalışabileceği gösterilmiştir. Bu bulgular, diferansiyel mahremiyetin fonksiyonel veriler üzerindeki uygulamalarının genişletilebileceğini ve bu alanda daha fazla araştırma yapılması gerektiğini vurgulamaktadır.

Raab ve ekibi (2024), Yerel Diferansiyel Mahremiyet (Local Differential Privacy) altında tanımlayıcı istatistiklerin hesaplanmasını incelemektedir. Yerel Diferansiyel Mahremiyetin, bireylerin verilerini ifşa etmeden hassas verilerin toplanmasını ve analiz edilmesini sağlayan bir gizlilik garantisi sunduğunu aktarmışlardır. Çalışma, Yerel Diferansiyel Mahremiyet yöntemlerinin tanımlayıcı istatistikler için sistematik bir şekilde sınıflandırılmasını ve bu yöntemlerin özelliklerinin ve gereksinimlerinin karşılaştırılmasını amaçlamaktadır.

Makalenin sonuçları, Yerel Diferansiyel Mahremiyet yöntemlerinin tanımlayıcı istatistikler için kullanılabilirliğini ve etkinliğini göstermektedir. Çalışma, Bernoulli dağılımından örnekleme yoluyla ortalama tahmin yöntemlerinin bir boyutlu durumda eşdeğer olduğunu ve varyans tahmini için yeni yöntemler sunduğunu ortaya koymuştur. Ampirik karşılaştırmalar, ortalama, varyans ve frekans tahmini için Yerel Diferansiyel Mahremiyet yöntemlerinin performansını değerlendirmiştir. Sonuç olarak, Yerel Diferansiyel Mahremiyet yöntemlerinin tanımlayıcı istatistikler için kullanımı konusunda önerilerde bulunulmuş ve bu yöntemlerin sınırlamaları ve açık sorular tartışılmıştır.

Sörensen ve Glover (2013), meta sezgisel algoritmaların tanımını ve uygulamalarını kapsamlı bir şekilde ele almaktadırlar. Meta sezgisel algoritmaların, belirli bir problem türüne özgü olmayan, genel bir çerçeve sunarak çeşitli optimizasyon problemlerini çözmek için kullanılan yüksek seviyeli algoritmik stratejiler olduğunu belirtmişlerdir. Bu algoritmaların, genetik algoritmalar, tabu arama, benzetimli tavlama ve karınca kolonisi optimizasyonu gibi çeşitli yöntemleri içerdiğini; özellikle karmaşık ve büyük ölçekli problemlerde, çözüm kalitesi ve hesaplama süresi arasında iyi bir denge sağlayarak, geleneksel yöntemlere kıyasla üstün performans gösterebildiğini aktarmışlardır.

Makalenin sonuçları, meta sezgisel algoritmaların esnekliği ve uyarlanabilirliği sayesinde, geniş bir uygulama yelpazesinde etkili çözümler sunduğunu göstermektedir. Bu algoritmaların, optimizasyon problemlerinin çözümünde kombinatoryal patlama sorununu yani bir problemin çözüm uzayının boyutunun çok hızlı bir şekilde büyümesi durumunu önleyerek, daha hızlı ve yeterince iyi çözümler üretebildiklerini söylemişlerdir. Ancak, bu esneklik ve uyarlanabilirlik, algoritmaların belirli problemlere uyarlanması için önemli ölçüde özel ayarlamalar gerektirebildiğini aktarmışlardır. Son olarak, meta sezgisel algoritmaların, bilimsel topluluk tarafından kabul gören ve sıklıkla tercih edilen bir alternatif olduğunu vurgulamışlardır.

Alcaraz ve arkadaşları (2024), makinelerin etik kararlar alabilmesi için yeni bir yöntem geliştirmişlerdir. Bu yöntem, felsefi düşünceleri, tartışma teorisini ve makine öğrenimini birleştirmektedir. Temel olarak, bir eylemin doğru mu yanlış mı olduğunu belirlemek için çeşitli "normatif nedenler" (ahlaki gerekçeler) arasındaki etkileşimleri kullanmışlardır. Bu modelde, normatif nedenlerin önem derecelerini tahmin etmek için ağırlıklı tartışma grafikleri kullanılmıştır. Bu tahminler, daha önce bilinen ahlaki durumlara dayanarak yapılmakta ve yeni durumlarda eylemlerin ahlaki statüsünü belirlemek için kullanılmaktadır.

Sonuç olarak bu yöntem, makinelerin etik kararlar alırken hem yukarıdan aşağıya (teoriden pratiğe) hem de aşağıdan yukarıya (pratikten teoriye) yaklaşımların avantajlarını birleştirmektedir. İlk deneyler, bu modelin hem eğitim sürecinde verimli olduğunu, hem de kararların şeffaflığını artırdığını göstermiştir. Bu bulgular, makine etiği alanında daha gelişmiş ve uyarlanabilir sistemlerin oluşturulmasına yönelik önemli bir adım olarak görülmektedir.

Dokeroglu ve arkadaşları (2022), özellik seçimi için son yıllarda geliştirilen meta sezgisel algoritmaların kapsamlı bir incelemesini sunmaktadır. Özellik seçiminin, büyük veri kümelerinde gereksiz veya alakasız özellikleri çıkararak makine öğrenimi modellerinin performansını artırmak için kullanılan bir teknik olduğunu belirtmişlerdir. Makale, son yirmi yılda geliştirilen en dikkat çekici metasezgisel algoritmaları, keşif ve sömürü operatörleri, seçim yöntemleri, transfer fonksiyonları, uygunluk değeri değerlendirmeleri ve parametre ayarlama teknikleri açısından değerlendirmektedir.

Sonuçlar, metasezgisel algoritmaların özellik seçimi sürecinde üstün performans gösterdiğini ve bu alanda gelecekteki araştırmalar için çeşitli zorluklar ve potansiyel konular sunduğunu ortaya koymaktadır. Bu algoritmaların, büyük veri kümelerinde daha etkili ve verimli özellik seçimi yaparak, makine öğrenimi modellerinin doğruluğunu ve genel performansını artırmada başarılı olduğunu belirtmişlerdir.

Akay ve ekibi (2022), derin öğrenme modellerinin optimizasyonu için meta sezgisel algoritmaların kapsamlı bir incelemesini sunmaktadır. Derin Sinir Ağları'nın (Deep neural networks), yapay sinir ağlarının genişletilmiş versiyonları olup, ham veri alanını daha karmaşık bir özellik alanına dönüştürerek daha yüksek seviyeli özellik hiyerarşilerini öğrenebildiklerini aktarmışlardır. Ancak, model parametrelerinin uygun değerlerinin seçilmesi, optimal mimarinin belirlenmesi ve ağırlık ve önyargı değerlerinin belirlenmesi gibi Derin Sinir Ağlarının performansını etkileyen bazı sorunlar olduğunu belirtmişlerdir. Son yıllarda, bu görevleri otomatikleştirmek için meta sezgisel algoritmaların önerildiğini söylemişlerdir.

Çalışma, yaygın olarak kullanılan Derin Sinir Ağları mimarileri hakkında kısa bilgiler vermekte ve Derin Sinir Ağları tasarımında optimizasyon problemlerini formüle etmektedir. Ayrıca, meta sezgisel algoritmaların Derin Sinir Ağlarını optimize etmek için nasıl kullanılabileceğini açıklar. İlk deneyler, bu modelin hem eğitim sürecinde verimli olduğunu, hem de kararların şeffaflığını artırdığını göstermektedir. Bu bulgular, makine etiği alanında daha gelişmiş ve uyarlanabilir sistemlerin oluşturulmasına yönelik önemli bir adım olarak değerlendirilmektedir.

Talbi (2021) araştırmasında, makine öğrenimi ve meta sezgisel algoritmaların entegrasyonunun, optimizasyon problemlerinin çözümünde önemli fırsatlar sunmakta olduğunu belirtmiştir. Çalışmasında, makine öğreniminin meta sezgisel algoritmalara

uygulanması için çeşitli sinerjiler ve fırsatlar incelenmiştir. Makine öğrenimi destekli meta sezgisel algoritmalar, yüksek kaliteli sonuçlar üretmiş ve en son optimizasyon algoritmalarını temsil etmektedir. Ancak, makine öğrenimi tekniklerinin entegrasyonu ek maliyetler getirdiğinden, bu maliyetlerin düşük tutulmasının gerektiğini belirtmiştir.

Gelecekte, makine öğrenimi ve meta sezgisel algoritmaların etkileşiminin artmasını beklemektedir. Derin öğrenme modelleri gibi daha sofistike makine öğrenimi tekniklerinin kullanımının, daha karmaşık problemleri çözmek için ilginç bir alternatif olacağını aktarmıştır. Transfer öğreniminin, dinamik ve çapraz alan optimizasyon problemleri için meta sezgisel arama sürecini hızlandırabileğini belirtmiştir. Ayrıca, makine öğrenimi destekli meta sezgisel algoritmalar için paralel modellerin tasarımı ve uygulanmasının, yüksek performanslı hesaplama sistemleriyle birlikte önemli bir araştırma alanı olarak öne çıkmakta olduğunu aktarmıştır.

Vasa ve Thakkar (2023), büyük veri çağında derin öğrenme modellerinde gizlilik korumasını incelemektedir. Özellikle, diferansiyel mahremiyetin derin öğrenme modellerinde nasıl uygulanabileceği ve meta sezgisel algoritmaların bu süreçteki rolü üzerinde durulmuştur. Diferansiyel mahremiyetin, bireylerin verilerinin anonimleştirilmiş veri setlerinde bile tanınmasını engellemek için kullanılan bir Teknik olduğu aktarılmıştır. Makalede, diferansiyel mahremiyetin sağlanması için Laplace ve üstel mekanizmalar gibi yöntemler detaylandırılmıştır. Ayrıca, derin öğrenme modellerinde diferansiyel mahremiyetin korunması için çeşitli saldırı türleri ve bu saldırılara karşı geliştirilen yöntemler ele alınmıştır. Örneğin, adaptif Laplace mekanizması kullanılarak derin öğrenme modellerine gürültü eklenmesi ve böylece gizliliğin korunması sağlanmıştır.

Meta sezgisel algoritmaların ise büyük veri analizinde gizlilik koruması sağlamak için kullanılan diğer önemli bir yaklaşım olduğu söylenmiştir. Bu algoritmaların, optimizasyon problemlerini çözmek için doğadan ilham alan yöntemleri kullandığı belirtilmiştir. Makalede, kaos ve bozulma tekniklerine dayalı anonimleştirme algoritmaları gibi çeşitli meta sezgisel yöntemler incelenmiştir. Ayrıca, gri kurt optimizasyonu ve kedi sürüsü optimizasyonu gibi doğadan ilham alan algoritmaların büyük veri güvenliğini sağlamak için nasıl kullanılabileceği tartışılmıştır. Bu yöntemlerin, verilerin anonimleştirilmesi ve gizliliğin korunması için etkili çözümler sunmakta olduğu bildirilmiştir.

Sonuç olarak makale, derin öğrenme modellerinde diferansiyel mahremiyetin korunması ve meta sezgisel algoritmaların bu süreçteki rolü hakkında derinlemesine bir inceleme sunmaktadır. Bu yöntemlerin büyük veri analizinde gizlilik koruması sağlamak için etkili olduğu vurgulanmış, ancak bazı sınırlamaları olduğu ve gelecekte bu sınırlamaların aşılması için daha fazla araştırma yapılması gerektiği belirtilmiştir. Bu bulguların, büyük veri çağında güvenli ve gizliliği koruyan veri analiz yöntemlerinin önemini bir kez daha ortaya koymakta olduğu aktarılmıştır.

Yukarıda bahsedilen çalışmalar incelendikten sonra; bu çalışmaya bakıldığında, sürekli ve kategorik veriler için farklı bir yaklaşım benimsenerek daha kapsamlı bir çözüm sunulmuştur. Meta sezgisel yöntemler, diferansiyel mahremiyet izninin optimize edilmesi için kullanılmış ve başarılı sonuçlar elde edilmiştir. Diferansiyel mahremiyet ve meta sezgisel yöntemlerin birleştirilmesiyle, gizlilik-fayda dengesi konusunda önemli bir adım atılmıştır. Bu yaklaşım, özellikle sağlık verileri gibi hassas verilerin korunması açısından büyük önem taşımaktadır. Sürekli ve kategorik veriler için farklı gizlilik oranları belirlenerek, veri özelliklerine özgü bir yaklaşım benimsenmiştir. Bu durum, veri analizlerinin daha doğru ve güvenilir olmasını sağlamaktadır.

Yine bu çalışmada, genetik algoritmanın Epsilon parametrelerini belirlemede başarılı olduğu görülmüştür. Genetik algoritmalar, veri mahremiyetini maksimum seviyede korurken, aynı zamanda veri analizinin doğruluğunu da koruma avantajına sahiptir. Elde edilen sonuçlar, farklı veri türleri ve daha karmaşık veri yapılarında da uygulanabilir. Özellikle derin öğrenme gibi gelişmiş makine öğrenmesi teknikleriyle birleştirilerek daha farklı çözümler geliştirilebilir. Farklı meta sezgisel yöntemlerin performanslarının karşılaştırılması ve yeni hibrit yöntemlerin geliştirilmesi, gelecekteki çalışmalar için önemli bir araştırma alanı olabilir. Bu çalışma, diferansiyel mahremiyet ve meta sezgisel yöntemlerin entegrasyonunun, veri mahremiyeti ve analiz doğruluğu arasındaki dengeyi optimize etmede başarılı olduğunu göstermiştir. Bu bulgular, gelecekte yapılacak araştırmalara rehberlik edecek nitelikte olup, farklı mahremiyet alanlarında yeni bakış açısı sunmaktadır. Literatüre hem teknik hem de uygulamalı katkılar sağlayan bu çalışma, diferansiyel mahremiyetin optimize edilmesi konusunda önemli bir adım olarak değerlendirilebilir.

7. SONUÇ VE ÖNERİLER

7.1. Sonuç

7.1.1. Gizlilik - Fayda Dengesi

Bu çalışma, diferansiyel mahremiyet ve meta sezgisel yöntemlerin entegre edilmesi ile özellikle gizlilik - fayda arasındaki dengeyi sağlamakta ve gizlilik iznini optimize etme konusunda başarı sağlamaktadır. Ayrıca, bu yaklaşımlar, verilerin analitik değeri korunurken bireylerin gizliliğini artırmayı hedeflemektedir.

7.1.2. Sürekli ve Kategorik Veriler

Literatürdeki çoğu uygulama, aynı çalışma veri türlerine odaklanmış veya tek bir mekanizmayı ayrım gözetmeden uygulamıştır. Bu çalışmada ise, aynı veri setindeki sürekli ve kategorik veri türleri arasında farklı gizlilik oranları ve veri özelliklerine özgü bir yaklaşım benimsenmiştir. Bu durum, sürekli ve kategorik veriler üzerinde ayrı ayrı analizler yapılan, farklı mahremiyet yöntemlerinin her iki veri tipi üzerinde farklı etkiler yarattığını göstermiştir. Özellikle, sürekli veri türlerinde daha yüksek hassasiyet ve doğruluk sağlanırken, kategorik verilerde gizlilik ve fayda arasında optimal bir denge kurulmuştur.

7.1.3. Meta Sezgisel Yöntemler, Genetik Algoritmalar ve Epsilon Parametreleri

Meta sezgisel yöntemler sayesinde diferansiyel mahremiyet izninin geleneksel yöntemlere göre daha etkin bir şekilde optimize edildiği görülmüştür. Çalışmada kullanılan genetik algoritma yaklaşımı, özellikle epsilon parametrelerini belirlerken hem veri mahremiyetini maksimum seviyede sağlamakta hem de veri analizinin doğruluğunu korumaktadır.

7.1.4. Sağlık Verileri ve Esnek Çerçeve

Bu yaklaşım, sağlık verilerinin sürdürülmesinde güvenilir ve kullanışlı bir çözüm sunmakla birlikte, diğer veri türlerine de genişletilebilir esnek bir çerçeve sağlamaktadır. Örneğin, finansal veriler, eğitim verileri ve sosyal medya verileri gibi

farklı veri türleri üzerinde de benzer mahremiyet koruma yöntemleri uygulanabilir. Sağlık verilerinde, özellikle hasta kayıtları ve genetik bilgiler gibi hassas verilerin gizliliğini korumak büyük önem taşımaktadır.

7.1.5. Geleneksel Yöntemlerle Karşılaştırma

Önerilen yöntem, geleneksel diferansiyel mahremiyet algoritmalarına kıyasla hem veri kayıplarını en aza indirmiş hem de veri özelliklerini korumada üstün performans sergilemiştir. Geleneksel yöntemlerin aksine, bu yaklaşım verilerin analitik değerini korurken, gizlilik ihlallerini minimuma indirmeyi hedeflemektedir. Özellikle, büyük veri setleri ve karmaşık veri yapılarına sahip uygulamalarda, önerilen yöntemlerin avantajları daha belirgin hale gelmektedir.

7.1.6. Literatüre Katkı ve Gelecek Araştırmalar

Çalışma, literatüre hem teknik bir katkı sağlamış hem de sağlık verileri gibi hassas veri türleri üzerinde mahremiyet korumasının etkin bir şekilde sağlanabileceğini göstermiştir. Çalışma, gelecekte yapılacak araştırmalara rehberlik sunacak kalitede olup, farklı mahremiyet alanında yeni bir bakış açısı sunmaktadır. Gelecek çalışmalar, önerilen metodolojinin farklı veri türleri ve uygulama alanlarında test edilmesi ve iyileştirilmesi üzerine odaklanabilir.

7.2. Öneriler

7.2.1. Daha Fazla Parametre Seçeneği ile Optimizasyon

Gelecek çalışmalar için, sadece epsilon değil, delta gibi diğer diferansiyel gizlilik değerinin de optimize edilmesi değerlendirilebilir. Özellikle Gauss mekanizmaları için delta parametreleri önemli bir performansa sahiptir ve bunun optimize edilmesi, genel gizlilik-fayda dağılımını daha da iyileştirilebilir.

7.2.2. Alternatif Meta Sezgisel Algoritmaların Kullanımı

Genetik uygulamaların yanı Parçacık Sürü Optimizasyonu (PSO) gibi diğer meta sezgisel yöntemlerinin kullanılması, değişim sürecini hızlandırabilir veya farklı veriler için daha uygun sonuçlar verilebilir. Bu tip farklı yöntemlerin denenmesi, yöntemin genellenebilirliğini artıracak ve daha iyi test etme imkânı sunacaktır.

7.2.3. Farklı Veri Kümesi Çalışmaları

Bu araştırmada tıbbi sağlık verileri kullanılmış olsa da, aynı teknoloji farklı sektörlerdeki verilerle de test edilebilir. Örneğin finans veya sosyal medya verileri gibi farklı nitelikteki veri setleri üzerinde bu yöntemin uygulanabilirliği araştırılabilir.

7.2.4. Farklı Gizlilik Seviyelerinin İncelenmesi

Çalışmada optimize edilen gizlilik haklarına ek olarak, farklı gizliliklerin açıklanmasının doğrulanması üzerinde ayrıntılı olarak analiz yapılabilir. Bu yetenek, düşük, orta ve yüksek düzeydeki gizlilik standartlarına uygun özelleştirilmiş çözümler geliştirilebilir.

7.2.5. Hibrit Yöntemlerinin Geliştirilmesi

Gelecekte, diferansiyel mahremiyet yöntemi ile meta sezgisel yöntemlerin birlikte büyüyebilmesi için entegre sistemin iyileştirilmesi değerlendirilebilir. Örneğin, genetik çözümler ile derin öğrenme modellerinin birleşimi, karmaşık veri kümelerinde daha etkili sonuçlar elde edebilme olanağı sağlayabilir.

7.2.6. Yöntemin Gerçek Zamanlı Uygulamaları

Önerilen yöntemin gerçek zamanlı veri değişimlerine uygulanabilirliği incelenebilir. Özellikle dinamik veri ortamlarında gizlilik - fayda dengesini sağlamak için anlık ağ yapan sistemlerin tasarlanması, yöntemin pratik kullanımı açısından önemli bir adım olacaktır.

7.2.7. Yapay Zeka Tabanlı İyileştirmeler

Meta sezgisel güncellemelerin arttırılması için, yapay zeka ve makine modülleri ile desteklenen algoritmalar önerilebilir. Örneğin, gizlilik ve fayda arasındaki dengenin daha hızlı ve etkili bir şekilde optimize edilmesi için takviyeli öğrenme (pekiştirmeli öğrenme) gibi yöntemler kullanılabilir. Bu yöntemler, sürekli geri bildirimler alınarak değişim sürecini dinamik bir şekilde yönlendirebilir ve daha başarılı sonuçlar elde edilebilmesini sağlayabilir.

7.2.8. Enerji ve Hesaplama Verimliliği Üzerine Çalışmalar

Meta sezgisel yöntemlerin kullanımıyla optimize edilen aydınlatma sistemleri, enerji tüketimi ve hesaplama maliyetleri açısından detaylı analizlere tabi tutulabilir. Bu analizler, sistemlerin enerji verimliliğini artırmaya yönelik olarak gerçekleştirilebilir ve çevre dostu çözümler geliştirilmesine katkı sağlayabilir. Bu bağlamda, enerji verimliliğine odaklanan çeşitli analiz tekniklerinin değerlendirilmesi, daha sürdürülebilir ve maliyet etkin aydınlatma çözümlerinin tasarımına imkân tanıyacaktır.

KAYNAKÇA

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep Learning With Differential Privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 308-318.

Abowd, J., & Schmutte, I. (2019). An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. *American Economic Review*.

Akay, B., Karaboga, D., & Akay, R. (2022). A Comprehensive Survey On Optimizing Deep Learning Models by Metaheuristics. *Artificial Intelligence Review*, 55(2), 829-894.

Akpusugh V. E. & Dekpoghol, T. (2024). Comparing Discrete and Continuous Data: Concepts, Differences, and Applications. *ScienceOpen Preprints.*

Alcaraz, B., Knoks, A., & Streit, D. (2024). Estimating Weights of Reasons Using Metaheuristics: A Hybrid Approach to Machine Ethics. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (7). 27-38.

Atalay, H. N. (2021). Mahremiyet Kapsamında Kişisel Sağlık Verilerinin Korunması ve Depolanması. *Journal of Academic Perspective on Social Studies (1)*, 1-20. https://doi.org/10.35344/japss.786353.

Atalay, M., & Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamalari-Artificial İntelligence and Machine Learning Applications in Big Data Analysis. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9(22), 155-172.

Aygün, S. Akıllı Kart ve Biyometrik Tabanlı genel Amaçlı Erişim Kontrolü Sistemi Tasarımı. Yayımlanmamış Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi. İstanbul, 2015.

Bailey, C., Chadwick, D. W., & De Lemos, R. (2011). Self-adaptive Authorization Framework for Policy Based RBAC/ABAC Models. In *2011 IEEE Ninth International Conference on Dependable, Autonomic and Secure Computing*. 37-44.

Balle, B., Barthe, G., & Gaboardi, M. (2018). Privacy Amplification by Subsampling: Tight Analyses via Couplings and Divergences. *Proceedings of the 35th International Conference on Machine Learning*.

Balle, B., & Wang, Y. X. (2018). Improving The Gaussian Mechanism for Differential Privacy: Analytical Calibration and Optimal Denoising. *International Conference on Machine Learning.* 394-403.

Bastı, M. (2012). P-medyan Tesis Yeri Seçim Problemi ve Çözüm Yaklaşımları. *AJIT-e: Academic Journal of Information Technology*, 3(7), 47-75.

Beşkirli, A., Özdemir, D. & Beşkirli, M. (2019). Şifreleme Yöntemleri ve RSA Algoritması Üzerine Bir İnceleme. *European Journal of Science and Technology, (Special Issue),* 284-291.

Canbay, Y., İsmetoğlu, A., & Canbay, P. (2021). Covid-19 Hastalığının Teşhisinde Derin Öğrenme ve Veri Mahremiyeti. *Mühendislik Bilimleri ve Tasarım Dergisi*, 9(2), 701-715.

Canbay, Y., & Sağıroğlu, Ş. (2020). Derin Öğrenmede Diferansiyel Mahremiyet. *Uluslararası Bilgi Güvenliği Mühendisliği Dergisi*, 6(1), 1-16.

Chase, M., Gilad-Bachrach, R., Laine, K., Lauter, K., & Rindal, P. (2017). Private Collaborative Neural Network Learning. *Cryptology ePrint Archive*.

Chaudhuri, K., Monteleoni, C., & Sarwate, A. D. (2011). Differentially Private Empirical Risk Minimization. *Journal of Machine Learning Research*, 12(3).

Dokeroglu, T., Deniz, A., & Kiziloz, H. E. (2022). A Comprehensive Survey On Recent Metaheuristics For Feature Selection. *Neurocomputing*, 494, 269-296.

Domingo-Ferrer, J., Sánchez, D., & Blanco-Justicia, A. (2021). The Limits of Differential Privacy (And its Misuse in Data Release And Machine Learning). *Communications of The ACM*, 64(7), 33-35.

Dorigo, M., & Stützle, T. (2019). *Ant Colony Optimization: Overview and Recent Advances*. 311-351.

Dwork, C. (2006). Differential Privacy. *International Colloquium on Automata, Languages and Programming.* 1-12.

Dwork, C., & Roth, A. (2014). The Algorithmic Foundations of Differential Privacy. *Foundations and Trends in Theoretical Computer Science*.

Dursun, Ö. O., & Toraman, S. (2021). Uzun Kısa Vadeli Bellek Yöntemi ile Havayolu Yolcu Tahmini. *Journal of Aviation*, 5(2), 241-248.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning From İmbalanced Data: Progress and Challenges, Marking The 15-year Anniversary. *Journal of Artificial Intelligence Research*, (61), 863-905.

Fortin, F. A., De Rainville, F. M., Gardner, M. A. G., Parizeau, M., & Gagné, C. (2012). DEAP: Evolutionary Algorithms Made Easy. *The Journal of Machine Learning Research,* 13(1), 2171-2175.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction.* 2, 1-758.

Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., ... & Amira, A. (2023). AI-Big Data Analytics for Building Automation and Management Systems: A Survey, Actual Challenges and Future Perspectives. *Artificial Intelligence Review*, 56(6), 4929-5021.

Huang, J., & Ling, C. X. (2005). Using AUC and Accuracy in Evaluating Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 17(3), 299-310.

İnan, A., & Var, E. (2018). Sınıflandırma İçin Diferansiyel Mahremiyete Dayalı Öznitelik Seçimi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, 33(1).

Kairouz, P., Oh, S., & Viswanath, P. (2015). The Composition Theorem For Differential Privacy. *International Conference on Machine Learning*. 1376-1385.

Kara, B. C., & Eyüpoğlu, C. (2020). Sağlık 4.0'da Mahremiyet ve Güvenlik Sorunları. *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. 1-12.

Kennedy, J., & Eberhart, R. (1995). Particle Swarm Optimization. *Proceedings of ICNN'95-International Conference on Neural Networks*. 4. 1942-1948.

Kodaz, H., & BOTSALI, F. M. (2010). Simetrik Ve Asimetrik Şifreleme Algoritmalarinin Karşilaştirilmasi. *Selcuk University Journal of Engineering Sciences*, 9(1), 10-23.

Konecny, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated Learning: Strategies for Improving Communication Efficiency. arXiv preprint arXiv:1610.05492.

Kurita, T. (2019). Principal Component Analysis (PCA). *Computer Vision: A Reference Guide*, 1-4.

Kuş, M. C. Differential Privacy in Financial Distributed Ledger Applications. Yayımlanmamış Doktora Tezi, Sabancı Üniversitesi. İstanbul, 2022.

Lau, K. K., & Taweel, F. M. (2007). Data Encapsulation in Software Components. *International Symposium on Component-Based Software Engineering*. 1-16.

Liang, J. (2022). Confusion Matrix: Machine Learning. *POGIL Activity Clearinghouse*, 3(4).

Liu, W., Zhang, Y., Yang, H., & Meng, Q. (2024). A Survey On Differential Privacy For Medical Data Analysis. *Annals of Data Science*, 11(2), 733-747.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS).*

Mironov, I. (2017). Rényi Differential Privacy. *Proceedings of the 30th International Conference on Neural Information Processing Systems.*

Mulder, V., & Humbert, M. (2023). Differential Privacy. *Trends in Data Protection and Encryption Technologies.* 157-161.

Murtagh, J., & Vadhan, S. (2015). The Complexity of Computing The Optimal Composition of Differential Privacy. *Theory of Cryptography Conference.* 157-175.

Narayanan, A., & Shmatikov, V. (2008). Robust de-Anonymization of Large Sparse Datasets. *2008 IEEE Symposium on Security and Privacy*. 111-125.

Neel, S. V., Roth, A. L., & Wu, Z. S. (2019). How To Use Heuristics for Differential Privacy. *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS).* 72-93.

Neubauer, T., & Heurix, J. (2011). A Methodology For The Pseudonymization of Medical Data. *International Journal Of Medical Informatics*, 80(3), 190-204.

Özdemir, M. (2017). Genetik Algoritma İle Doğrusal Regresyonda Tahmin Amaçli Model Seçimi. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (28), 213-233.

Raab, R., Berrang, P., Gerhart, P., & Schröder, D. (2024). SoK: Descriptive Statistics Under Local Differential Privacy. *IACR Cryptology ePrint Archive*. 1464.

Sağır, Ö. & Vural, U. (2024(I)). Veri ve Erdem: Makine Öğrenmesinde Etik Bir Veri Kültürü Oluşturma. *Hukuk, Bilişim, Teknoloji Sempozyumu* (Yayımlanmamış).

Sağır, Ö. & Vural, U. (2024(II)). Diferansiyel Mahremiyet Yöntemlerinin Sağlık Verilerine Uygulanması: Laplace ve Gaussian Mekanizmalarının Karşılaştırılması. *1. Kocaeli Uluslararası Mühendislik-Doğa Bilimleri ve Sağlık Teknolojileri Konferansı* (Yayımlanmamış).

Shakeel, A., Tanaka, T., & Kitajo, K. (2020). Time-series Prediction of The Oscillatory Phase of EEG Signals Using The Least Mean Square Algorithm-based AR Model. *Applied Sciences*, 10(10), 3616.

Sörensen, K., & Glover, F. (2013). Metaheuristics. *Encyclopedia of Operations Research and Management Science*, 62, 960-970.

Vanneschi, L., & Silva, S. (2023). Genetic Algorithms. *Lectures on Intelligent Systems.* 45-103.

Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion Matrix-Based Feature Selection. *Maics*, 710(1), 120-127.

Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., ... & Vadhan, S. (2018). Differential Privacy: A Primer For a Non-technical Audience. *Vanderbilt Journal of Entertainment & Technology Law*, 21, 209.

Yang, J., Zhao, X., Lu, X., Lin, X., & Xu, G. (2015). A Data Preprocessing Strategy for Metabolomics to Reduce The Mask Effect in Data Analysis. *Frontiers in Molecular Biosciences*, *2*, 4.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2), 1-19.

Yare, A. (2024). *Health Metrics Dataset* [Data set]. Kaggle. https://www.kaggle.com/datasets/abhayayare/health-metrics-dataset?resource=download Erişim tarihi: 01.09.2024

KİŞİSEL YAYINLAR VE ESERLER

[1] Ozturk, C., **Sagir, O**., & Vural, U. (2024). Machine Learning Approaches to Predict Thyroid Cancer Recurrence: A Comparative Study. *In 2024 9th International Conference on Computer Science and Engineering (UBMK)*. IEEE. doi: 10.1109/UBMK63289.2024.10773518.

[2] **Sağır, Ö**., & Albayrak, N. B. (2024). Veri Madenciliğinin Siber Suçlarda Kötüye Kullanımı. *Journal of Kocaeli Health and Technology University*, 2(3), 68-80.

[3] **Sağır, Ö**. & Vural, U. (2024). Veri ve Erdem: Makine Öğrenmesinde Etik Bir Veri Kültürü Oluşturma. *Hukuk, Bilişim, Teknoloji Sempozyumu* (Yayın aşamasında).

[4] **Sağır, Ö**. & Vural, U. (2024). Diferansiyel Mahremiyet Yöntemlerinin Sağlık Verilerine Uygulanması: Laplace ve Gaussian Mekanizmalarının Karşılaştırılması. *1. Kocaeli Uluslararası Mühendislik-Doğa Bilimleri ve Sağlık Teknolojileri Konferansı*.