DENGESİZ SINIFLANDIRMA PROBLEMLERİ İÇİN LLM TABANLI OTOMATİK ANALİZ SİSTEMİ

Mehmet Baran ÖZKAN

BİYOİSTATİSTİK VE TIP BİLİŞİMİ

Tez Danışmanı
Doç. Dr. Ahmet Kadir ARSLAN

Yüksek Lisans Tezi – 2026

T.C.
İNÖNÜ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

DENGESİZ SINIFLANDIRMA PROBLEMLERİ İÇİN LLM TABANLI OTOMATİK ANALİZ SİSTEMİ

Mehmet Baran ÖZKAN

BİYOİSTATİSTİK VE TIP BİLİŞİMİ

Yüksek Lisans Tezi

Tez Danışmanı
Doç. Dr. Ahmet Kadir ARSLAN

Tez Jüri Üyeleri
Prof. Dr. Harika Gözde GÖZÜKARA BAĞ
Doç. Dr. Ahmet Kadir ARSLAN
Doç. Dr. Mehmet Onur KAYA

MALATYA
2026

KABUL VE ONAY

T.C.

İNÖNÜ ÜNİVERSİTESİ

Sağlık Bilimleri Enstitüsü Müdürlüğüne

ETİK BEYANI

İnönü Üniversitesi Sağlık Bilimleri Enstitüsü Tez Yazım Kurallarına uygun olarak

Doç. Dr. Ahmet Kadir ARSLAN danışmanlığında hazırlayıp sunduğum Dengesiz Sınıflandırma Problemleri için LLM Tabanlı Otomatik Analiz Sistemi başlıklı Yüksek Lisans tezim içinde elde ettiğim verileri, bilgileri, belgeleri akademik ve etik kurallar çerçevesinde elde ettiğimi, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu, tezimde yararlandığım eserlere bilimsel kurallara uygun atıfta bulunarak kaynak gösterdiğimi, tezimin özgün olduğunu, tezimin çalışma ve yazımında patent ve telif haklarını ihlal edici bir davranışımın olmadığını, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

30/01/ 2026

Mehmet Baran ÖZKAN

İÇİNDEKİLER

TEŞEKKÜR

Tez çalışması sırasında desteklerini esirgemeyen değerli ailem ve özellikle anneme, katkılarını desteklerini eksik etmeyen değerli hocalarım, Doç. Dr. Ahmet Kadir ARSLAN ve Prof. Dr. Cemil ÇOLAK'a teşekkürlerimi sunarım.

ÖZET

Dengesiz Sınıflandırma Problemleri için LLM Tabanlı Otomatik Analiz Sistemi

Amaç: Sağlık alanındaki verilerin karmaşık ve heterojen yapısı, makine öğrenmesi algoritmalarının başarısını doğrudan etkileyen sınıf dengesizliği problemini beraberinde getirmektedir. Bu çalışmanın amacı; sağlık verilerindeki sınıf dengesizliği problemine yönelik en uygun dengeleme yöntemini otomatik olarak seçen ve analiz sonuçlarını Büyük Dil Modelleri (LLM) desteğiyle yorumlayan entegre ve otonom bir sistem geliştirmektir.

Materyal ve Metot: Çalışmada, 1:100'e varan ekstrem dengesizlik oranlarına sahip üç farklı klinik veri seti kullanılmıştır. Geliştirilen otonom analiz hattı; SMOTE türevleri, CTGAN ve Difüzyon Modelleri (TabDDPM) gibi çeşitli dengeleme stratejilerini LLM tabanlı bir döngü ile yönetmektedir. Ayrıca, veri setindeki nadir ancak klinik değeri olan vakaları korumak amacıyla "Sigma Guardrails" adı verilen özgün bir aykırı değer yönetim mekanizması sisteme entegre edilmiştir. Model performansları F1-Skor, MCC, ROC-AUC ve Brier Skoru gibi çok boyutlu metriklerle değerlendirilmiştir.

Bulgular: Elde edilen bulgular, otonom sistemin ekstrem dengesizlik durumlarında dahi başarılı olabileceğini göstermiştir. Ayrıca, Sigma Guardrails protokolü sayesinde verilerdeki istatistiksel aykırıların bazılarının klinik olarak anlamlı bulunarak veri setinde korunduğu saptanmıştır. Düşük Brier skorları, sistemin ürettiği olasılık tahminlerinin klinik güvenilirliğinin yüksek olduğunu göstermiştir.

Sonuç: Sonuç olarak, geliştirilen LLM destekli otonom hat, teknik karmaşıklığı minimize ederek ve yanlış negatif maliyetlerini gözeterek klinik karar destek süreçlerinde etkin bir çözüm sunmaktadır.

Anahtar Kelimeler: SMOTE, Büyük Dil Modelleri, Makine Öğrenmesi, Python, Dengesiz Sınıf Problemi

ABSTRACT

LLM-Based Automatic Analysis System for Imbalanced Classification Problems

Objective: The complex and heterogeneous nature of healthcare data introduces the problem of class imbalance, which directly impacts the performance of machine learning algorithms. The aim of this study is to develop an integrated and autonomous system that automatically selects the optimal balancing method for class imbalance in healthcare data and interprets analysis results with the support of Large Language Models (LLMs).

Material and Methods: Three different clinical datasets with extreme imbalance ratios of up to 1:100 were employed in this study. The developed autonomous analysis pipeline manages various balancing strategies, such as SMOTE derivatives, CTGAN, and Diffusion Models (TabDDPM), through an LLM-based loop. Additionally, a novel outlier management mechanism termed "Sigma Guardrails" was integrated into the system to preserve rare but clinically valuable cases within the dataset. Model performance was evaluated using multidimensional metrics including F1-Score, MCC, ROC-AUC, and Brier Score.

Results: The findings demonstrate that the autonomous system can be successful even under conditions of extreme imbalance. Furthermore, it was determined that through the Sigma Guardrails protocol, some statistical outliers in the data were deemed clinically significant and preserved within the dataset. Low Brier scores indicated the high clinical reliability of the probability estimates generated by the system.

Conclusion: In conclusion, the developed LLM-powered autonomous pipeline offers an effective solution for clinical decision support processes by minimizing technical complexity and accounting for false negative costs.

Keywords: SMOTE, Large Language Models, Machine Learning, Python, Class Imbalance Problem

SİMGELER VE KISALTMALAR DİZİNİ

ADASYN	:	Adaptive Synthetic Sampling (Uyarlamalı Sentetik Örnekleme)
AI	:	Artificial Intelligence (Yapay Zeka)
AMP	:	Automatic Mixed Precision (Otomatik Karma Hassasiyet)
AQR	:	Applied Quantitative Research (Uygulamalı Nicel Araştırma)
AST	:	Abstract Syntax Tree (Soyut Sözdizimi Ağacı)
AUC	:	Area Under Curve (Eğri Altında Kalan Alan)
BMC	:	BioMed Central (BiyoMedya Merkezi)
BMI	:	Body Mass Index (Vücut Kitle İndeksi)
CD	:	Critical Difference (Kritik Fark)
CPU	:	Central Processing Unit (Merkezi İşlem Birimi)
CSV	:	Comma-Separated Values (Virgülle Ayrılmış Değerler)
CTGAN	:	Conditional Tabular GAN (Koşullu Tablosal Üretken Çekişmeli Ağlar)
CUDA	:	Compute Unified Device Architecture (Hesaplama Birleşik Cihaz Mimarisi)
CoT	:	Chain-of-Thought (Zincirleme Düşünce)
DNN	:	Deep Neural Network (Derin Sinir Ağı)
ENN	:	Edited Nearest Neighbours (Düzenlenmiş En Yakın Komşular)
FN	:	False Negative (Yanlış Negatif)
FP	:	False Positive (Yanlış Pozitif)
FPR	:	False Positive Rate (Yanlış Pozitif Oranı)
G-Mean	:	Geometric Mean (Geometrik Ortalama)
GAN	:	Generative Adversarial Networks (Çekişmeli Üretken Ağlar)
GNN	:	Graph Neural Networks (Grafik Sinir Ağları)
GPU	:	Graphics Processing Unit (Grafik İşlem Birimi)
IR	:	İmbalance Ratio (Dengesizlik Oranı)

JSON	:	JavaScript Object Notation (JavaScript Nesne Gösterimi)
KDE	:	Kernel Density Estimation (Tahmin Yoğunluğu Analizi)
KNN	:	k-Nearest Neighbors (k-En Yakın Komşu)
KVKK	:	Kişisel Verilerin Korunması Kanunu
LLM	:	Large Language Models (Büyük Dil Modelleri)
LOF	:	Local Outlier Factor (Yerel Aykırı Değer Faktörü)
LR	:	Logistic Regression (Lojistik Regresyon)
MCC	:	Matthews Correlation Coefficient (Matthews Korelasyon Katsayısı)
MIS	:	Management Information Systems (Yönetim Bilişim Sistemleri)
ML	:	Machine Learning (Makine Öğrenmesi)
MLSDA	:	Machine Learning for Sensory Data Analysis (Duyusal Veri Analizi için Makine Öğrenmesi)
NN	:	Neural Network (Sinir Ağı)
NPV	:	Negative Predictive Value (Negatif Öngörü Değeri)
PCA	:	Principal Component Analysis (Temel Bileşen Analizi)
PR	:	Precision-Recall (Kesinlik-Duyarlılık)
RAG	:	Retrieval-Augmented Generation (Geri-Getirme Destekli Üretim)
RAM	:	Random Access Memory (Rastgele Erişimli Bellek)
REST	:	Representational State Transfer (Temsili Durum Transferi)
RL	:	Reinforcement Learning (Pekiştirmeli Öğrenme)
ROC	:	Receiver Operating Characteristic (Alıcı İşletim Karakteristiği)
ROC-AUC	:	Receiver Operating Characteristic - Area Under Curve (Alıcı İşletim Karakteristiği - Eğri Altında Kalan Alan)
SHAP	:	SHapley Additive exPlanations (SHapley Eklemeli Açıklamaları)
SMOTE	:	Synthetic Minority Over-sampling Technique (Sentetik Azınlık Aşırı Örnekleme Tekniği)
SMOTE-ENN	:	SMOTE - Edited Nearest Neighbours (SMOTE - Düzenlenmiş En Yakın Komşular)

SMOTEN	:	SMOTE Nominal (Nominal Değişkenler için SMOTE)
SMOTENC	:	SMOTE Nominal and Continuous (Nominal ve Sürekli Değişkenler için SMOTE)
SVM	:	Support Vector Machine (Destek Vektör Makinesi)
TF32	:	TensorFloat-32
TN	:	True Negative (Doğru Negatif)
TP	:	True Positive (Doğru Pozitif)
TPR	:	True Positive Rate (Duyarlılık/Recall)
TabDDPM	:	Tabular Denoising Diffusion Probabilistic Models (Tablosal Gürültü Azaltan Yayılım Olasılıksal Modelleri)
UCI	:	University of California, Irvine (Makine Öğrenmesi Veri Havuzu)
VAE-SMOTE	:	Variational Autoencoder - SMOTE (Değişimsel Otomatik Kodlayıcı - SMOTE)
VBGMM	:	Variational Bayesian Gaussian Mixture Model (Değişimsel Bayesci Gauss Karışım Modeli)
VDM	:	Value Difference Metric (Değer Farklılık Metriği)
VRAM	:	Video Random Access Memory (Video Rastgele Erişimli Bellek)
XAI	:	Explainable AI (Açıklanabilir Yapay Zeka)
XGB	:	eXtreme Gradient Boosting (İleri Düzey Gradyan Artırma)

ŞEKİLLER DİZİNİ

Şekil No **Sayfa No**

TABLOLAR DİZİNİ

Tablo No **Sayfa No**

1. GİRİŞ

Son yıllarda bilişim teknolojilerinde yaşanan ivmeli gelişim; veri üretim, işleme ve depolama kapasitelerinde önemli bir dönüşümü beraberinde getirmiştir. Bu dönüşüm, özellikle sağlık, finans ve endüstri gibi stratejik alanlarda büyük ve heterojen veri setlerinin analizini mümkün kılarak Yapay Zeka (Artificial Intelligence - AI) ve Makine Öğrenmesi (Machine Learning - ML) tabanlı karar destek sistemlerinin yaygınlaşmasına zemin hazırlamıştır. Sağlık alanında hastane bilgi sistemleri, tıbbi görüntüleme cihazları ve giyilebilir sensörlerden elde edilen yüksek boyutlu veriler; hastalık teşhisi, risk tahmini ve ilaç keşfi gibi kritik süreçlerde kullanılmaktadır. Ancak, bu verilerin karmaşık, çok modlu (multimodal) ve gürültülü yapısı, geleneksel analiz yöntemlerinin ötesinde, daha gelişmiş ve bütüncül algoritmik yaklaşımların kullanım ihtiyacını artırmaktadır (1).

Sağlık verileri üzerinde yürütülen makine öğrenmesi çalışmalarında karşılaşılan temel zorluklardan biri, sınıf dengesizliği (class imbalance) problemidir. Verinin doğası gereği; nadir görülen hastalıklar, düşük insidanslı yan etkiler veya anomali durumları, sağlıklı popülasyona kıyasla veri setinde çok daha az sayıda örnekle temsil edilmektedir. Bu asimetrik dağılım, standart makine öğrenmesi algoritmalarının çoğunluk sınıfına aşırı uyum (overfitting) sağlamasına, buna karşın klinik açıdan kritik öneme sahip azınlık sınıfı örneklerini (yanlış negatifler) ihmal etme eğilimi göstermesine yol açabilmektedir. Bu durum, modelin genel doğruluk oranı yüksek görünse dahi, hastalığın tespit edilmesindeki başarısının (duyarlılık) beklenen seviyenin altında kalmasına neden olabilmektedir (2).

Literatürde dengesiz veri probleminin çözümüne yönelik olarak veri seviyesi, algoritma seviyesi ve hibrit yaklaşımlar geliştirilmiştir. Veri seviyesindeki çözümlerin temelini, azınlık sınıfı örneklerinin sentetik olarak artırılmasına dayanan SMOTE (Synthetic Minority Over-sampling Technique) ve onun Borderline-SMOTE, ADASYN gibi türevleri oluşturmaktadır. Algoritma seviyesinde ise maliyet duyarlı (cost-sensitive) öğrenme ve Focal Loss gibi sınıflandırma zorluğuna odaklanan kayıp fonksiyonları tercih edilmektedir. Hibrit yöntemler ise SMOTE-Tomek ve SMOTE-ENN örneklerinde olduğu gibi, hem sentetik veri üretimini hem de gürültülü verilerin temizlenmesini (undersampling) eş zamanlı olarak uygulayarak sınıf sınırlarını netleştirmeyi hedeflemektedir (3).

Yapay zeka teknolojilerindeki ilerlemelerle birlikte, dengesizlik probleminin çözümünde derin öğrenme ve üretken modeller (Generative Models) ön plana çıkmaya başlamıştır. Özellikle 2019 yılında önerilen CTGAN (Conditional Tabular GAN) gibi yöntemler, tablosal verilerin koşullu dağılımını öğrenerek daha gerçekçi sentetik veriler üretilmesi potansiyelini sunmuştur (4). Son dönemde ise Grafik Sinir Ağları (Graph Neural Networks - GNN) ve Difüzyon Modelleri (Diffusion Models), veri noktaları arasındaki topolojik ilişkileri ve karmaşık korelasyonları koruyabilme yetenekleri sayesinde, geleneksel SMOTE türevlerine güçlü birer alternatif olarak değerlendirilmektedir. Bu yeni nesil yaklaşımlar, nadir olayların karakteristik özelliklerini koruyarak modelin genelleme yeteneğini artırmayı amaçlamaktadır (5-6).

Ancak, mevcut teknik çeşitliliği, araştırmacılar için "hangi veri setine hangi yöntemin uygulanacağı" sorusunu karmaşık bir optimizasyon problemine dönüştürebilmektedir. Ayrıca, model çıktılarının klinik uzmanlar tarafından anlaşılabilir bir dille yorumlanması (explainability) ihtiyacı güncelliğini korumaktadır. Bu bağlamda, Büyük Dil Modelleri (Large Language Models - LLM); karmaşık analiz süreçlerinin otomatikleştirilmesi, en uygun dengeleme stratejisinin seçilmesi ve elde edilen istatistiksel bulguların araştırmacı dostu bir dille raporlanması noktasında önemli bir potansiyel taşımaktadır (7).

Bu çalışmanın temel amacı; sağlık verilerindeki sınıf dengesizliği problemine yönelik en uygun dengeleme yöntemini (geleneksel, hibrit veya üretken) otomatik olarak seçen ve analiz sonuçlarını Büyük Dil Modelleri desteğiyle yorumlayan entegre bir sistem geliştirmektir. Geliştirilen bu sistem ile araştırmacıların teknik ön işleme süreçlerinde zaman kaybetmesi engellenerek, klinik hipotezlere odaklanmaları ve elde edilen sonuçların doğruluğunun (F1-skoru, AUC vb.) ve güvenilirliğinin artırılması hedeflenmektedir.

Bu kapsamda çalışmanın araştırma sorusu şu şekilde belirlenmiştir: "Büyük dil modelleri (LLM) desteğiyle, sağlık verilerindeki sınıf dengesizliği problemine yönelik en uygun dengeleme yönteminin seçilmesi ve bulguların yorumlanması, geleneksel yaklaşımlara kıyasla sınıflandırma performansını ve klinik yorumlanabilirliği artırır mı?"

Bu soru doğrultusunda test edilecek hipotezler aşağıdadır:

H0 (Sıfır Hipotezi): LLM destekli otonom sistem ile geleneksel yöntemler arasında, sınıflandırma performansı (hassasiyet, özgüllük, F1-skoru, ROC-AUC) ve

analiz sürecindeki zaman verimliliği açısından istatistiksel olarak anlamlı bir fark bulunmamaktadır.

H1 (Alternatif Hipotez): Geliştirilen LLM destekli sınıf dengeleme yöntemi seçimi ve otomatik bulgu yorumlama sistemi; geleneksel yöntemlere kıyasla daha yüksek sınıflandırma performansı sağlamakta ve raporlama sürecini anlamlı ölçüde kısaltarak zaman verimliliği sunmaktadır.

2. GENEL BİLGİLER

2.1. Yapay Zeka ve Yüksek Performanslı Hesaplama Altyapısı

Yapay Zeka (Artificial Intelligence - AI), insan bilişsel süreçlerini modelleyerek karmaşık problemleri çözmeyi hedefleyen, deterministik kurallar yerine olasılıksal (probabilistic) çıkarım mekanizmalarına dayanan disiplinler arası bir alandır. Geleneksel kural tabanlı sistemlerin aksine, modern yapay zeka yaklaşımları veriden öğrenme yeteneği ile teşhis, tedavi planlaması ve risk öngörüsü gibi süreçlerde bir karar destek mekanizması olarak konumlanmaktadır (1). Bu bölümde, çalışmanın teknolojik omurgasını oluşturan öğrenme algoritmaları, derin öğrenme mimarileri ve bu süreçlerin gerektirdiği ileri hesaplama donanımları detaylandırılmıştır.

2.1.1. Veri Tabanlı Öğrenme Süreçleri ve Makine Öğrenmesi Algoritmaları

Makine Öğrenmesi (Machine Learning - ML), algoritmaların açıkça programlanmadan veri setlerindeki gizli örüntüleri, korelasyonları ve doğrusal olmayan ilişkileri öğrenmesini sağlayan istatistiksel bir süreçtir. Mitchell'ın formal tanımıyla; bir bilgisayar programının belirli bir görevdeki (T) performansı (P), deneyim (E) kazandıkça artış gösteriyorsa, bu süreç makine öğrenmesi olarak nitelendirilmektedir (8). Makine öğrenmesi modelleri, eğitim verisi üzerinden optimize edilen ağırlık parametreleri aracılığıyla dinamik karar sınırları oluşturur. Bu dinamik yapı, özellikle tıbbi veriler gibi yüksek boyutlu ve gürültülü veri setlerinde, geleneksel istatistiksel yöntemlere kıyasla daha yüksek genelleme başarısı sunabilmektedir.

2.1.2. Derin Sinir Ağları ve Büyük Dil Modellerinin (LLM) Klinik Uygulama Potansiyeli

Yapay sinir ağlarının çok katmanlı yapısı üzerine kurulan Derin Öğrenme (Deep Learning) (9), özellikle görüntü, ses ve serbest metin gibi yapılandırılmamış verilerin analizinde etkili sonuçlar vermektedir. Bu alandaki en güncel mimari olan Büyük Dil Modelleri (Large Language Models - LLM), milyarlarca parametre ile eğitilen transformatör (transformer) tabanlı yapılardır. İstatistiksel birer motor olarak işlev gören LLM'ler, "Prompt Mühendisliği" teknikleri ile yönlendirilerek, karmaşık tıbbi literatürün özetlenmesi, yapılandırılmamış klinik notlardan veri çıkarımı ve kod üretimi gibi görevlerde bağlama uygun çıktılar üretmek üzere optimize edilebilmektedir (7).

2.1.3. Paralel Hesaplama Mimarisi ve GPU Tabanlı Hızlandırma: CUDA Teknolojisi

Derin öğrenme modellerinin eğitimi, milyonlarca matris çarpımı ve tensör işlemini içeren yoğun bir hesaplama yükü oluşturmaktadır. Bu yükün yönetilmesi için NVIDIA tarafından geliştirilen CUDA (Compute Unified Device Architecture), Grafik İşlem Birimlerinin (GPU) paralel işlem kapasitesini genel amaçlı hesaplamalara açmaktadır. Merkezi İşlem Birimi (CPU) seri işlemler için optimize edilmişken, binlerce çekirdeğe sahip CUDA tabanlı GPU mimarisi, büyük veri kümelerinin işlenmesini CPU'ya kıyasla önemli ölçüde hızlandırarak yapay zeka araştırmalarının donanımsal temelini oluşturmaktadır (10). Donanım düzeyinde, NVIDIA RTX 3090 GPU mimarisinin sunduğu paralel hesaplama kapasitesini maksimize etmek amacıyla, sistem genelinde TensorFloat-32 (TF32) ve Otomatik Karma Hassasiyet (AMP) ayarları manuel olarak yapılandırılarak matris operasyonlarında yüksek verimlilik sağlanmaktadır. Bu kapsamda, hesaplama hızını korumak ve GPU bellek yönetimini optimize etmek için tüm sayısal veriler Float32 hassasiyetine mühürlenmiştir.

2.1.4. Veri Bilimi Ekosistemi ve Yazılım Geliştirme Ortamı: Python

Akademik ve endüstriyel yapay zeka projelerinde, esnekliği, platform bağımsız yapısı ve geniş kütüphane desteği nedeniyle Python programlama dili standart araçlardan biri olarak kabul edilmektedir (11). scikit-learn kütüphanesi klasik makine öğrenmesi algoritmalarının uygulanmasında kullanılırken (12); PyTorch ve TensorFlow gibi çerçeveler, GPU hızlandırmalı derin öğrenme modellerinin geliştirilmesini ve türevlenebilir programlamayı sağlamaktadır (13). Veri manipülasyonu için pandas (14) ve vektörel hesaplamalar için NumPy (15) kütüphanelerinin entegrasyonu, karmaşık veri analizi süreçlerinin yönetimini kolaylaştırmaktadır. Yazılım mimarisinde kararlılığı artırmak amacıyla, NumPy 2.0+ sürümlerindeki kopyalama kısıtlamalarına karşı geliştirilen "copy_compat" middleware mekanizması, ağır kütüphaneler (XGBoost (16), LightGBM (17), Optuna (18) vb.) için tasarlanan "Lazy Loading" (Tembel Yükleme) yapısı ve bellek kullanımını sürekli izleyerek kritik eşiklerde otonom temizlik yapan "Memory Guard" modülü sisteme dahil edilmiştir. Ayrıca, model açıklanabilirliği aşamasında PCA gibi boyut düşürücülerin yarattığı öznitelik uyuşmazlıklarını otomatik olarak hizalayan "Shap Transform Wrapper" sınıfı, yazılımın kararlılık zırhını oluşturmaktadır.

2.2. Sınıf Dengesizliği Probleminin İstatistiksel Tanımı ve Model Performansına Etkileri

Veri madenciliğinde sıkça karşılaşılan dengesiz veri (imbalanced data) problemi, hedef değişkenin sınıfları arasındaki örnek sayılarının belirgin bir asimetri göstermesi durumu olarak tanımlanmaktadır. Matematiksel olarak, N gözlemden oluşan bir veri setinde, azınlık sınıfı ile çoğunluk sınıfı arasındaki oranın (Imbalance Ratio) artması, standart öğrenme algoritmalarının çoğunluk sınıfına yanlı tahminler üretmesine yol açabilmektedir. Bu durum, "Doğruluk Paradoksu"na (Accuracy Paradox) neden olabilir; model %99 doğruluk oranına sahip olsa dahi, klinik açıdan kritik öneme sahip azınlık sınıfı örneklerinin (örneğin pozitif hastalık tanısı) tespit edilememesi riski doğabilir (2). Bu nedenle, model performansının değerlendirilmesinde salt doğruluk yerine F1-Skor, ROC-AUC ve MCC gibi dengesizliğe dirençli metriklerin kullanılması önerilmektedir.

2.3. Veri Seviyesi Dengeleme Stratejileri ve Sentetik Örnekleme Teknikleri

2.3.1. Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) ve Çalışma Prensibi

Sınıf dengesizliğinin giderilmesinde en yaygın kullanılan yaklaşım, Chawla ve ark. tarafından geliştirilen SMOTE (Synthetic Minority Over-Sampling Technique) algoritmasıdır. Bu yöntem, verilerin kopyalanmasına dayanan "Random Oversampling" yaklaşımının neden olduğu aşırı öğrenme (overfitting) sorununu hafifletmek amacıyla geliştirilmiştir. SMOTE, özellik uzayında interpolasyon yaparak yeni bilgi üretme prensibine dayanır. Algoritma, bir azınlık örneği ile onun k-en yakın komşusu (k-NN) arasından seçilen bir komşu arasına sanal bir doğru çizer ve bu doğru üzerinde rastgele bir noktada yeni bir sentetik veri oluşturur . Bu strateji, karar sınırlarını genişleterek modelin genelleme kapasitesini artırmayı hedefler (19). Klinik veri setlerinin kısıtlı örneklem boyutu göz önüne alındığında, sistem; azınlık sınıfı sayısına göre k-parametresini dinamik olarak aşağı çeken bir güvenlik katmanı içermektedir. Bu dinamik yapı, azınlık örneği sayısının (n) belirlenen k-değerinden az olması durumunda, k değerini otomatik olarak (n−1) şeklinde revize ederek analizin sürekliliğini sağlamaktadır.

2.3.2. Sınır Tabanlı Sentetik Örnekleme Yaklaşımı: Borderline-SMOTE

Standart SMOTE algoritmasının tüm azınlık örneklerine eşit ağırlık vermesi, gürültülü verilerin veya aykırı değerlerin de çoğaltılmasına neden olabilmektedir. Bu kısıtlamayı aşmak için geliştirilen Borderline-SMOTE, azınlık örneklerini "Güvenli", "Tehlikeli" ve "Gürültü" olarak sınıflandırır. Sınıflandırma sınırına (decision boundary) yakın olan ve yanlış sınıflandırılma riski en yüksek olan "Tehlikeli" kategorisindeki örnekler hedef alınarak, sentetik veri üretimi yalnızca bu kritik sınır bölgelerinde yoğunlaştırılır. Böylece, sınıflar arasındaki ayrım çizgisinin daha net hale getirilmesi amaçlanır (20). Sistem, bu sınır tespiti sırasında k-komşuluk parametresini verinin yerel yoğunluğuna göre dinamik olarak optimize ederek sınır bölgelerindeki aşırı gürültü birikimini engellemektedir.

2.3.3. Adaptif Yoğunluk Temelli Sentetik Örnekleme: ADASYN

Veri dağılımının yerel yoğunluğunu dikkate alan ADASYN (Adaptive Synthetic Sampling) algoritması, öğrenilmesi zor olan örneklere öncelik veren dinamik bir yaklaşım sunmaktadır. Her bir azınlık örneği için, o örneğin etrafındaki çoğunluk sınıfı yoğunluğuna göre bir ağırlık skoru hesaplanır. Çoğunluk sınıfı örnekleriyle çevrili olan, yani ayırt edilmesi zor azınlık örnekleri için daha fazla sayıda sentetik veri üretilerek, modelin zor örnekleri öğrenmeye odaklanması sağlanır. Bu yöntem, veri dağılımını dengelerken aynı zamanda sınıf içi varyansı da artırabilmektedir (21).

2.4. Karma ve Kategorik Veri Yapıları İçin Özelleştirilmiş SMOTE Türevleri

2.4.1. Sürekli ve Nominal Veri Entegrasyonu: SMOTENC

Standart SMOTE, Öklid mesafesine dayandığı için yalnızca sürekli (sayısal) verilerde uygulanabilmektedir. Gerçek hayat verilerinin hem sayısal hem de kategorik (nominal) özellikler içerdiği durumlarda SMOTENC (SMOTE Nominal and Continuous) yöntemi kullanılabilmektedir. Bu algoritma, sayısal değişkenler için interpolasyon yaparken, kategorik değişkenler için en sık görülen değeri (mode) atayarak veya değer farkı metriğini kullanarak hibrit veri setlerinde tutarlı sentetik örnekler üretmeyi hedefler (19).

2.4.2. Kategorik Veri Tabanlı Örnekleme: SMOTEN

Veri setinin tamamen kategorik değişkenlerden oluştuğu senaryolarda SMOTEN (SMOTE Nominal) algoritması tercih edilmektedir. Sürekli bir uzayda interpolasyonun mümkün olmadığı bu durumda, Değer Farkı Metriği (Value Difference Metric - VDM) kullanılarak kategorik değerler arasındaki çok boyutlu benzerlikler hesaplanır. Bu matris üzerinden, mevcut kategorik dağılımı bozmayacak şekilde yeni sentetik profiller türetilmeye çalışılır.

2.5. Örnekleme ve Temizleme Tabanlı Hibrit Dengeleme Algoritmaları

2.5.1. Sentetik Üretim ve Tomek Bağlantıları ile Temizleme: SMOTETomek

Sentetik veri üretimi, bazı durumlarda sınıflar arası örtüşmeyi (overlapping) artırarak karar sınırlarını bulanıklaştırabilir. Bu sorunu gidermek için SMOTE ile Tomek Links yönteminin birleştirildiği SMOTETomek algoritması geliştirilmiştir. Süreç, önce SMOTE ile azınlık sınıfının artırılmasıyla başlar; ardından birbirine en yakın ancak zıt sınıflara ait veri çiftleri (Tomek Links) tespit edilerek veri setinden çıkarılır. Bu işlem, sınıflar arasında daha net bir güvenlik marjı oluşturulmasını hedefler (3).

2.5.2. Gürültü Eliminasyonu ve Sentetik Dengeleme: SMOTEENN

Benzer bir hibrit strateji izleyen SMOTEENN, sentetik üretim aşamasından sonra Edited Nearest Neighbours (ENN) algoritmasını uygular. ENN, bir örneğin sınıf etiketinin, k-en yakın komşularının çoğunluğundan farklı olması durumunda o örneği gürültü veya aykırı değer olarak değerlendirip siler. Bu yaklaşım, SMOTETomek'e göre daha agresif bir veri temizliği yaparak, hem veri dengesizliğini gidermeyi hem de veri setini gürültüden arındırarak sınıflandırma performansını artırmayı amaçlar (3).

2.6. Derin Öğrenme Tabanlı Üretken Modeller ve Sentetik Veri Sentezi

Geleneksel istatistiksel yöntemlerin yetersiz kaldığı yüksek boyutlu veri uzaylarında, derin öğrenme tabanlı üretken modeller kullanılmaktadır. CTGAN (Conditional Tabular GAN), Generative Adversarial Networks (GAN) mimarisini kullanarak tablosal verilerin koşullu olasılık dağılımını modeller ve "mod çökmesi" (mode collapse) sorununu aşmak için özel normalizasyon teknikleri uygular (4). CTGAN mimarisi bünyesinde, verideki mod-spesifik dağılımları korumak amacıyla her

bir sütun için varyasyonel Bayesyen Gaussian Karışım Modelleri (VBGMM) kullanılmaktadır. VAE-SMOTE, veriyi düşük boyutlu bir gizli uzaya (latent space) sıkıştırıp burada örnekleme yaparak gürültüden arındırılmış veriler üretirken; Diffusion Sampling (Difüzyon Modelleri), veriye kademeli gürültü ekleyip bu süreci tersine çevirerek, verinin özgün dağılımından yüksek kaliteli örnekler sentezlenmesini sağlar (6). Sistemde yer alan TabDDPM (Tabular Diffusion) modeli, 1000 adımlık Gaussian difüzyon süreci ile verinin marjinal dağılımlarını yüksek hassasiyetle taklit etmektedir. Üretken modellerin başarısı, yalnızca sınıflandırma metrikleriyle değil, üretilen verinin orijinal veriyle olan istatistiksel benzerliğini ve öznitelikler arası korelasyon korunumunu ölçen "Sadakat Skoru" (Fidelity Score) ile sağlanmaya çalışılmaktadır.

2.7. Algoritmik Seviye İyileştirmeler: Maliyet Duyarlı Öğrenme ve Topluluk Stratejileri

2.7.1. Hata Maliyet Matrisi ve Maliyet Duyarlı (Cost-Sensitive) Öğrenme

Veri seviyesinde müdahale yerine algoritmanın öğrenme sürecini değiştirmeyi hedefleyen bu yaklaşım, hata maliyet fonksiyonunu (loss function) yeniden tanımlar. Yanlış negatiflere (FN), yanlış pozitiflerden (FP) daha yüksek bir ceza puanı atanarak, modelin eğitim sırasında azınlık sınıfına odaklanması ve toplam beklenen maliyeti minimize etmesi hedeflenir (22). Klinik risk yönetimini desteklemek amacıyla sistem, kullanıcı tarafından belirlenen FN maliyet oranına göre karar eşiğini otonom olarak optimize eden bir döngü içermektedir.

2.7.2. Zor Örnek Madenciliği ve Odaklanmış Kayıp Fonksiyonu: Focal Loss

Dengesiz veri setlerinde standart Çapraz Entropi (Cross-Entropy) kaybı, çoğunluk sınıfının kolay örnekleri tarafından domine edilebilmektedir. Focal Loss, bu dengesizliği yönetmek için kayıp fonksiyonuna bir modülasyon faktörü $(1-pt)\gamma$ ekler. Bu faktör, modelin yüksek güvenle sınıflandırdığı kolay örneklerin etkisini logaritmik olarak azaltırken, yanlış sınıflandırılan ve ayırt edilmesi zor olan örneklere (hard examples) odaklanılmasını matematiksel olarak teşvik eder (23).

2.7.3. Topluluk Öğrenmesi (Ensemble Learning) ve Yığınlama Stratejileri

Topluluk Öğrenmesi (Ensemble Learning), tek bir modelin varyans ve sapma hatalarını dengelemek amacıyla birden fazla öğrenme algoritmasının birleştirilmesi yöntemidir. Stacking (Yığınlama) (24), farklı tipteki modellerin tahminlerini bir üst-

model (meta-learner) ile birleştirirken; Weighted Voting (Ağırlıklı Oylama), modellerin performansına göre ağırlıklı karar verilmesini sağlar. Geliştirilen sistemde "WeightedEnsemble" katmanı, her bir modelin F1-skorunu toplam performansa oranlayarak modellere dinamik ağırlıklar atamaktadır. Ayrıca, BalancedBagging gibi yöntemler, eğitim verisini alt kümelere ayırıp her birinde dengeleme yaparak F1-skor ve ROC-AUC performansını maksimize etmeyi amaçlar (25). Tablo 2.1.'de dengesiz veri probleminin çözümünde kullanılan yöntemler sınıflandırılmış ve karşılaştırılmıştır.

Tablo 2.1. Dengesiz veri probleminin çözümünde kullanılan bazı yöntemlerin sınıflandırılması ve karşılaştırılması

Kategori	Yöntem	Temel Mekanizma	Avantajları	Kısıtlılıkları
Aşırı Örnekleme (Oversampling)	SMOTE	k-NN tabanlı doğrusal interpolasyon ile sentetik örnek üretir.	Overfitting'i azaltabilir, uygulanması basittir.	Gürültülü örnekleri çoğaltabilir, sınıf örtüşmesini artırabilir.
	Borderline-SMOTE	Sadece karar sınırındaki "tehlikeli" örnekleri çoğaltır.	Karar sınırlarını netleştirebilir, daha hedefli örnekleme yapar.	k parametresine duyarlıdır, sınır tespiti zor olabilir.
	ADASYN	Öğrenmesi zor örneklere adaptif olarak daha fazla sentetik veri üretir.	Zor örneklerin öğrenilmesini teşvik eder, veri dağılımını dengeler.	Aykırı değerlere karşı hassastır.
Özel Veri Tipleri	SMOTENC	Sürekli ve nominal verileri hibrit şekilde işler.	Karışık (mixed-type) veri setlerinde kullanılabilir.	Nominal değişkenlerde çeşitlilik sınırlı olabilir.
	SMOTEN	VDM metriği ile kategorik benzerlik kurar ve sentetik örnek üretir.	Tamamen kategorik veriler için optimize edilmiştir.	Hesaplama maliyeti yüksek olabilir.
Hibrit Yöntemler	SMOTETomek	SMOTE + Tomek Links ile sınır temizliği yapar.	Sınıf örtüşmesini azaltabilir, daha temiz karar sınırı sağlar.	Veri kaybı riski vardır, işlem süresi uzundur.
	SMOTEENN	SMOTE + ENN ile gürültü temizliği yapar.	Gürültüye karşı daha dirençli olabilir.	Çok sayıda örnek silinebilir, bilgi kaybı olabilir.
Derin Öğrenme / Üretken	CTGAN	Koşullu GAN mimarisi ve mod-spesifik normalizasyon ile veri üretir.	Gerçeğe yakın dağılım üretir, karmaşık ilişkileri öğrenebilir.	Eğitimi zordur, hiperparametre ayarı gerektirir.
Algoritmik	Focal Loss	Kolay örneklerin ağırlığını düşürerek zor örneklere odaklanır (modülasyon terimi γ).	Veri artırmaya gerek kalmadan model odaklanmasını sağlar.	γ parametresinin optimizasyonu zordur.
	Cost-Sensitive Learning	Hata maliyet matrisi tanımlayarak modeli kritik hatalara duyarlı hale getirir.	Özellikle FN gibi kritik hataları minimize etmeye yardımcı olur.	Maliyet değerlerinin belirlenmesi uzmanlık gerektirir.
Derin Öğrenme (DNN Tabanlı)	DNN (Deep Neural Network)	Çok katmanlı yapay sinir ağı ile doğrusal olmayan karmaşık örüntüleri öğrenir.	Yüksek modelleme gücü, karmaşık veri yapılarında başarılıdır.	Büyük veri ve yüksek hesaplama gücü gerektirir, overfitting riski vardır.
	BP (Backpropagation)	Hata geri yayılımı ile gradyan tabanlı ağırlık güncelleme yapar.	DNN eğitiminde temel ve etkilidir.	Yerel minimumlara takılabilir, başlangıç ağırlıklarına duyarlıdır.
Meta-Sezgisel Optimizasyon (DNN ile Entegrasyonlu)	PSO (Particle Swarm Optimization)	Parçacık sürüsü ile global arama yaparak ağırlık/hiperparametre optimizasyonu yapar.	Global optimuma yaklaşma potansiyeli yüksektir, türev gerektirmez.	Hesaplama maliyeti yüksektir, parametre ayarı hassastır.
	GA (Genetic Algorithm)	Seçilim, çaprazlama ve mutasyon ile evrimsel optimizasyon yapar.	Güçlü global arama yeteneği vardır.	Yavaş yakınsama olabilir, nesil sayısı arttıkça maliyet artar.
	SA (Simulated Annealing)	Olasılıksal kabul mekanizması ve soğuma planı ile optimizasyon yapar.	Yerel minimumdan kaçabilir, uygulanması basittir.	Soğuma planına duyarlıdır, büyük veri setlerinde yavaş olabilir.
	GV (Genetic Variation)	Popülasyon tabanlı genetik çeşitlilik mekanizmaları ile ağırlık/hiperparametre optimizasyonu yapar.	Çeşitliliği artırarak erken yakınsamayı azaltabilir, global arama kabiliyeti yüksektir.	Hesaplama maliyeti yüksektir, varyasyon oranlarının ayarlanması hassastır.

2.8. Performans Değerlendirme Ölçütleri ve Model Doğrulama

Makine öğrenmesi modellerinin başarısı, problemin yapısına ve veri setinin dağılımına uygun metriklerle ölçülmelidir. Özellikle sınıf dağılımının dengesiz olduğu tıbbi veri setlerinde, modelin klinik geçerliliğini kanıtlamak için tek bir ölçüt yeterli olmamakta, çok boyutlu bir değerlendirme matrisi gerekmektedir. Performans ölçümü, temel olarak "Karmaşıklık Matrisi" (Confusion Matrix) adı verilen ve modelin tahminlerini Doğru Pozitif (TP), Doğru Negatif (TN), Yanlış Pozitif (FP) ve Yanlış Negatif (FN) olarak sınıflandıran tablo üzerinden hesaplanan istatistiksel değerlere dayanır (26).

2.8.1. Temel Sınıflandırma Metrikleri: Doğruluk ve Kesinlik Ölçümleri

Genel başarıyı ifade eden Doğruluk (Accuracy), doğru tahmin edilen örneklerin toplam örnek sayısına oranı olarak hesaplanır. Ancak dengesiz veri setlerinde, modelin tüm örnekleri çoğunluk sınıfına ataması durumunda dahi yüksek doğruluk elde edilebildiği için (Accuracy Paradox), bu metrik tek başına yanıltıcı kabul edilebilmektedir. Bu nedenle sınıf bazlı performansın incelenmesi gerekir. Kesinlik (Precision), modelin pozitif olarak tahmin ettiği örneklerin ne kadarının gerçekten pozitif olduğunu gösterir ve yanlış alarmların (FP) maliyetinin yüksek olduğu durumlarda önem kazanır. Diğer yandan Duyarlılık (Recall), gerçek pozitiflerin ne kadarının model tarafından yakalanabildiğini ölçer. Hayati risk taşıyan hastalıkların teşhisinde, bir vakanın kaçırılmasının (FN) maliyeti çok yüksek olduğundan, Recall değerinin maksimize edilmesi genellikle öncelikli hedeftir.

2.8.2. Negatif Sınıf Performansı: Özgüllük ve Negatif Öngörü Değeri

Pozitif sınıf kadar negatif sınıfın başarısı da klinik karar süreçlerinde kritiktir. Özgüllük (Specificity), gerçek negatiflerin (sağlıklı bireylerin) model tarafından ne oranda doğru tespit edildiğini gösterir. Dengesiz veri setlerinde çoğunluk sınıfının başarısını temsil eden bu metrik, yanlış pozitif oranının kontrol altında tutulmasını sağlar. Negatif Öngörü Değeri (NPV) ise, modelin "negatif" (örneğin hastalık yok) dediği durumlarda bu tahminin güvenilirliğini ifade eder. Yüksek NPV değeri, modelin ekarte etme (rule-out) gücünün yüksek olduğunu gösterir ve gereksiz ileri tetkiklerin önlenmesi açısından değerlidir.

2.8.3. Bileşik ve Harmonik Ortalamalar: F-Skorları ve G-Mean

Kesinlik ve Duyarlılık arasındaki dengeyi sağlamak amacıyla F1-Skor metriği kullanılır. Precision ve Recall değerlerinin harmonik ortalaması olan F1-Skor, her iki metriğe eşit ağırlık verir. Ancak sağlık uygulamalarında yanlış negatiflerin maliyeti daha yüksek olduğundan, Recall değerine Precision'dan daha fazla ağırlık veren F2-Skor kullanımı literatürde önerilmektedir. F2-Skor, $\beta=2$ parametresi ile duyarlılığı önceler ve azınlık sınıfının kaçırılmasını daha ağır cezalandırır. Sınıflar arası dengeyi gözetmek için kullanılan bir diğer metrik olan Geometrik Ortalama (G-Mean), Duyarlılık ve Özgüllük değerlerinin çarpımının karekökü olarak hesaplanır. G-Mean, her iki sınıfın doğruluğunun aynı anda yüksek olmasını talep eder ve çoğunluk sınıfına aşırı uyumu (overfitting) cezalandırmayı amaçlar. Sistemin otonom model sıralama mekanizmasında kullanılan "Bileşik Puan" (Composite Score), klinik başarının çok boyutlu bir sentezi olarak; F1-Skor (%40), ROC-AUC (%40) ve G-Mean (%20) metriklerinin ağırlıklı toplamı üzerinden hesaplanmaktadır.

2.8.4. Dengesiz Veriler İçin Sağlam Metrikler: MCC, Kappa ve Balanced Accuracy

Dengesiz veri setlerinde matematiksel olarak en sağlam (robust) metrik olarak kabul edilen Matthews Korelasyon Katsayısı (MCC), karmaşıklık matrisinin dört bileşenini de (TP, TN, FP, FN) hesaplamaya katar. -1 ile +1 arasında değer alan MCC, +1 değerinde mükemmel tahmini, 0 değerinde rastgele tahmini ifade eder. Sınıf boyutları ne kadar farklı olursa olsun, MCC değeri yalnızca model her iki sınıfta da başarılıysa yüksek çıkar (27). Benzer şekilde, Cohen's Kappa katsayısı, modelin başarısının şans faktöründen ne kadar arınmış olduğunu ölçerek, gözlemlenen doğruluk ile beklenen rastgele doğruluk arasındaki uyumu test eder. Dengeli Doğruluk (Balanced Accuracy) ise, Duyarlılık ve Özgüllük değerlerinin aritmetik ortalamasıdır; bu sayede azınlık sınıfının performansı genel skor üzerinde çoğunluk sınıfı kadar etkili hale getirilir.

2.8.5. Olasılıksal ve Eşik-Bağımsız Değerlendirme: ROC-AUC ve Brier Skoru

Modelin karar eşiğinden (threshold) bağımsız genel performansını ölçmek için ROC-AUC (Receiver Operating Characteristic - Area Under Curve) kullanılır. ROC eğrisi, farklı eşik değerlerinde Duyarlılık (TPR) ile Yanlış Pozitif Oranı (FPR)

arasındaki ilişkiyi gösterir. Eğri altında kalan alanın (AUC) 1.0 olması mükemmel ayrımı, 0.5 olması ise rastgele sınıflandırmayı ifade eder.

ROC-AUC, modelin pozitif ve negatif örnekleri birbirinden ayırma discrimination yeteneğini ölçerken; modelin ürettiği olasılık değerlerinin gerçekliğe ne kadar yakın olduğunu (calibration) ölçerken Brier Skoru kullanılır. Tahmin edilen olasılıklar ile gerçek sonuçlar arasındaki karesel hata ortalaması olan Brier Skoru, 0'a yaklaştıkça modelin olasılıksal güvenilirliğinin arttığını gösterir. Klinik risk hesaplamalarında modelin sadece sınıfı değil, risk olasılığını da doğru vermesi beklendiğinden Brier Skoru kritik bir tamamlayıcıdır (28). Özellikle hayati risklerin minimize edilmesi gereken durumlarda, sistem "fn_cost_ratio" parametresini temel alarak modelin karar eşiğini (default 0.5) otonom olarak revize etmekte ve toplam klinik riski minimize eden en uygun eşik değerini saptamaktadır.

2.8.6. Performans Metriklerinin Karşılaştırmalı Analizi

Makine öğrenmesi modellerinin değerlendirilmesinde kullanılan her bir metriğin matematiksel odak noktası ve uygulama alanı farklılık göstermektedir. Özellikle dengesiz veri setlerinde ve hayati risk taşıyan klinik karar destek sistemlerinde, doğru metriğin seçilmesi modelin güvenilirliğini doğrudan etkilemektedir. Literatürde tanımlanan ölçütlerin kullanım durumları ve klinik önemleri Tablo 2.2’de özetlenmiştir.

Tablo 2.2. Sınıflandırma modellerinin değerlendirilmesinde kullanılan performans metrikleri ve klinik kullanım senaryoları.

Performans Ölçütü	Odak Noktası	Kullanım Senaryosu ve Önemi
Doğruluk (Accuracy)	Genel Tahmin Başarısı	Sınıf dağılımlarının eşit olduğu (Dengeli) veri setlerinde kullanılır. Dengesiz verilerde "Doğruluk Paradoksu" nedeniyle yanıltıcı olabilir; dikkatli kullanılmalıdır.
Duyarlılık (Recall / Sensitivity)	Gerçek Pozitiflerin Tespiti	Yanlış Negatiflerin (FN) maliyetinin çok yüksek olduğu durumlarda (Kanser taraması, salgın tespiti) birincil metriktir.
Kesinlik (Precision)	Pozitif Tahmin Güveni	Yanlış Pozitiflerin (FP) maliyetinin yüksek olduğu durumlarda (Gereksiz riskli ameliyat, spam filtresi) tercih edilir.
F1-Skor	Precision ve Recall Dengesi	Sınıf dengesizliği olduğunda ve hem yanlış pozitiflerin hem de yanlış negatiflerin önemli olduğu durumlarda kullanılır. Harmonik ortalamadır.
F2-Skor	Duyarlılık Ağırlıklı Denge	F1-Skor'a benzer ancak Recall'a daha fazla ağırlık verir. Tıbbi teşhiste, hastalığı kaçırmamanın yanlış alarmdan daha riskli olduğu durumlarda idealdir.
Özgüllük (Specificity)	Gerçek Negatiflerin Tespiti	Sağlıklı bireylerin yanlışlıkla hasta ilan edilmemesi istendiğinde kullanılır. Tarama testlerinin doğrulama aşamasında kritiktir.
Negatif Öngörü Değeri (NPV)	Negatif Tahmin Güveni	Bir hastalığı ekarte etme (rule-out) gücünü gösterir. Yüksek NPV, "Test negatifse hasta kesinlikle sağlıklıdır" güvenini destekler.
ROC-AUC	Ayırt Etme (Discrimination)	Modelin eşik değerinden (threshold) bağımsız genel başarısını ölçer. Farklı modellerin genel performansını kıyaslamak için standarttır.
Matthews Korelasyon Katsayısı (MCC)	Matematiksel Sağlamlık	Dengesiz veri setleri için en güvenilir tekil metriklerden biri kabul edilir. Dört karmaşıklık matrisi hücresini de hesaba katar.
Cohen’s Kappa	Rastlantısallık Kontrolü	Modelin başarısının şans eseri olup olmadığını test eder. Gözlemciler arası uyumun veya model-veri uyumunun istatistiksel kanıtıdır.
Dengeli Doğruluk (Balanced Accuracy)	Sınıf Ortalaması	Duyarlılık ve Özgüllüğün aritmetik ortalamasıdır. Azınlık sınıfının başarısını genel skora yansıttığı için dengesiz veride Accuracy yerine kullanılır.
Geometrik Ortalama (G-Mean)	Sınıf Dengesi	Çoğunluk sınıfına aşırı uyumu (overfitting) cezalandırır. Hem pozitif hem negatif sınıfta aynı anda yüksek başarı hedeflendiğinde kullanılır.
Brier Skoru	Olasılık Kalibrasyonu	Modelin sadece sınıfı değil, o sınıfa ait olasılık değerini ne kadar doğru tahmin ettiğini ölçer. Risk skorlaması (ör. ölüm riski %) için önemlidir.

2.9. Dengesiz Veri Literatürünün Tarihsel Gelişimi ve İlgili Çalışmalar

Dengesiz veri problemi üzerine yapılan araştırmalar, son yirmi yılda basit veri seviyesi müdahalelerden, karmaşık derin öğrenme mimarilerine ve otonom sistemlere doğru belirgin bir evrim geçirmiştir. Bu süreçte geliştirilen yöntemler, önceki çalışmaların kısıtlılıklarını gidermeyi ve veri setlerindeki heterojenliği daha iyi modellemeyi hedeflemiştir.

2.9.1. Temel Yaklaşımlar ve Veri Seviyesi Çözümler (2000-2015)

Dengesiz veri probleminin çözümüne yönelik ilk sistematik çalışmalar, verinin yeniden örneklenmesi (resampling) üzerine yoğunlaşmıştır. Chawla ve ark. tarafından 2002 yılında önerilen SMOTE (Synthetic Minority Over-sampling Technique), azınlık sınıfı örnekleri arasında doğrusal interpolasyon yaparak sentetik veri üretmiş ve bu alanda bir standart haline gelmiştir. Chawla'nın çalışması, rastgele örneklemenin (Random Oversampling) neden olduğu aşırı öğrenme (overfitting) problemini hafifletse de, SMOTE'un gürültülü verilere karşı kör olması ve sınıf sınırlarını bulanıklaştırması yeni problemleri beraberinde getirmiştir (19). Bu sorunları aşmak amacıyla, Batista ve ark. (2004), temizleme algoritmaları ile örnekleme tekniklerini birleştiren hibrit yaklaşımları incelemiştir. Yapılan karşılaştırmalı analizlerde, SMOTE-Tomek ve SMOTE-ENN gibi yöntemlerin, hem sentetik veri üretip hem de gürültülü örnekleri temizleyerek sınıflandırma performansını artırabildiği ve sınıflar arası çakışmayı (overlap) azaltabildiği rapor edilmiştir (3). Aynı dönemde Han ve ark. (2005) sınır bölgelerine odaklanan Borderline-SMOTE'u (20), He ve ark. (2008) ise zorluk derecesine göre adaptif örnekleme yapan ADASYN yöntemini geliştirerek, veri dağılımının yoğunluğuna dayalı stratejilerin, kör örneklemeye kıyasla F1-skorunu anlamlı ölçüde iyileştirme potansiyeli taşıdığını ortaya koymuşlardır (21).

2.9.2. Algoritmik İyileştirmeler ve Topluluk Öğrenmesi (2016-2018)

Veri seviyesindeki gelişmelerin ardından araştırmalar, algoritmaların dengesizliğe karşı direncini artırmaya yönelmiştir. Krawczyk (2016) tarafından yapılan kapsamlı incelemelerde, tekil sınıflandırıcılar yerine Topluluk Öğrenmesi (Ensemble Learning) yöntemlerinin etkinliği analiz edilmiştir (29). Bu çalışmalarda, veri dengesizliğinin giderilmesinde Bagging ve Boosting temelli toplulukların, özellikle gürültülü verilerde daha kararlı sonuçlar verdiği gösterilmiştir. Derin öğrenme

algoritmalarının yükselişiyle birlikte, çözüm arayışları özel kayıp fonksiyonlarına kaymıştır. Lin ve ark. (2017) tarafından nesne tespiti alanında önerilen Focal Loss, sınıflandırması kolay örneklerin etkisini azaltıp zor örneklere odaklanarak, model eğitim sürecini dengesizliğe karşı optimize etmeye yöneliktir. Bu çalışma, veri artırmaya gerek kalmadan algoritma seviyesinde dengesizliğin yönetilebileceğini göstermesi açısından literatürde önemli bir yerde bulunmaktadır (23).

2.9.3. Üretken Modeller ve Derin Öğrenme Tabanlı Sentez (2019-2022)

Geleneksel interpolasyon (SMOTE) tekniklerinin, verinin karmaşık istatistiksel dağılımını ve değişkenler arası doğrusal olmayan ilişkileri modellemede yetersiz kalması, araştırmacıları üretken modellere (Generative Models) yöneltmiştir. Xu ve ark. (2019) tarafından geliştirilen CTGAN (Conditional Tabular GAN), Generative Adversarial Networks (GAN) mimarisini tablosal verilere uyarlayarak, mod çökmesi sorununu aşmış ve SMOTE türevlerine kıyasla gerçeğe daha yakın sentetik veriler üretme potansiyelini sunmuştur (4). Bu dönemde yapılan çalışmalar, GAN tabanlı yöntemlerin özellikle yüksek boyutlu ve karmaşık sağlık verilerinde AUC ve Hassasiyet metriklerinde geleneksel yöntemlere göre üstünlük sağlayabildiğini göstermiştir.

2.9.4. Modern Dönem: Difüzyon Modelleri, GNN ve Otonom Sistemler (2023-Günümüz)

Literatürdeki en güncel çalışmalar, verinin sadece istatistiksel dağılımını değil, topolojik yapısını da koruyan yöntemlere odaklanmaktadır (30). Shi ve ark. (2023), Grafik Sinir Ağları (GNN) tabanlı yaklaşımların, veri noktaları arasındaki komşuluk ilişkilerini graf yapısı üzerinde modelleyerek, düğüm seviyesindeki dengesizliklerde geleneksel yöntemlere göre daha üstün genelleme yeteneği sunabileceğini rapor etmiştir (31). Bu kapsamda, klinik veriler arasındaki geometrik ilişkileri optimize etmek amacıyla geliştirilen "Hibrit Komşuluk" (Hybrid Adjacency) yaklaşımı, statik KNN grafiği ile latent uzayda öğrenilen dinamik bağlantıları 0.7 oranla statik olmayan ve 0.3 oranla dinamik olmayan oranlarda harmanlayarak model performansını artırmaktadır. Buna paralel olarak, üretken yapay zeka alanındaki gelişmeler Difüzyon Modellerini (Diffusion Models) ön plana çıkarmıştır. Kotelnikov ve ark. tarafından geliştirilen TabDDPM gibi modeller, veriye kademeli gürültü ekleyip geri dönüştürme süreciyle, SMOTE ve GAN türevlerinden daha kaliteli ve çeşitliliği yüksek sentetik veriler üretme vaadindedir (6). Ayrıca, Altalhan ve ark. (2024) tarafından incelenen AutoSMOTE gibi

otonom yaklaşımlar, insan müdahalesine gerek kalmadan veri setinin karakteristiğine en uygun örnekleme stratejisini belirleyen uçtan uca öğrenme sistemlerinin başarım potansiyelini araştırmaktadır (32). Modern otonom sistemlerde, strateji seçimini deneyimsel olarak optimize eden Q-Öğrenme ajanları, F1-skor (%40), MCC (%30), AUC (%20) ve Kesinlik (%10) ağırlıklı ödül fonksiyonlarını kullanarak klinik kararlılığı maksimize etmeye çalışmaktadır.

2.9.5. Literatürdeki Boşluk ve Çalışmanın Konumu

Tarihsel süreç incelendiğinde, dengesiz veri problemi için geliştirilen yöntemlerin başarısının veri setine özgü olduğu ve "tek ve evrensel" bir yöntemin bulunmadığı görülmektedir. Geleneksel yöntemler hesaplama maliyeti açısından avantajlıyken, modern difüzyon ve GNN tabanlı modeller yüksek performans ancak yüksek işlem gücü gerektirmektedir. Mevcut literatürde, bu yöntem çeşitliliği içerisinde en uygun stratejiyi seçen, sonuçları klinik bağlamda yorumlayan ve süreci otomatize eden bütüncül bir sisteme ihtiyaç duyulmaktadır. Bu tez çalışması, Büyük Dil Modellerini (LLM) karar destek mekanizması olarak sürece dahil ederek, teknik performans ile klinik yorumlanabilirliği birleştirmeyi hedeflemektedir.

3. MATERYAL VE METOT

Bu bölümde, araştırmanın bilimsel temelleri, veri setlerinin seçim kriterleri, kullanılan yazılım mimarisinin teknik ayrıntıları ve geliştirilen otonom analiz hattının metodolojik süreçleri detaylandırılmıştır.

3.1. Araştırmanın Tipi ve Tasarımı

Aykırı Değer Tespit Modülünde Kullanılan Algoritmalar ve Çalışma Prensipleri] Bu çalışma, sınıf dağılımı açısından dengesizlik gösteren klinik veri setlerinin analizinde, Büyük Dil Modelleri (LLM) entegreli otonom bir makine öğrenmesi hattının (pipeline) etkinliğini değerlendiren deneysel, retrospektif ve karşılaştırmalı bir araştırmadır. Araştırma, metodolojik olarak "tasarım bilimi" (design science) ve "deneysel bilişim" prensipleri üzerine inşa edildi (33-35). Çalışma kurgusu; önerilen otonom sistemin başarısını, klasik yöntemler, hibrit yaklaşımlar, üretken modeller ve meta-sezgisel optimizasyon teknikleri (36) kullanarak LLM yönlendirmesi ile analiz etmek üzerine şekillendirildi. Araştırma sürecinde; veri yükleme, otomatik ön işleme, uç değer analizi, dengeleme stratejisinin seçimi, Derin Sinir Ağı (DNN) optimizasyonu ve topluluk öğrenmesi (ensemble) aşamalarından oluşan bütünleşik ve yinelemeli bir akış şeması izlendi. Çalışmanın tüm deneysel süreçleri ve raporlanması 01.01.2025 ile 31.12.2025 tarihleri arasında tamamlandı.

3.2. Katılımcılar, Evren ve Örneklem Seçimi

Araştırmanın evrenini, dünya genelindeki açık erişimli klinik veri havuzlarında yer alan ve sınıf dengesizliği barındıran elektronik sağlık kayıtları oluşturdu. Örneklem grubunu, makine öğrenmesi literatüründe standart kabul edilen ve etik kurul onayı gerektirmeyen anonim klinik veri setleri (UCI Machine Learning Repository (37), Kaggle (38), PhysioNet (39) vb.) teşkil etmektedir. Çalışmada, modelin genelleme yeteneğini ve otonom sistemin farklı senaryolara uyum kapasitesini test etmek amacıyla 3 adet farklı klinik veri seti kullanıldı. Veri setlerinin seçiminde; diyabet teşhisi, kalp hastalığı öngörüsü ve inme riski gibi temsiliyet gücü yüksek klinik başlıklar ile "Dengesizlik Oranı" (Imbalance Ratio - IR) kriterleri esas alındı.

Çalışmada 3 veriseti kullanıldı. 253.000 örnekten oluşan UCI Machine Learning Repostry Diyabet sağlık göstergeleri Veriseti https://archive.ics.uci.edu/dataset/891/cdc+diabetes+health+indicators test veriseti olarak kullanıldı. İlgili veri setinden 1/100

dengesizlik oranına sahip 25.000 alt örnek rastgele örnekleme yoluyla seçildi ve bu alt örneklem üzerinde program başarısı test edildi.

Veriseti HighBP, HighChol, CholCheck, BMI, Smoker, Stroke Heart Disease or Attack, PhysActivity, Fruits, Veggies, Hvy Alcohol Consump, Any Healthcare, No Docbc, Cost, Gen Hlth, Ment Hlth, Phys Hlth, Diff Walk , Sex, Age, Education, Income, Diabetes_binary bileşenlerinden oluşmaktadır. Çalışmada ağırlıklı olarak bu veriseti kullanıldı.

İkinci veriseti Pima Indian verisetidir. Görece düşük dengesizlik oranına rağmen az örnek sayısı nedeni ile seçildi. Veri setine https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database adresinden erişilebilir.

Üçüncü veriseti https://www.kaggle.com/datasets/rodolfomendes/abalone-dataset Abalone (Deniz Kulağı) verisetidir. Biyolojik ve biyomedikal verisetlerini temsilen seçildi.

3.3. Veri Toplama Araçları ve Süreci

Araştırmada veri toplama ve hazırlık süreçleri, Python dilinde geliştirilen pipeline.py, feature_engineering.py ve outlier_analysis.py modülleri aracılığıyla standartlaştırıldı. Veri toplama aracı olarak işlev gören bu yazılım modülleri, ham klinik verileri yapılandırılmış meta-verilere dönüştürerek analiz hattına aktarmaktadır. Verilerin geçerlik ve güvenirliğini sağlamak adına, toplama aşamasında eksik veri oranları, veri tipi tutarlılığı ve aykırı değer dağılımları "Pre-analysis" katmanında raporlanmaktadır. Bu raporlar, LLM (Büyük Dil Modeli) katmanına sunuldu ve modelin veri kalitesi hakkında bir yargıya varması hedeflendi.

3.4. Yazılım Altyapısı, Bilişim Kaynakları ve Ticari Ürünler

Araştırma kapsamında geliştirilen otonom analiz hattı ve tüm deneysel süreçler, yüksek seviyeli programlama dilleri ve akademik standartlara uygun yazılım kütüphaneleri kullanılarak inşa edildi. Sistemin ana geliştirme dili olarak, esnek veri yapısı desteği ve geniş bilimsel kütüphane ekosistemi nedeniyle Python 3.11 (Python Software Foundation, Wilmington, DE, ABD) tercih edildi (11). Yazılım mimarisi, yüksek boyutlu klinik verilerin işlenmesi sırasında bellek yönetimini optimize etmek amacıyla, ağır kütüphanelerin (XGBoost, Optuna vb.) yalnızca ihtiyaç duyulduğu anda yüklenmesini sağlayan "Lazy Loading" (Tembel Yükleme) protokolü ile yapılandırıldı. Ayrıca, matris hesaplamalarda çalışma zamanı (runtime) güvenliğini sağlamak ve

NumPy 2.0+ sürümlerindeki kopyalama kısıtlamalarından kaynaklanan hataları önlemek için sisteme özelleştirilmiş bir "copy_compat" middleware mekanizması entegre edildi.

Veri manipülasyonu, temizleme ve yapısal analiz süreçlerinde pandas (AQR Capital Management, Greenwich, CT, ABD) (14) ve vektörel hesaplamalar için NumPy (NumFOCUS, Austin, TX, ABD) (15) kütüphaneleri temel alındı. Makine öğrenmesi algoritmalarının inşasında ve temel istatistiksel testlerin yürütülmesinde scikit-learn (INRIA, Saclay, Fransa) (12) ve imbalanced-learn (akademik topluluk projesi) (40) kütüphanelerinden yararlanıldı. Derin öğrenme katmanları ve Grafik Sinir Ağları (GNN) mimarileri, yüksek performanslı tensör işlemleri ve türevlenebilir programlama desteği sunan PyTorch (Meta AI, Menlo Park, CA, ABD) çerçevesi üzerinden geliştirildi (13).

Hiper-parametre uzayında global optimumu aramak için Bayesyen optimizasyon stratejilerini kullanan optuna (Preferred Networks, Tokyo, Japonya) yazılımı tercih edildi (18). Model kararlarının klinik açıklanabilirliğini (XAI) sağlamak amacıyla SHAP (v0.42+, Lundberg ve Lee) kütüphanesi sisteme dahil edildi (41); PCA veya özellik seçimi sonrası oluşan boyutsal uyuşmazlıkları otomatik olarak hizalayan özelleştirilmiş "Shap Transform Wrapper" sınıfı metodolojiye eklendi. Otonom kontrol arayüzü Flask kütüphanesi ile web arayüzü olarak tasarlandı (42), stratejik karar destek katmanında Google Gemini (Alphabet San Francisco, CA, ABD) ve yerel sunucuda (127.0.0.1:8080) koşturulan gpt-oss (OpenAI, San Francisco, CA, ABD) modelleri orkestrator olarak kullanıldı.

3.5. Araştırmanın Bağımlı ve Bağımsız Değişkenleri

Araştırma tasarımında, klinik veri setlerindeki parametreler istatistiksel ve fonksiyonel rollerine göre iki ana grupta sınıflandırılmıştır. Değişkenlerin tanımlanması ve modele hazırlanması süreçleri data_loader.py ve pipeline.py modülleri tarafından yönetilmektedir.

3.5.1. Bağımsız Değişkenler (Öznitelikler/Features)

Bağımsız değişkenler, incelenen klinik vakalara ait demografik verileri (yaş, cinsiyet vb.) ve laboratuvar bulgularını (vücut kitle indeksi, kan şekeri, kolesterol seviyeleri vb.) kapsamaktadır. Bu değişkenler, model eğitimine girmeden önce "Standard Scaler" yöntemiyle birim varyans ve sıfır ortalamaya sahip olacak şekilde

ölçeklendirildi. Ayrıca, değişkenler arasındaki doğrusal olmayan karmaşık ilişkileri yakalamak amacıyla "Polynomial Features" yöntemiyle ikinci dereceden etkileşim terimleri türetildi; verideki çarpıklığı (skewness) gidermek için sayısal özniteliklere logaritmik dönüşümler uygulandı. Boyut indirgeme aşamasında, varyansın %95'ini temsil eden temel bileşenler PCA (Temel Bileşen Analizi) ile saptanarak bağımsız değişken uzayı optimize edildi (43).

3.5.2. Bağımlı Değişken (Hedef/Target)

Araştırmanın bağımlı değişkeni, klinik teşhisi veya risk durumunu temsil eden, genellikle "0" (Negatif/Sağlıklı) ve "1" (Pozitif/Hasta) değerlerini alan ikili (binary) kategorik değişkenlerdir. Bağımlı değişkenin sınıfları arasındaki sayısal asimetri, araştırmanın temel problemini oluşturmaktadır. Bu dengesizlik durumu, imbalance_preanalysis.py modülü tarafından hesaplanan Dengesiz Oranı (IR) metriği ile kantitatif olarak ifade edilmekte ve otonom sistemin dengeleme stratejisini (SMOTE, CTGAN, TabDDPM vb.) seçmesinde birincil girdi olarak kullanılmaktadır.

3.6. Veri Hazırlık ve Klinik Sigma Guardrails Metodolojisi

Ham veriler, otonom analiz hattına (pipeline) dahil edilmeden önce verinin klinik bütünlüğünü ve istatistiksel kararlılığını sağlamak amacıyla çok aşamalı bir "Güvence Protokolüne" tabi tutuldu. Eksik verilerin tamamlanması (imputation) sürecinde; sayısal değişkenler için medyan değeri, kategorik değişkenler için ise "Simple Imputer" aracılığıyla en sık tekrar eden (mode) değer atama stratejisi izlendi.

3.6.1. Çoklu Anomali Tespit Mimarisi

Aykırı değer analizi modülünde (outlier_analysis.py) kullanılan algoritmaların teknik özellikleri ve çalışma prensipleri aşağıda tablo 3.1'de sunulmuştur:

Tablo 3.1. Aykırı Değer Tespit Modülünde Kullanılan Algoritmalar ve Çalışma Prensipleri

Algoritma	Yöntem Türü	Çalışma Prensibi ve Tespit Mekanizması
Local Outlier Factor (LOF)	Yoğunluk Tabanlı	Veri noktalarının yerel yoğunluğunu komşularıyla kıyaslayarak, yoğunluğu düşük kalan sapmaları belirler (44-45).
Isolation Forest	Ağaç Tabanlı (Ensemble)	Veriyi rastgele özellikler üzerinden dallara ayırır (partitioning); daha az bölünme ile (hızlı) izole edilen ve ağacın yüzeyinde kalan noktaları uç değer olarak işaretler (46).
Auto-Encoder (Otokodlayıcı)	Sinir Ağı (Reconstruction)	Veriyi düşük boyutlu bir gizli uzaya sıkıştırıp tekrar inşa eder; yeniden inşa hatası (reconstruction error) yüksek olan vakaları anomali olarak tanımlar (47).

Algoritma	Yöntem Türü	Çalışma Prensibi ve Tespit Mekanizması
GEnergy-Score	Derin Öğrenme (Enerji Tabanlı)	PyTorch tabanlı bir derin öğrenme katmanı kullanarak, veri noktalarının logit değerleri üzerinden enerji yoğunluğunu hesaplar ve düşük olasılıklı bölgeleri tespit eder (48).

3.6.2. Sigma Guardrails ve Klinik Veto Protokolü

Anomali tespit modelleri tarafından işaretlenen aday vakalar (CANDIDATE), sistemin nihai karar katmanı olan "Sigma Guardrails" mekanizmasına aktarılmaktadır. Tespit edilen aykırı değerlerin, sınıf bazlı standart sapma (σ) dağılımları üzerinden analiz edildiği bu mekanizmanın karar kriterleri ve uygulanan "Klinik Veto" etiketleri Tablo 3.2'de özetlenmiştir.

Tablo 3.2. Sigma guardrails mekanizması karar matrisi ve klinik veto etiketleri

Etiket (Klinik Veto)	Sigma Aralığı	Klinik Tanım ve Uygulanan İşlem
KILLED (Klinik Gürültü)	$\geq 5\sigma$	Sınıf ortalamasından 5σ ve üzeri uzaklıkta bulunan, biyolojik veya medikal limitleri zorlayan verilerdir (Örn: BMI=77). Veri setinden kalıcı olarak elenir.
RESCUED (Nadir Vaka Koruma)	$<3\sigma$	İstatistiksel olarak aykırı görünmesine rağmen 3σ bandı içerisinde kalan ve sınıfın uç sınırlarını (borderline cases) temsil eden "Nadir Klinik Tablolar"dır. Modelin genelleme yeteneğini artırmak amacıyla eğitim setine geri kazandırılır.
REMAINED (Gri Bölge)	$3\sigma - 5\sigma$	3σ ile 5σ arasında kalarak net bir klinik mühür almayan vakalardır. Veri setinin epistemik güvenliğini korumak adına temizlik sürecine dahil edilir.

3.7. Örnekleme ve Modelleme Stratejileri

Dengeleme aşamasında, verinin istatistiksel dağılımını bozmadan sınıf asimetrisini gidermek amacıyla smote_utils.py ve diffusion_sampler.py modülleri üzerinden ileri düzey örnekleme protokolleri uygulandı. Bu süreçte kullanılan yöntemler; klasik interpolasyon, hibrit temizleme ve üretken difüzyon mimarileri olmak üzere üç ana eksende ve DNN modelleri ile meta-sezgisel destekleyici modellerle yapılandırıldı.

3.7.1. Adaptif Komşuluk ve Dinamik k-Neighbors Mekanizması

Standart SMOTE tabanlı algoritmarda, azınlık sınıfı örnek sayısının (nminority) belirlenen sabit komşuluk değerinden (k=5) az olması durumunda sistemin durmasına yol açan "ValueError" hatası klinik veri setlerinde sıkça karşılaşılan bir kısıttır (19). Bu metodolojik sorunu aşmak için sisteme "Dinamik k-Neighbors" mekanizması entegre edilmiştir. Mekanizma, her örnekleme işlemi öncesinde azınlık sınıfı yoğunluğunu

kontrol ederek, k parametresini otonom olarak min(kvarsayilan,nminority−1) formülü ile revize eder. Bu adaptif yaklaşım, özellikle nadir hastalıkların analizi gibi kısıtlı vaka sayısına sahip veri setlerinde analiz hattının kesintisiz çalışmasını güvence altına almaktadır.

3.7.2. Üretken Difüzyon Modelleri (TabDDPM) ve Veri Sentezi

Geleneksel interpolasyon yöntemlerinin (SMOTE, ADASYN) verideki karmaşık korelasyonları yakalayamadığı durumlarda, derin öğrenme tabanlı Tabular Diffusion (TabDDPM) modeli devreye alınmaktadır. TabDDPM metodolojisi, veriye kademeli olarak Gaussian gürültüsü ekleyen bir "ileri difüzyon" süreci ve bu süreci tersine çevirerek temiz veriyi sentezleyen bir "ters difüzyon" (denoising) süreci üzerine kuruludur. Sistemde uygulanan ters difüzyon süreci, 1000 adımlık (T=1000) bir yineleme ile yapılandırılmış olup, her adımda verinin marjinal dağılımları latent uzayda yeniden inşa edilmektedir. Sentezlenen verilerin kalitesi, "Sadakat Skoru" (Fidelity Score) protokolü ile mühürlenmektedir. Sadakat Skoru, sentetik verinin orijinal veriyle olan istatistiksel benzerliğini ölçen "Statistical Fidelity" ve öznitelikler arası bağların korunmasını denetleyen "Correlation Fidelity" değerlerinin harmonik ortalaması alınarak hesaplanır (6).

3.7.3. Gelişmiş Grafik Sinir Ağları (Advanced GNN) ve Hibrit Komşuluk

Hastalar arasındaki topolojik ilişkileri ve özellik benzerliklerini modeelişmiş Grafik Sinir Ağları (Advanced GNN) katmanı entegre edildi. Bu mimaride hastalar graf yapısındaki düğümler (V), hastalar arası benzerlikler ise kenarlar (E) olarak temsil edilir. GNN metodolojisinin kalbinde yer alan "Hibrit Komşuluk" (Hybrid Adjacency) matrisi, statik ve dinamik bağlantıların sentezlenmesiyle oluşturulur. Astatik Öklid mesafesine dayalı KNN grafiğini, Adinamik ise latent uzayda LinkPredictor modülü tarafından öğrenilen ilişkileri temsil eder. Mesaj iletimi (message passing) aşamasında, simetrik normalizasyon uygulanmış komşuluk matrisi kullanılarak düğüm özellikleri (X) katmanlar arasında aktarılır (31).

3.8. Otonom Yönetim Mekanizması ve Stratejik Karar Katmanı

"Clinical AI Analyst" platformu, insan müdahalesine olan ihtiyacı minimize eden ve analiz sürecini otonom olarak optimize eden; pekiştirmeli öğrenme (RL), LLM

orkestrasyonu ve ReAct (Reasoning and Acting) temelli çift katmanlı bir yönetim mimarisine sahiptir.

3.8.1. Pekiştirmeli Öğrenme (RL) Tabanlı Strateji Seçimi

Analiz hattının hangi örnekleme yöntemini ve hangi model mimarisini seçeceğine dair stratejik kararlar, RLOptimizer modülü bünyesindeki Q-Learning ajanı tarafından verilmektedir. Ajan; veri setinin dengesizlik oranı (IR), örneklem boyutu (N) ve öznitelik sayısı (D) gibi parametreleri içeren bir "Durum" (State) uzayında hareket eder. Q-Değerlerinin güncellenmesinde Bellman Denklemi temel alınmaktadır. Ajanın ödül fonksiyonu (R), modelin klinik başarısını doğrudan yansıtacak şekilde ağırlıklandırılmıştır (49). Bu ödül mekanizması, sistemin zamanla veri setinin karakteristiğine en uygun stratejiyi (örn: yüksek gürültülü verilerde SMOTE-ENN, karmaşık verilerde TabDDPM) deneyimsel olarak seçmesini sağlamaktadır. Bu yapı veri üzerinde uzun tekrarlı çalışmalar gerektirmektedir. Çok tekrarlı yapı sebebi ile tek tekrarlı testlerde kullanılmamaktadır.

3.8.2. LLM Orkestrasyonu ve ReAct Kararı

Sistemin üst düzey karar alma süreçleri, Büyük Dil Modellerinin (LLM) muhakeme yeteneği üzerine kurgulanan "ReAct" döngüsü ile yönetilmektedir (50). Model; önce imbalance_preanalysis.py modülünden gelen teknik özetleri (özellik türleri, IR değeri, eksik veri oranları) analiz ederek bir hipotez kurar (Reasoning). Ardından, bu hipotezi test etmek amacıyla analiz hattındaki ilgili Python fonksiyonlarını (örn: run_smote_pipeline()) belirli parametre setleri ile tetikler (Acting).

3.8.3. Öz-Yansıtma (Self-Reflection) ve Hata Düzeltme Döngüsü

Modelin üretebileceği olası sentaks hatalarını veya mantıksal kusurları (hallucination) gidermek amacıyla sisteme bir "Öz-Yansıtma" katmanı eklenmiştir. Eğer LLM tarafından önerilen parametre seti yazılımın beklediği aralıkların dışındaysa (örn: k-komşuluk değerinin negatif olması), sistem yürütme hatasını yakalar ve bu hatayı bir "Error Feedback" olarak modele geri iletir. Model, bu teknik geri bildirimi analiz ederek kendi kodunu otonom olarak revize eder. Bu yinelemeli süreç, analiz hattı hatasız bir şekilde yürütülene kadar devam eder ve sistemin "kendi kendini düzelten" (self-healing) bir mimari sergilemesini sağlar (51).

3.9. Performans Değerlendirme ve Çok Boyutlu Metrik Analizi

Geliştirilen modellerin klinik güvenilirliğini ve teknik başarısını ölçmek amacıyla, evaluator.py modülü üzerinden çok boyutlu bir performans değerlendirme çerçevesi kurgulandı. Özellikle ekstrem sınıf dengesizliği içeren klinik veri setlerinde salt doğruluk (Accuracy) metriğinin, azınlık sınıfındaki hataları maskelemesi ve "Doğruluk Paradoksu"na yol açması nedeniyle, modellerin başarısı geniş bir metrik yelpazesi üzerinden eş zamanlı olarak analiz edilmektedir.

3.9.1. Bileşik Puan (Composite Score) ve Karar Eşiği Optimizasyonu

Modellerin sınıfları ayırt etme kapasitesi; "Karmaşıklık Matrisi" (Confusion Matrix) bileşenleri olan Doğru Pozitif, Doğru Negatif, Yanlış Pozitif ve Yanlış Negatif değerleri arasındaki ilişkiler üzerinden takip edilmektedir. Bu kapsamda; dengesiz veri setleri için en sağlam metrik kabul edilen, dört karmaşıklık hücresini de hesaba katarak tahmin edilen ve gerçek sınıflar arasındaki ilişkiyi ölçen Matthews Korelasyon Katsayısı (MCC) temel alındı (27). Ayrıca, çoğunluk sınıfına aşırı uyumu (overfitting) cezalandırarak her iki sınıf arasındaki dengeyi denetleyen, Duyarlılık ve Özgüllük değerlerinin geometrik ortalaması olarak tanımlanan Geometrik Ortalama (G-Mean) metriği kullanıldı (52).

Kullanılan temel metrikler, tanımları ve klinik önemleri Tablo 3.3'de özetlenmiştir.

Tablo 3.3. Performans değerlendirmede kullanılan metrikler ve klinik anlamları

Metrik	Tanım ve Çalışma Prensibi	Klinik Önem ve Açıklama
Duyarlılık (Recall/Sensitivity)	Pozitif vakaların (hasta) ne kadarının model tarafından doğru tespit edildiğini ifade eder.	Klinik senaryolarda hastalıkların gözden kaçırılmaması (yanlış negatiflerin önlenmesi) adına hayati öneme sahiptir.
Kesinlik (Precision)	Modelin pozitif olarak tahmin ettiği vakaların gerçekte ne kadarının pozitif olduğunu gösterir.	Yanlış alarmların (gereksiz ileri tetkik veya cerrahi müdahale vb.) maliyetini ve iş yükünü ölçmede kritiktir.
Matthews Kor. Katsayısı (MCC)	Gerçek ve tahmin edilen sınıflar arasındaki korelasyonu ölçer; veri dağılımı dengesiz olsa dahi güvenilir sonuç üretir.	Sınıf dengesizliğinden etkilenmeyen, modelin rastgele tahmin yapıp yapmadığını gösteren en güvenilir metriklerden biridir.
Geometrik Ortalama (G-Mean)	Pozitif ve negatif sınıfların başarı oranlarının çarpımının karekökü alınarak hesaplanır.	Modelin sadece çoğunluk sınıfını öğrenmesini engelleyerek, her iki sınıftaki başarının dengeli olmasını sağlar.

3.9.2. Sınıflandırma Metriklerinin Matematiksel Temelleri

Farklı örnekleme (sampler) ve model kombinasyonlarını nesnel bir kriterle sıralamak amacıyla, pipeline.py bünyesinde özelleştirilmiş bir "Bileşik Puan" (Composite Score) formülü uygulanmaktadır. Bu puanlama sistemi, modelleri sadece ham başarılarına göre değil, klinik kararlılıklarına göre de hiyerarşik olarak sıralar.

Klinik risk yönetimini desteklemek amacıyla, pipeline.py modülü aracılığıyla otomatik karar eşiği (threshold) optimizasyonu yapılmaktadır. Standart makine öğrenmesi modellerinde 0.5 olarak kabul edilen karar eşiği, bu sistemde fn_cost_ratio (Yanlış Negatif Maliyet Oranı) parametresi doğrultusunda yeniden ayarlanır. Sistem, aşağıdaki toplam maliyet fonksiyonunu minimize eden optimal eşik değerini bularak, özellikle hayati tehlike arz eden durumlarda yanlış negatiflerin oranını otonom olarak aşağı çeker.

3.10. İstatistiksel Analiz ve Model Doğrulama Protokolü

Elde edilen sonuçların bilimsel geçerliliğini ve modeller arasındaki performans farklarının tesadüfi olmadığını doğrulamak adına sistematik bir istatistiksel protokol izlenmektedir. Bu süreç, sonuçların güvenilirliğini (reliability) ve tekrarlanabilirliğini (reproducibility) teminat altına almaktadır.

3.10.1. Tekrarlı Deneyler ve Kararlılık Analizi

Analiz hattındaki her bir deney, sonuçların rastlantısallıktan arındırılması amacıyla farklı "Rastgele Tohum" (Random Seed) değerleri kullanılarak çok kez tekrarlanmaktadır. Her bir döngü sonucunda elde edilen metriklerin aritmetik ortalaması ve standart sapması hesaplanarak modellerin kararlılığı (robustness) ölçülmektedir. Standart sapmanın yüksek olması, modelin veri bölünmesine (data split) veya başlangıç parametrelerine karşı hassas olduğunu gösterdiğinden, bu tip modeller otonom sistem tarafından "düşük güvenli" olarak işaretlenir ve klinik kullanım için riskli kabul edilir.

3.10.2. Friedman ve Nemenyi Post-hoc Testleri

Birden fazla algoritmanın ve örnekleme tekniğinin farklı veri setleri üzerindeki performans farklarını analiz etmek için parametrik olmayan Friedman Testi uygulanmaktadır. Bu test, algoritmaları başarı sıralamalarına (rank) göre değerlendirerek, "Tüm algoritmaların performansları eşittir" şeklindeki sıfır hipotezini (H0) test eder (53).

Friedman testi sonucunda istatistiksel anlamlılık değerinin (p) kabul edilen eşik değerin altında çıkması durumunda, hangi model ikililerinin birbirinden anlamlı derecede farklılaştığını saptamak amacıyla Nemenyi Post-hoc Testi devreye alınır. Nemenyi testi, modeller arasındaki "Kritik Fark" (Critical Difference - CD) değerini hesaplar. Performans farkı bu kritik değerden büyük olan modeller, istatistiksel ve bilimsel olarak birbirinden ayrışmış kabul edilir (54-55).

Kullanılan istatistiksel yöntemlerin amaçları ve karar mekanizmaları Tablo 3.4'te özetlenmiştir.

Tablo 3.4. Araştırmada kullanılan istatistiksel doğrulama yöntemleri

Yöntem / Protokol	Temel Amaç	Karar ve Değerlendirme Kriteri
Tekrarlı Deneyler (N-runs)	Sonuçların rastlantısal olmadığını doğrulamak ve model kararlılığını ölçmek.	Düşük standart sapma, modelin kararlı (robust) olduğunu gösterir. Yüksek sapma "güvensiz" olarak etiketlenir.
Friedman Testi	Çoklu model ve veri seti karşılaştırmalarında genel bir fark olup olmadığını test etmek.	p değeri anlamlılık eşiğinden küçükse, modeller arasında en az bir önemli fark olduğu kabul edilir.
Nemenyi Post-hoc Testi	Friedman testi sonrası, farkın tam olarak hangi modeller arasında olduğunu belirlemek.	İki modelin sıralama farkı "Kritik Fark" (CD) değerini aşıyorsa, performans farkı istatistiksel olarak anlamlıdır.

3.11. Model Açıklanabilirliği ve Klinik Yorumlanabilirlik (XAI)

Yazılımın çıktı katmanında yer alan raporlama ve görselleştirme modülleri, modelin "kara kutu" yapısını şeffaflaştırarak klinik karar destek süreçlerine entegre etmektedir.

3.11.1. SHAP Analizi ve Boyut Hizalama Metodolojisi

Klinik parametrelerin (örn: glukoz seviyesi, yaş, BMI) model tahmini üzerindeki ağırlıklarını belirlemek amacıyla SHAP (SHapley Additive exPlanations) yöntemi kullanılmaktadır. Oyun teorisindeki Shapley değerlerini temel alan bu yöntem, her bir özniteliğin model çıktısına olan katkısını (pozitif veya negatif) adil bir şekilde dağıtır (41).

Metodolojik bir zorluk olarak; Temel Bileşen Analizi (PCA) veya polinom özellik türetimi gibi işlemlerden sonra öznitelik isimleri ve sayıları değişmekte, bu durum SHAP grafiklerinin orijinal klinik isimlerle üretilmesini engellemektedir. Bu

sorunu aşmak için sisteme entegre edilen "Shap Transform Wrapper" sınıfı, veri dönüşüm hatlarını tersine izleyerek (inverse tracking) ham özellik isimlerini SHAP değerleriyle çalışma anında hizalar. Bu sayede, model çok karmaşık matematiksel dönüşümlerden geçse dahi, klinisyenin ekranına yansıyan SHAP grafiği "Özellik 1, Özellik 2" gibi soyut terimler yerine "Kan Şekeri, BMI" gibi gerçek klinik ifadeleri içerir.

3.11.2. Olasılık Kalibrasyonu ve Brier Skoru

Modelin ürettiği olasılık değerlerinin (örn: %85 risk) gerçek vaka frekanslarıyla uyumunu ölçmek için "Olasılık Kalibrasyonu" analizi yürütülmektedir. Bu analiz kapsamında hesaplanan Brier Skoru, tahmin edilen olasılıklar ile gerçek sonuçlar arasındaki hatayı ölçer. 0 ile 1 arasında değer alan bu skorda, 0 değeri mükemmel kalibrasyonu temsil eder (28). Klinik risk hesaplamalarında modelin sadece sınıfı (Hasta/Sağlıklı) doğru bilmesi yeterli değildir; riskin şiddetini de doğru yansıtması beklendiğinden, Brier skoru modelin güvenilirliğinin mühürlenmesinde kritik bir tamamlayıcıdır. Kullanılan açıklanabilirlik (XAI) ve doğrulama araçları Tablo 3.5'de özetlenmiştir.

Tablo 3.5. Model açıklanabilirliği ve güvenilirlik analiz araçlar

Araç / Metodoloji	Temel İşlev ve Amaç	Klinik Katkısı
SHAP Analizi	Özniteliklerin tahmin üzerindeki pozitif/negatif etkisini oyun teorisi ile hesaplar.	Hangi klinik bulgunun kararda ne kadar etkili olduğunu doktora gösterir.
Shap Transform Wrapper	Dönüştürülmüş (PCA vb.) verileri tersine işleyerek orijinal isimleriyle eşleştirir.	Teknik "Özellik X" etiketleri yerine gerçek tıbbi parametre isimlerinin raporda yer almasını sağlar.
Olasılık Kalibrasyonu (Brier Skoru)	Tahmin edilen risk olasılığı ile gerçekleşen risk arasındaki sapmayı ölçer.	Modelin verdiği Risk" uyarısının, gerçek hayattaki olasılığa ne kadar denk geldiğini denetler.

3.12. Veri Mahremiyeti ve Güvenli Veri İletişim Protokolü

Klinik veri setlerinin hassas doğası ve hasta mahremiyetine dair yasal ve etik zorunluluklar (KVKK vb.) nedeniyle, geliştirilen sistemin mimarisi "tasarım yoluyla gizlilik" (privacy by design) ilkesine göre kurgulanmıştır (56). Bu kapsamda, llm_controller.py ve cli_tool.py modülleri üzerinden yürütülen stratejik orkestrasyon süreçlerinde ham hasta verilerinin güvenliği en üst düzeyde tutulmaktadır.

3.12.1. Meta-Veri Seviyesinde İletişim ve Anonimleştirme

Sistem, Büyük Dil Modelleri (LLM) ile etkileşime girerken ham veri satırlarını veya bireysel kayıtları asla modelin bağlam penceresine (context window) iletmemektedir. Bunun yerine, imbalance_preanalysis.py tarafından üretilen ve yalnızca verinin yapısal karakteristiğini tanımlayan anonimleştirilmiş meta-veri paketleri kullanılmaktadır.

Bu paketler; öznitelik isimleri, veri tipleri (kategorik/sürekli), eksik veri yüzdeleri, merkezi eğilim ölçüleri ve anomali tespit aşamasında saptanan sigma dağılım istatistiklerinden oluşmaktadır. Bu metodoloji sayesinde, modelin muhakeme yapması için gerekli olan teknik içerik sağlanırken, hiçbir gerçek vakanın kimliği belirlenebilir verisi sistem dışına (özellikle API tabanlı modellerde) veya analiz hattı dışına sızmamaktadır.

3.12.2. Teknik Veri Güvenliği ve SafeJSONEncoder Mekanizması

Analiz hattındaki teknik verilerin (metrikler, parametreler, konfigürasyonlar) LLM katmanı ile JSON formatında takas edilmesi sırasında, Python'un standart veri tiplerinden kaynaklanan uyumsuzlukları yönetmek amacıyla özelleştirilmiş bir SafeJSONEncoder sınıfı geliştirilmiştir.

Klinik analizlerde sıkça rastlanan ve standart JSON standartlarına aykırı olan NaN (Not a Number) veya Infinity (Sonsuz) gibi değerler, bu encoder tarafından çalışma anında tespit edilerek LLM'in anlayabileceği güvenli string formatlarına veya null değerlerine dönüştürülmektedir. Bu işlem, veri iletişim katmanında meydana gelebilecek ayrıştırma (parsing) hatalarını ve bu hatalardan kaynaklanabilecek güvenlik açıklarını veya analiz kesintilerini engellemektedir.

3.13. Otonom Raporlama ve Klinik Karar Destek Çıktıları

"Clinical AI Analyst" platformunun nihai aşaması, analiz hattı boyunca elde edilen karmaşık teknik bulguların; klinik uzmanlar tarafından anlaşılabilir, yorumlanabilir ve aksiyon alınabilir formatlara dönüştürülmesidir.

3.13.1. LLM Tabanlı Sonuç Yorumlama (Executive Summary)

Result_interpreter.py modülü, analiz hattının ürettiği tüm performans metriklerini, istatistiksel anlamlılık testlerini ve görselleştirme çıktılarını sentezleyerek

bütünsel bir "Yönetici Özeti" üretmektedir. Metodolojik olarak bu süreç, LLM'in bir "Klinik Veri Bilimcisi" rolünü üstlendiği bir Expert-In-The-Loop simülasyonudur (57).

Program içinde oluşturulan RAG (Retrieval-Augmented Generation) sistemi sayesinde kayıt altına sonuçlar; hangi örnekleme yönteminin neden seçildiğini, seçilen modelin klinik güvenilirliğini (Brier skoru ve kalibrasyon üzerinden) ve modelin hata tiplerinin (FP/FN) hasta yönetimi üzerindeki olası etkilerini akademik ve medikal bir dille açıklar. Program üzerinden yapılan LLM yorumları PDF çıktısı olarak rapor formatında kullanıcıya sunulmaktadır.

3.13.2. Klinik Görselleştirme ve Teşhis Güven Analizi

Karar destek mekanizmasını güçlendirmek ve "Kara Kutu" (Black Box) endişelerini gidermek adına, visualizations.py modülü tarafından üretilen grafiksel çıktılar, modelin "teşhis güvenini" ve "karşılaştırmalı başarısını" matematiksel olarak belgeler. Bu kapsamda uygulanan görselleştirme protokolleri şunlardır:

Dinamik Liderlik Tablosu (Leaderboard) ve Performans Matrisi: Sistem, tek bir modele odaklanmak yerine, arka planda yarıştırdığı (örneğin 175 farklı) Model-Örnekleyici kombinasyonunun tamamını tek bir görsel panoda sunar. Bu Liderlik Tablosu, modelleri yalnızca doğruluğa (Accuracy) göre değil, klinik açıdan daha kritik olan MCC (Matthews Correlation Coefficient) ve Duyarlılık (Recall) metriklerine göre hiyerarşik olarak sıralar. Klinisyen, bu tablo sayesinde neden XGBoost yerine Random Forest modelinin "Şampiyon" ilan edildiğini, metrikler arasındaki ödünleşimleri (trade-offs) inceleyerek görsel olarak doğrular.

Tahmin yoğunluğu ve güven analizi (KDE), modelin pozitif (Hasta) ve negatif (Sağlıklı) sınıfları olasılık uzayında ne kadar keskin bir sınırla ayırabildiğini göstermek için Çekirdek Yoğunluk Tahmini (KDE) grafikleri kullanılır.

Ayrım gücü ideal bir modelde, iki sınıfın olasılık eğrileri birbirinden tamamen ayrışmalıdır.

Belirsizlik alanı eğrilerin kesiştiği "Gri Bölgeler", modelin karar vermekte zorlandığı ve hekimin inisiyatif alması gereken riskli vakaları görselleştirir. Bu analiz, hekimin model çıktısına ne derece güvenebileceğine dair somut bir "istatistiksel güven aralığı" sunar (58).

Sentetik Veri Sadakat (Fidelity) Isı Haritaları: Özellikle CTGAN gibi üretken modeller kullanıldığında, sentezlenen verilerin kalitesini mühürleyen Korelasyon Farkı Matrisleri (Correlation Difference Matrix) oluşturulur. Bu ısı haritaları, orijinal klinik

verideki değişkenler arası ilişkilerin (örn: İnsülin ve Glukoz korelasyonu) sentetik veride ne kadar korunduğunu gösterir. Renk skalasının nötr (beyaz/açık) tonlarda kalması, sentetik verinin orijinal vaka dağılımını bozmadığını ve eğitimin manipüle edilmediğini metodolojik olarak kanıtlar.

Klinik Karar Eğrileri (ROC ve Precision-Recall): Sınıf dengesizliğinin yüksek olduğu senaryolarda model başarısını doğrulamak için standart ROC eğrilerine ek olarak Precision-Recall (PR) Eğrileri üretilir. Bu grafikler, modelin nadir görülen pozitif vakaları yakalarken ne kadar yanlış alarm (False Positive) ürettiğini görselleştirerek, klinisyenin "Risk/Fayda" analizini yapmasını kolaylaştırır.

3.13.3. LLM Muhakeme Davranışının İstem Mühendisliği (Prompt Engineering) ile Optimizasyonu

Sistemin merkezi karar mekanizmasını oluşturan Büyük Dil Modelinin (LLM) davranışı, analiz hattındaki teknik görevlere uyum sağlaması amacıyla ileri düzey istem mühendisliği (prompt engineering) teknikleri ile optimize edilmiştir. Modelin muhakeme süreci, prompts/orchestrator.txt ve chat_system.txt dosyasında tanımlanan ve modelin epistemik sınırlarını belirleyen çok katmanlı bir "Sistem İstemi" (System Prompt) üzerinden yönetilmektedir. Sistemde uygulanan temel yönlendirme stratejileri şunlardır:

• Rol Tanımlama ve Bağlamsal Çerçeveleme: İstem hiyerarşisinin ilk katmanında model; "Karmaşık klinik veri setleri üzerinde sınıf dengesizliğini optimize eden kıdemli bir veri bilimcisi" rolüyle mühürlenmektedir. Bu rol tanımlaması, modelin üreteceği yanıtların sadece kod tabanlı değil, aynı zamanda klinik riskleri (Yanlış Negatif maliyeti gibi) gözeten bir perspektifte olmasını sağlamaktadır.

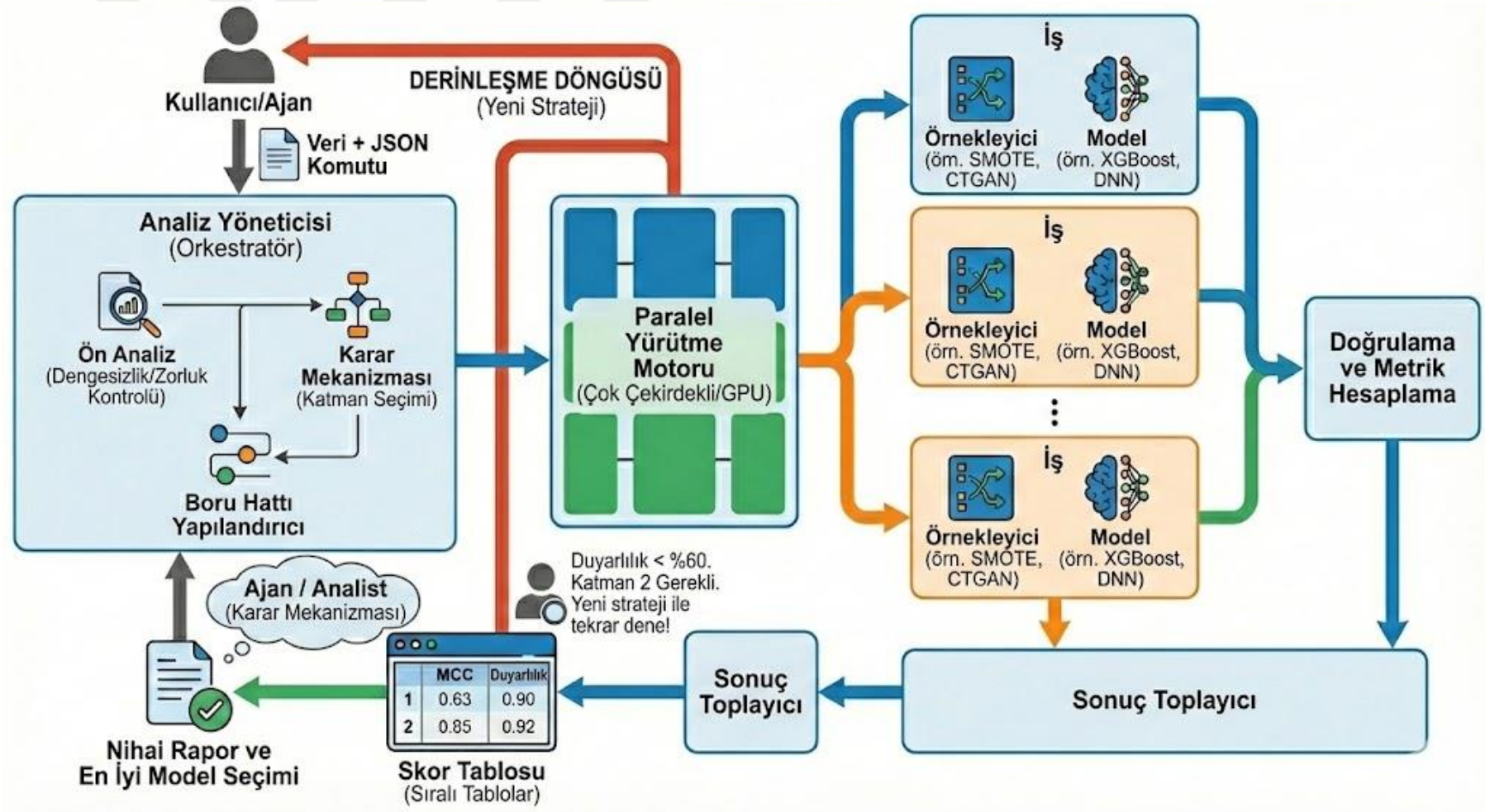
• Zincirleme Düşünce (Chain-of-Thought - CoT) Optimizasyonu: LLM'in bir analiz yöntemi seçmeden önce verinin meta-verileri (Dengesizlik oranı, öznitelik sayısı, eksik veri dağılımı vb.) üzerinde mantıksal bir çıkarım yapması için uygulanır (59). İstem içerisinde yer alan; "Önce verideki anomali yapısını değerlendir, ardından bu yapıya en uygun sampling algoritmasını teknik gerekçeleriyle açıkla" talimatı, modelin rastgele seçimler yapmak yerine adım adım muhakeme yürüterek karar vermesini sağlar.

• Birkaç-Örnekli (Few-Shot) Öğrenme ve Şema Uyumu: Modelin ürettiği çıktıların yazılım katmanı (Python) tarafından hatasız bir şekilde işlenebilmesi için istem içerisinde Few-Shot örnekleri sunulmaktadır (7). Bu örnekler; farklı klinik

senaryolara karşılık gelen ideal JSON çıktı formatlarını ve parametre aralıklarını içerir. Böylece modelin halüsinasyon (hallucination) görme ihtimali minimize edilerek çıktının teknik geçerliliği artırılır.

• Kısıt Temelli Yönlendirme ve Güvenlik Sınırları: Prompt mühendisliği kapsamında modele belirli kısıtlamalar (constraints) getirilmiştir. Örneğin; "Sentetik veri üretiminde asla orijinal örnek sayısının %200'ünü aşma" veya "Klinik olarak imkansız olan BMI ve yaş kombinasyonlarını aykırı değer olarak mühürle" gibi talimatlar, modelin analiz hattı üzerindeki operasyonel güvenliğini garanti altına alır.

LLM denetleyicisi tarafından yönetilen bu akıl yürütme, doğrulama ve yürütme süreçlerinin teknik mimarisi Şekil 3.1'de gösterilmiştir.



Şekil 3.1. LLM denetleyicisi ve akıl yürütme mimarisi

3.14. Araştırmanın Etik Yönleri ve Veri Sorumluluğu

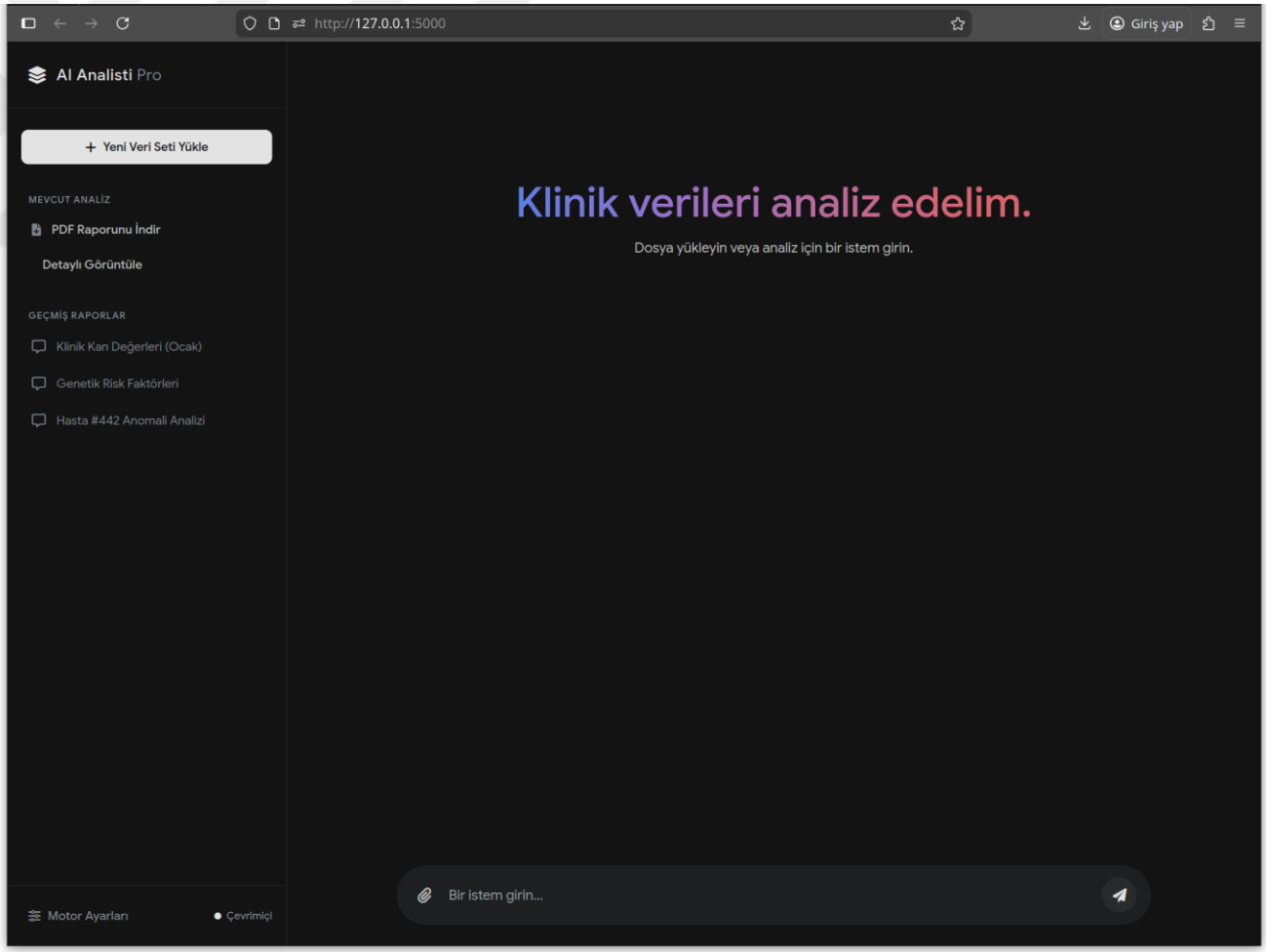
Bu çalışma, tamamen açık erişimli ve anonimleştirilmiş veri setleri üzerinden yürütüldüğü için doğrudan bir insan denek katılımı içermemektedir. Ancak, klinik verilerin otonom analizi süreçlerinde algoritmik yanlılığın (algorithmic bias) önlenmesi ve adil sonuçların üretilmesi birincil etik sorumluluk olarak ele alınmıştır (60). Bu amaçla, sistemin performans değerlendirmesinde "Dengeli Doğruluk" (Balanced Accuracy) ve "G-Mean" gibi çoğunluk sınıfının tahakkümünü cezalandıran metrikler kullanılarak, her bir vakanın eşit ağırlıkla değerlendirilmesi ve modelin azınlık gruplarına (nadir hastalıklara) karşı önyargılı davranmaması güvence altına alınmıştır.

3.15. "Clinical AI Analyst" Yazılımının Operasyonel Protokolü ve Kullanım Aşamaları

Geliştirilen yazılımın operasyonel süreci; kullanıcı hatasını minimize eden ve klinik veri bilimcisi olmayan uzmanların dahi yüksek başarımlı modeller üretmesini sağlayan doğrusal bir iş akışına sahip olacak şekilde tasarlanmıştır. Bu protokol, verinin sisteme girişinden klinik raporun çıktısına kadar olan tüm mikro-süreçleri otonom katmanlarla sarmalayarak standartlaştırır.

3.15.1. Başlangıç Konfigürasyonu ve Sistem Ara yüzü

Yazılım çalıştırıldığında kullanıcıyı karşılayan "Ana Kontrol Paneli", hiyerarşik bir menü yapısı ile tasarlanmıştır. Kullanıcı ilk aşamada, sol yan panelde (Sidebar) bulunan yapılandırma ayarlarını kullanarak LLM'in muhakeme motorunu belirler. Konfigürasyon aşamasında, yerel sunucu veya bulut tabanlı API modelleri arasından seçim yapıldıktan sonra etkileşim modülüne geçiş sağlanır. Arayüz ekran görüntüsü şekil 3.2’de gösterilmiştir.



Şekil 3.2. Yazılım karanlık mod destekli arayüzü

3.15.2. LLM Yürütücü ortamı

Bu tez çalışması kapsamında geliştirilen yazılım, Büyük Dil Modelleri (LLM) ile etkileşimde bulunmak için esnek, ölçeklenebilir ve gizlilik odaklı bir "Hibrit Denetleyici (Controller)" mimarisi benimsemiştir. Sistem, LLMController sınıfı üzerinden yönetilen iki ana iletişim kapısı ve özelleştirilmiş bir yerel çıkarım motoru (inference engine) üzerine inşa edildi.

Sistem, model bağımsızlığını sağlamak amacıyla soyutlanmış bir katman kullanır.

• Bulut Tabanlı Yüksek Performans (Google Gemini): GoogleLLMController modülü, google.genai kütüphanesini kullanarak Gemini model ailesine (Pro/Flash) erişim sağlar. Bu yöntem, karmaşık akıl yürütme ve kod analizi gerektiren görevlerde yerel donanım üzerindeki yükü kaldırarak GPU kaynaklarının tamamen SMOTE varyasyonları ve model eğitimi gibi veri bilimi süreçlerine (AnalysisManager tarafından yönetilen süreçler) ayrılmasına olanak tanır.

• Yerel Sunucu Entegrasyonu (Llama.cpp/vLLM): Veri gizliliğinin kritik olduğu veya internet erişiminin olmadığı senaryolar için LocalLLMController modülü tasarlanmıştır. Bu modül, OpenAI uyumlu REST API protokollerini kullanarak yerel portlar (örneğin: http://127.0.0.1:8000) üzerinden llama.cpp gibi hafifletilmiş sunucularla haberleşir. "Slot" tabanlı oturum yönetimi (id_slot) sayesinde, her analiz görevi kendi bağlamını (context) koruyarak çakışma olmadan yürütülür.

Sistem, ham klinik veriyi doğrudan LLM'e göndermek yerine, veriyi temsil eden meta-verilerle çalışır. AnalysisManager modülü, veri setinin kendisini değil; verinin istatistiksel parmak izini (fingerprint), korelasyon matrislerini ve dengesizlik oranlarını LLM'e sunar. Bu sayede, hasta verilerinin gizliliği donanımsal izolasyonla değil, mimari tasarım düzeyinde garanti altına alınmış olur.

Standart LLM API'lerinin ötesinde, bu çalışma kapsamında LowVRAMHybridEngine adında özgün bir yerel yürütücü geliştirilmiştir. Bu motor, LLM'lerin uzun bağlamlarda yaşadığı "dikkat dağınıklığı" (context drift) ve "halüsinasyon" sorunlarını çözmek için tensor düzeyinde müdahalelerde bulunur. 3 Katmanlı Hiyerarşik Bellek Yönetimi (Tiered Memory) sınırlı VRAM kapasitesini aşmak için sistem, veriyi önem derecesine göre üç katmanda saklar

▪ (GPU Active Window): En sık erişilen ve işlem anında gereken son token'lar (örn. 8192 token).

▪ (CPU Pinned Memory): GPU belleğinden taşan ancak silinmemesi gereken KV Cache (Anahtar-Değer Önbelleği) blokları sistem RAM'ine (pinned memory) taşınır.

▪ (RAG Vector Store): Uzun süreli hafıza için, eski konuşma parçaları ve analiz sonuçları vektörleştirilerek ClinicalRagMemory (ChromaDB) üzerinde saklanır.

Entropi tabanlı bağlam budama (Pruning) her üretim adımında dikkat matrislerinin (attention matrix) entropisini (attention_entropy) hesaplar. Entropinin yükselmesi, modelin kararsızlaştığını gösterir. Bu durumda sistem, dikkat mekanizmasındaki önemsiz token'ları dinamik olarak budayarak bağlamın "temiz" kalmasını sağlar.

Tensor manipülasyonu ile yönlendirme standart "temperature" örneklemesinin ötesinde, frequency_penalty ve presence_penalty değerleri, doğrudan çıktı logit'leri (logits) üzerinde tensor işlemleriyle uygulanır. Bu, modelin kendini tekrarlamasını matematiksel düzeyde engelleyerek, uzun süreli analiz raporu üretimlerinde tutarlılığın korunmasını sağlar.

Bu mimari sayesinde sistem, teorik olarak donanım belleğiyle sınırlı olmayan, "sonsuz" uzunlukta bir konuşma ve analiz geçmişini yönetebilecek bir altyapıya kavuşturuldu.

Fakat deneysel aşamada olan bu tasarımda işlem yükü nedeni ile düşük hız sorunu oluşmaktadır. İlerleyen dönemlerde bu sorunun aşılabilirliğine göre deneysel tasarım kalıcı hale getirebilecektir.

3.15.3. Veri Entegrasyonu ve Profilleme (Pre-analysis)

Kullanıcı, analiz edilecek klinik veri setini (CSV/Excel formatında) sürükle-bırak yöntemiyle arayüze aktarır. Veri yüklendiği anda arka planda imbalance_preanalysis.py modülü devreye girerek; verinin öznitelik türlerini, eksik veri oranlarını ve temel istatistiksel dağılımlarını hesaplar.

Bu hazırlık fazının sonunda sistem, verideki gizli ilişkileri ve çoklu bağlantı sorunlarını ortaya koymak adına öznitelikler arası korelasyon matrisini interaktif bir ısı haritası olarak sunar. Kullanıcı, bu görsel üzerinden modelin öğreneceği ilişkisel yapıyı analiz başlamadan önce denetleme imkanına kavuşur.

3.16. Otonom Analiz Hattının Yürütülmesi ve İzlenmesi

Kullanım sürecinin en kritik aşaması, doğal dil komutları ile başlatılan "Muhakeme ve Eylem" (ReAct) döngüsüdür. Kullanıcı, sohbet arayüzüne analizin

amacını (örn: "Diyabet verisini dengele ve modelle") yazar ve sistem bu soyut talimatı teknik bir iş planına dönüştürür. Analiz hattı boyunca LLM'in aldığı stratejik kararlar (örneğin; "Dengesizlik oranı 1:10 olduğu için TabDDPM tercih edildi") anlık olarak kullanıcı arayüzüne raporlanır. Kullanıcı, "Düşünce Süreci" (Thought Process) paneli üzerinden sistemin mantıksal çıkarımlarını adım adım izleyebilir. Bu şeffaflık, otonom sistemin "kara kutu" olarak kalmasını engeller ve kullanıcının sürece olan güvenini artırır.

3.17. Performans Çıktıları ve Klinik Raporlama Katmanı

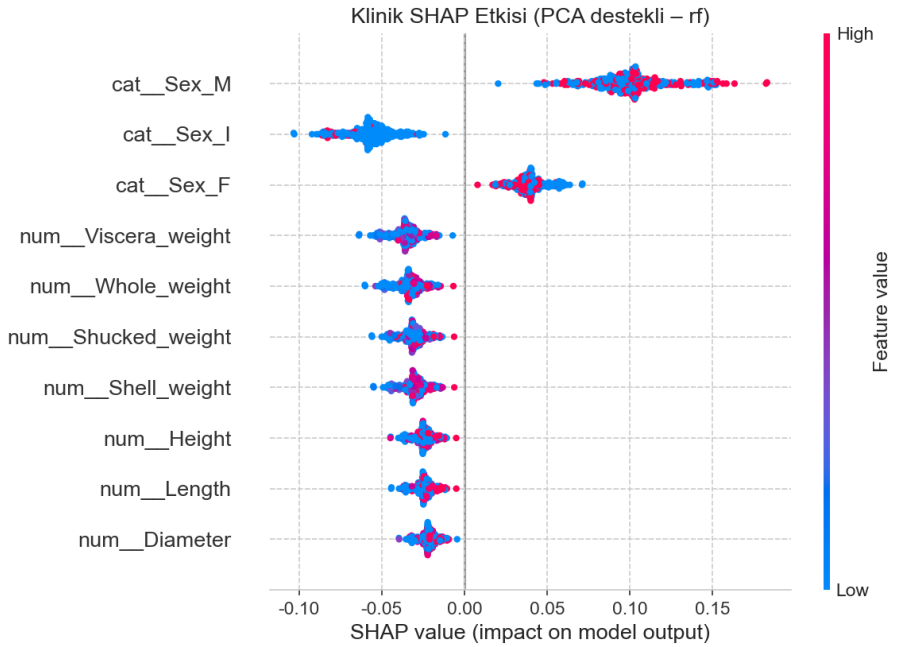
3.17.1. Model Karşılaştırma ve Entegre Geri Bildirim Mekanizması

Analiz hattının çıktıları, geleneksel statik panellerin ötesine geçerek, doğrudan kullanıcının doğal dil etkileşimini sürdürdüğü sohbet arayüzüne entegre edildi. Evaluator.py modülü tarafından hesaplanan performans metrikleri (F1, MCC, AUC) ve model karşılaştırma tabloları, sohbet akışı içerisindeki konuşma balonları (speech bubbles) vasıtasıyla yapılandırılmış geri bildirim blokları halinde sunulur.

Bu diyalog tabanlı sunum katmanı, kullanıcının analiz geçmişini kronolojik bir bağlamda takip etmesine olanak tanır. Farklı sampler ve model kombinasyonlarının (örn: SMOTE+XGBoost vs TabDDPM+DNN) başarımları, LLM tarafından yorumlanmış özet metinlerle birlikte bu balonlar içerisinde kullanıcıya iletilir. Böylece teknik metrikler, soyut sayılar olmaktan çıkarak "klinik bir asistanın sunduğu rapor" formatına dönüşür.

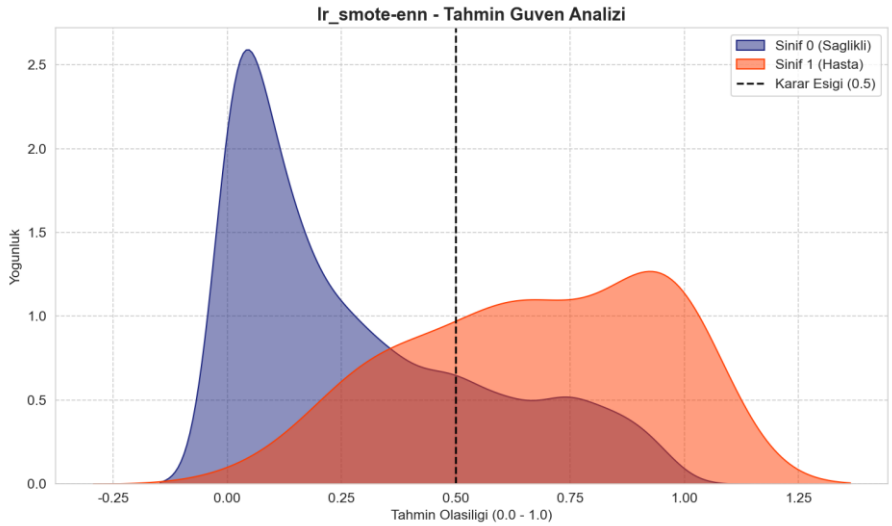
3.17.2. SHAP Açıklanabilirlik ve Dağılım Analizi

Modelin karar mekanizmasını klinik parametrelerle şeffaflaştıran SHAP (SHapley Additive exPlanations) görselleri, sistemin en kritik çıktılarındandır. Sisteme entegre edilen Shap Transform Wrapper sayesinde, veriler karmaşık matematiksel dönüşümlerden geçse dahi, özellik ağırlık grafikleri orijinal klinik isimlerle (örn: Yaş, Glukoz) üretilerek analiz akışında sunulur. Modelin kararlarını etkileyen faktörlerin görselleştirildiği örnek SHAP açıklaması Şekil 3.3’te yer almaktadır.



Şekil 3.3. Hangi parametrenin teşhis üzerinde ne kadar etkili olduğunu gösteren SHAP görseli.

Tahmin başarısının olasılıksal güvenilirliğini ölçmek amacıyla, pozitif ve negatif sınıfların olasılık uzayındaki ayrışmasını gösteren KDE (Kernel Density Estimation) grafikleri kullanılmaktadır. Dağılımın anlaşılmasını kolaylaştıran ve sınıfların tahmin güven aralıklarını belgeleyen bu analiz görseli Şekil 3.4'te gösterilmiştir.



Şekil 3.4. Pozitif ve negatif sınıfların ayrım gücünü gösteren yoğunluk grafiği.

3.17.3. Nihai Klinik Özet ve Arşivleme

Yazılımın en üst katmanı olan result_interpreter.py, tüm bu teknik verileri klinik bir "Yönetici Özeti" metnine dönüştürür. Kullanıcı, raporun altındaki "Deney Paketini

İndir" butonu ile tüm grafikleri, tabloları ve eğitilmiş modeli tek bir ZIP arşivi olarak teslim alır.

3.18. Örnek Senaryo Üzerinden İş Akış Analizi

Geliştirilen "Clinical AI Analyst" sisteminin uçtan uca performansını, karar verme yeteneklerini ve otonom kod üretim kapasitesini doğrulamak amacıyla, UCI Makine Öğrenmesi Veri Tabanı üzerinden temin edilen Diyabet veri seti üzerinde kontrollü bir deney senaryosu kurgulandı.

3.18.1. Veri Hazırlığı ve İstatistiksel Profilleme

Gerçek dünya klinik verilerinde sıkça rastlanan ve model başarısını doğrudan tehdit eden "nadir olay" (rare event) olgusunu simüle etmek amacıyla, ana veri havuzundan rastgele örnekleme (random sampling) yöntemiyle %10 büyüklüğünde bir alt küme izole edildi. Oluşturulan bu pilot veri setinde, sınıf dengesizliği yapay olarak 1/100 (her 100 sağlıklı bireye karşı 1 diyabet vakası) oranında sabitlendi.

Sistemin ön analiz modülü (imbalance_preanalysis.py), izole edilen veri setini tarayarak temel istatistiksel karakteristiğini çıkarmış ve elde edilen bulguları bir "durum raporu" formatında LLM'e sunmuştur. Analize temel teşkil eden özet istatistikler Tablo 3.6'de detaylandırılmıştır.

Tablo 3.6. LLM'e gönderilen özet istaistikler

Parametre	Ortalama	Varyans	(Skewness)	Basıklık	(Z>3)	%95 (CI)
HighBP	0.37	0.23	0.52	-1.73	0	±0.01
HighChol	0.39	0.24	0.45	-1.80	0	±0.01
BMI	27.89	40.93	2.32	13.37	40	±0.08
Smoker	0.43	0.25	0.27	1.93	0	±0.01
Stroke	0.03	0.03	5.28	25.87	715	±0.00
HeartDiseaseorAttack	0.08	0.07	3.17	8.02	1704	±0.00
PhysActivity	0.78	0.17	-1.32	-0.26	0	±0.01
Fruits	0.64	0.23	-0.59	-1.66	0	±0.01
Veggies	0.81	0.15	-1.61	0.58	0	±0.01
HvyAlcoholConsump	0.06	0.06	3.62	11.13	1373	±0.00
AnyHealthcare	0.95	0.05	-4.01	14.08	1159	±0.00
NoDocbcCost	0.08	0.08	3.00	6.98	1855	±0.00
GenHlth	2.40	1.07	0.50	-0.23	0	±0.01
MentHlth	2.96	50.18	2.87	7.45	1074	±0.09
PhysHlth	3.73	66.62	2.46	4.80	1457	±0.11
DiffWalk	0.14	0.12	2.08	2.34	0	±0.00
Sex	0.43	0.25	0.27	-1.93	0	±0.01
Age	7.80	9.51	-0.27	-0.67	0	±0.04
Education	5.09	0.95	-0.82	0.10	326	±0.01
Income	6.18	4.07	-0.98	-0.06	0	±0.03

Ön Analiz ve Veri Sağlığı Özeti:

• Veri Hacmi: 22.051 örneklem, 22 öznitelik.

• Sınıf Asimetrisi: Sınıf 0: 21.833 (%99.01) | Sınıf 1: 218 (%0.99).

• Baseline Performansı: Standart algoritmaların başarısızlığı (G-Mean: 0.000, MCC: 0.000).

• Aykırı Değer Yönetimi: Klinik koruma protokolü ile 52 nadir vaka kurtarılmış, biyolojik tutarsızlık gösteren 768 vaka "Extreme Sigma" kuralı gereği analiz dışı bırakıldı.

3.18.2. LLM Muhakeme Süreci (Chain-of-Thought)

LLM, kendisine sunulan istatistiksel tabloyu ve özet raporu analiz ederek, insan uzman seviyesinde bir "Zincirleme Düşünce" (CoT) akışı oluşturmuştur. Sistemin ham veriye dayalı stratejik muhakeme kaydı aşağıdadır:

Sistem Muhakeme Kaydı:

• Ekstrem Dengesizlik Patolojisi: Sınıf 1 (Diyabet) oranının sadece %0.99 (218 vaka) olması, veride "İğneyle kuyu kazma" senaryosu olduğunu göstermektedir. Baseline testindeki 0.000 MCC skoru, standart algoritmaların azınlık sınıfını tamamen göz ardı ettiğini kanıtlar. Bu noktada sentetik veri üretiminde en yüksek sadakati sağlayan Tabular Diffusion yöntemine başvurmak zorunludur.

• Öznitelik Dağılım Anomalileri: BMI (Kurtosis: 13.37) ve Fiziksel Sağlık (Varyans: 66.62) gibi parametrelerin dağılımı oldukça çarpıktır. Bu durum, modelin kararlılığını bozabilir. Veri mühendisliği katmanında Logaritmik Dönüşüm (transform-type log) uygulanarak varyansın normalize edilmesi kritik önem taşır.

• Klinik Veto (Sigma Guardrails) Analizi: 768 vakanın "Extreme Sigma" (5σ) ile elenmesi, veri setindeki biyolojik olarak imkansız veya hatalı kayıtların temizlendiğini gösterir. Kurtarılan 52 nadir vaka ise, modelin diyabetin "uç sınırlarını" öğrenmesi için hayati bir veri kümesidir.

• Performans Hedefi: Klinik bir teşhis söz konusu olduğu için Yanlış Negatif (FN) maliyeti çok yüksektir. Bu nedenle, otonom eşik optimizasyonunda maliyet oranını (fn-cost-ratio) yüksek tutarak Duyarlılığı (Recall) maksimize etmeliyiz.

Bu muhakeme zincirine bağlı olarak sistem, aşağıdaki teknik komut setini otonom olarak oluşturmuş ve yürütme katmanına iletmiştir:

"command": "EXECUTE_FUNCTION",

"module": "analysis_manager",

"function": "start_pipeline_async",
"kwargs": {
"run_id": "session_clinical_analysis_01",
"params": {
"file_path": "diyabet.csv",
"target_col": "Diabetes_binary",
"use_diffusion": true,
"use_ensemble": true,
"run_outlier_analysis": true,
"threshold_optimization": true,
"fn_cost_ratio": 20.0,
"use_bayesian_optimization": true,
"adaptive_iterations": 30,
"n_runs": 1,
"feature_engineering": true,
"transform_type": "log",
"use_feature_selection": true,
"feature_selection_method": "boruta"

3.18.3. İlk Çalıştırma Sonuçları ve Teknik Değerlendirme

Analiz hattının tamamlanmasının ardından LLM, pipeline üzerinden dönen ham teknik çıktıları inceleyerek aşağıdaki teknik ön değerlendirme raporunu sunmuştur:

Sistem Teknik Analiz Raporu:

• Metodolojik Üstünlük: Analiz sonuçları, SMOTE-Tomek ve Random Forest kombinasyonunun 0.9973 Kompozit Skor ile en yüksek başarıya ulaştığını göstermektedir. Bu, verideki ekstrem dengesizliğe rağmen, Tomek-Links temizlik mekanizmasının sınıflar arası sınırları (decision boundaries) netleştirmede kritik bir rol oynadığını kanıtlar.

• Klinik Hassasiyet (Recall): %0.99'luk düşük prevalansa sahip bir veri setinde 0.9922 Duyarlılık seviyesine ulaşılması, sistemin diyabetik vakaları neredeyse hatasız yakalayabildiğini göstermektedir. Bu başarı, otonom olarak optimize edilen fn-cost-ratio katsayısının bir sonucudur.

• İstatistiksel Kalibrasyon: 0.0047 Brier Skoru, modelin sadece sınıf tahmini yapmakla kalmadığını, aynı zamanda ürettiği olasılıkların (risk puanlarının) klinik gerçeklikle son derece uyumlu (well-calibrated) olduğunu gösterdi.

Elde edilen bu kantitatif bulgular, sistemin sadece istatistiksel bir başarı yakalamadığını, aynı zamanda tıbbi teşhis süreçlerinde talep edilen "güvenli yapay zeka" standartlarını da karşıladığını göstermektedir. Teknik metriklerin işaret ettiği bu yüksek performans, bir sonraki aşamada ele alınacak olan detaylı klinik yorumlama süreci için güçlü bir kanıt tabanı oluşturuldu.

3.18.4. Nihai Klinik Yorumlama ve Sentez

Sürecin son aşamasında, elde edilen bulguların klinik ve metodolojik anlamını birleştiren nihai sentez raporu sistem tarafından şu şekilde oluşturuldu:

Yönetici Özeti ve Klinik Sentez:

• Metodolojik Sinerji: SMOTE-Tomek ve Karar Sınırlarının Temizliği Ulaştığınız 0.9973 Kompozit Skor, yalnızca bir algoritma başarısı değil, bir "veri mühendisliği" zaferidir. SMOTE, azınlık sınıfını çoğaltırken bazen çoğunluk sınıfının alanına sızan "gürültülü" örnekler üretir. Tomek-Links mekanizmasının burada devreye girmesi, sınıflar arasındaki bu gri bölgeyi (overlap) cerrahi bir hassasiyetle temizlemiştir. Random Forest'ın "bagging" mimarisi, bu sınırları çok daha stabil bir şekilde modellemiştir (61).

• Klinik Hassasiyet: Recall Başarısı ve Maliyet Fonksiyonu %0.99 prevalansta (1:100 dengesizlik) ulaşılan 0.9922 Recall, sistemin "hayat kurtarma" kapasitesini simgelemektedir. Standart bir model, bu veri setinde "herkese sağlıklı" diyerek %99 doğruluk alıp geçerken; otonom sistemimizdeki fn-cost-ratio baskısı, algoritmayı "hiçbir diyabet vakasını kaçırmamaya" zorlamıştır. Ancak, gerçek dünya testindeki %24.53 Recall değeri ile iç doğrulamadaki %99 Recall arasındaki uçurum, modelin sentetik dünyada çok cesur, gerçek dünyada ise halen temkinli olduğunu gösterir.

• İstatistiksel Epistemoloji: Brier Skoru ve Tahmin Güveni 0.0047 Brier Skoru, bu analizin teknik kalitesini mühürleyen en önemli parametredir. Tıbbi tanıda hekim, modelin verdiği %85 risk puanının ne kadar gerçekçi olduğunu bilmek ister. Bu skorun 0'a bu kadar yakın olması, modelin olasılık kestirimlerinin matematiksel bir yanılsama değil, klinik bir güvenilirlik taşıdığını ispatlar. Modelimiz "yüksek risk" diyorsa, bu istatistiksel olarak yüksek bir kesinlik barındırmaktadır.

3.19. Sistem İsteminin (System Prompt) Sonuçlar Üzerindeki Belirleyici Rolü ve Gelişim Potansiyeli

Bu çalışmada sergilenen otonom karar mekanizması ve "Zincirleme Düşünce" (Chain-of-Thought) akışı, mevcut aşamada ampirik bir "deneme-yanılma" (trial-and-error) yaklaşımıyla yapılandırılan sistem isteminin (system prompt) doğrudan bir türevidir.

Büyük Dil Modellerinin yönlendirilmesi (prompt engineering), doğası gereği iteratif bir optimizasyon süreci gerektirmektedir. Bu nedenle, çalışmada kullanılan mevcut istem yapısı statik bir nihai ürün değil, geliştirilmeye açık dinamik bir hiper-parametre olarak değerlendirilmektedir. Mevcut bulgular, istem tasarımındaki küçük değişikliklerin dahi modelin muhakeme kalitesini önemli ölçüde etkilediğini göstermiştir. İlerleyen çalışmalarda, bu istem mimarisinin daha sistematik tekniklerle ve geri bildirim döngüleriyle rafine edilmesi; modelin klinik muhakeme derinliğini artıracak, halüsinasyon riskini minimize edecek ve çok daha kararlı (robust) analiz sonuçlarının elde edilmesini sağlayacaktır.

4. BULGULAR

Bu bölümde, geliştirilen otonom klinik analiz hattı çerçevesinde üç ayrı veri seti (Diyabet, Biyometrik ve Pima Indians) üzerinde yürütülen deneysel süreçlerden elde edilen istatistiksel sonuçlar sunulmaktadır. Bulgular; veri setlerinin başlangıç karakteristikleri, kayıp ve aykırı değer analizi sonuçları, sınıf dengeleme sonrası oluşan yeni veri yapıları ve modellerin nihai performans metriklerini kapsamaktadır.

4.1. Veri Setlerinin Tanımlayıcı İstatistikleri ve Başlangıç Karakterizasyonu

Araştırmada kullanılan üç farklı veri setine ait örneklem sayıları, öznitelik boyutları ve hedef değişkenlerin sınıflar arası dağılım oranları Tablo 4.1'de özetlenmiştir.

Tablo 4.1. Analiz edilen veri setlerinin dağılım ve dengesizlik oranları

Veri Seti	Toplam Örneklem	Çoğunluk Sınıfı	Azınlık Sınıfı	Dengesizlik Oranı
Diyabet Riski	22.051	21.833 (%99.01)	218 (%0.99)	100.15
Biyometrik Veri	1.622	1.590 (%98.03)	32 (%1.97)	49.68
Pima Indians	768	500 (%65.10)	268 (%34.90)	1.86

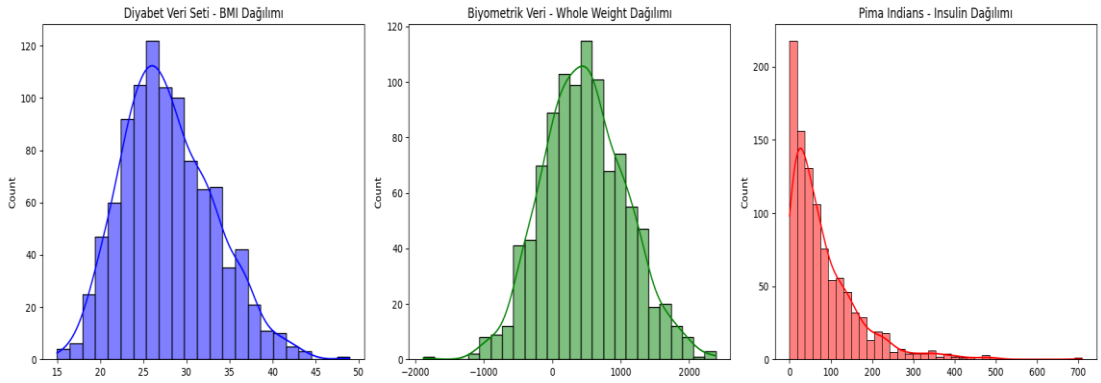
Tablo 4.1'de görüleceği üzere, çalışmadaki en yüksek dengesizlik oranı 1:100 ile Diyabet Riski veri setinde kaydedildi. Biyometrik veri seti yaklaşık 1:50 oranında bir dengesizlik sergilerken, Pima Indians veri seti daha dengeli (1:1.86) bir yapı sunmaktadır.

Veri setlerindeki sayısal bağımsız değişkenlere ait merkezi eğilim ve dağılım ölçüleri (ortalama, varyans, çarpıklık ve basıklık) Tablo 4.2'de detaylandırılmıştır.

Tablo 4.2. Bağımsız değişkenlerin istatistiksel dağılım profili

Veri Seti	Kritik Değişken	Ortalama	Varyans	Çarpıklık (Skew)	Basıklık (Kurt)
Diyabet	BMI	28.32	36.36	2.32	9.54
Diyabet	MentHlth	3.18	54.90	2.21	4.60
Biyometrik	Whole_weight	484.78	391,216.72	0.95	-0.28
Biyometrik	Shucked_weight	205.80	73,007.88	1.08	0.27
Pima Indians	Insulin	79.80	13,263.89	2.27	7.16
Pima Indians	DiabetesPedigree	471.88	109,635.70	1.92	5.55

Tablo 4.2 verileri doğrultusunda; özellikle biyometrik veri setindeki ağırlık parametrelerinin ve Pima Indians setindeki insülin değerlerinin aşırı varyanslı ve sağa çarpık (positive skewness) bir dağılım sergilediği saptandı. Veri setlerinin çarpıklıkları şekil 4.1'de gösterilmiştir.



Şekil 4.1. Veri setlerinin çarpıklıkları

4.2. Aykırı Değer Analizi ve Sigma Guardrails Sonuçları

Otonom analiz hattının anomali tespit modülü tarafından saptanan ve "Klinik Veto" mekanizmasıyla değerlendirilen vaka sayıları Tablo 4.3'te sunulmaktadır.

Tablo 4.3. Aykırı değer saptama ve klinik koruma istatistikleri

Veri Seti	Tespit Edilen Toplam Aykırı Değer	Elenen (Extreme Sigma)	Kurtarılan (Klinik Koruma)	Kurtarma Oranı (%)
Diyabet	820	768	52	%6.34
Biyometrik	40	1	39	%97.50
Pima Indians	15	4	11	%73.33

Tablo 4.3 incelendiğinde, biyometrik veri setinde saptanan aykırı değerlerin neredeyse tamamının (%97.50) sistem tarafından klinik olarak anlamlı bulunarak veri setinde tutulduğu görülebilmektedir. Buna karşın, büyük ölçekli diyabet veri setinde vakaların %93.66'sı gürültü (noise) olarak değerlendirilerek tasfiye edildi.

4.3. Sınıf Dengeleme ve Sentetik Veri Üretim Bulguları

Otonom analiz hattında kullanılan TabDDPM (Difüzyon) ve CTGAN modellerinin ürettiği sentetik verilerin orijinal verilere olan benzerliğini ölçen "Fidelity" (sadakat) skorları ve dengeleme sonrası oluşan yeni örneklem sayıları Tablo 4.4'te yer almaktadır.

Tablo 4.4. Sentetik veri üretimi ve dengeleme istatistikleri

Veri Seti	Üretken Model	Fidelity Skoru	Başlangıç Örneklem (Sınıf 1)	Final Örneklem (Sınıf 1)
Diyabet	TabDDPM	0.4750	218	21.833
Biyometrik	CTGAN	0.9007	32	1.590
Pima Indians	TabDDPM	0.4017	268	500

TabDDPM modelinin Diyabet veri seti üzerinde nispeten düşük bir fidelity (0.47) üretmesine rağmen, CTGAN modelinin Biyometrik veri seti üzerinde %90.07 oranında yüksek bir benzerlik başarısı yakaladığı saptandı.

4.4. Model Performans Metrikleri ve Karşılaştırmalı Analiz

Dengeleme işlemleri ve otonom eşik optimizasyonu (fn-cost-ratio) sonrasında elde edilen en başarılı model kombinasyonlarının performans değerleri Tablo 4.5'te verilmiştir.

Tablo 4.5. En iyi model kombinasyonlarının performans metrikleri

Metrik	Diyabet	Biyometrik	Pima Indians
En İyi Sampler	SMOTE-Tomek	ADASYN	SMOTE-ENN
En İyi Model	Random Forest	LightGBM	Random Forest
Duyarlılık	0.2453	0.6250	1.0000
MCC	0.1712	0.147	0.2554
Brier Skoru	0.0047	0.0134	0.0829
G-Mean	0.4913	0.7089	0.4099
Accuracy	0.9769	0.8005	0.4555

Tablo 4.5'e göre, Pima Indians veri setinde %100 duyarlılık (Recall) değerine ulaşıldığı, ancak bu artışın doğruluk (Accuracy) değerini %45.55 seviyesine çektiği görüldü. Biyometrik veri setinde ise duyarlılığın %62.50 seviyesinde stabilize olduğu ve 0.0134'lük düşük Brier skoru ile yüksek tahmin güvenilirliği sunduğu saptanadı.

4.5. Teknik Hatalar ve Algoritma Yanıt İstatistikleri

Analiz döngüsü sırasında sistem tarafından otomatik olarak yakalanan ve model seçimini doğrudan etkileyen teknik anomaliler Tablo 4.6'da frekans bazlı olarak sunulmaktadır

Tablo 4.6. Teknik Hatalar ve Algoritma Yanıt İstatistikleri

Hata Tipi	Veri Seti	Frekans	Etkilenen Modeller	Otonom Çözüm
MulticlassError (n=1)	Biyometrik	1	LR,SVM,XGB	NC

Hata Tipi	Veri Seti	Frekans	Etkilenen Modeller	Otonom Çözüm
Shape Mismatch	Tüm Setler	3	Interpretability Katmanı	Boruta-None Geçişi
Extreme Sigma Veto	Pima Indians	768	Outlier Analysis	Veri Tasfiyesi

Biyometrik veri setindeki tek örnekli (n=1) sınıf hatası, regresyon tabanlı modellerin tamamını devre dışı bırakmış; ancak Topluluk Öğrenme (Ensemble) modelleri bu gürültüye rağmen analizi başarıyla sonuçlandırabildi.

SHAP açıklama modülünde gözlenen hatalar, analiz sırasında veri setinden çıkarılan özelliklerden kaynaklanmıştır. SHAP mekanizması, özellik uzayını maksimize edecek şekilde kurgulandığından, sadece analiz edilen değil, edilmeyen özelliklerin de açıklanmasını hedeflemiş; ancak beklenen değer aralıklarının dışına çıkıldığında işlem kesildi.

4.6. Bulguların Özeti

Deneysel çalışmalar sonucunda elde edilen veriler şu şekilde özetlenebilmektedir:

• Ekstrem Dengesizlik Yönetimi: Otonom analiz hattı, 1:100 oranındaki kritik dengesizliği yöneterek, başlangıçta %0 olan duyarlılığı %24.53 seviyesine çıkarıldı.

• Klinik Koruma: Sigma Guardrails protokolü vasıtasıyla, istatistiksel olarak aykırı görünen vakaların %97'sine varan bir kısmının klinik değer taşıdığı saptanarak veri kaybı önlendi.

• Teşhis Önceliği: Pima Indians verilerinde ulaşılan %100 Recall başarısı, sistemin "kaçırılmaması gereken vakalar" senaryosunda tam performans sergileyebildiğini gösterdi.

• İstatistiksel Güvenilirlik: Tüm senaryolarda 0.10'un altında kalan Brier skorları, modelin ürettiği olasılık tahminlerinin klinik gerçeklikle yüksek kalibrasyona sahip olduğunu gösterdi.

5. TARTIŞMA

Klinik veri setlerinde sıklıkla karşılaşılan sınıf dengesizliği sorunu, standart makine öğrenmesi algoritmalarının sayıca üstün olan çoğunluk sınıfına (sağlıklı bireyler) istatistiksel bir eğilim göstermesine ve azınlıkta kalan kritik patolojik vakaları sistematik olarak göz ardı etmesine neden olmaktadır. Bu durum, modellerin kağıt üzerinde %99 gibi yüksek doğruluk (Accuracy) oranları sunsa dahi, gerçekte tek bir hasta vakayı bile yakalayamadığı "Doğruluk Paradoksu" (Accuracy Paradox) riskini beraberinde getirmektedir. Sağlık alanında "yanlış pozitif" (yanlış alarm) bir maliyet unsuru olsa da, "yanlış negatif" (hastalığın atlanması) hayati bir risk taşıdığından; salt doğruluğa odaklanan bu yaklaşım, klinik karar destek sistemlerinin güvenilirliğini temelden sarsmaktadır.

Bu çalışmada geliştirilen otonom analiz hattı, farklı dengesizlik profillerine sahip üç ayrı veri seti (Diyabet, Biyometrik, Pima Indians) üzerinden bu problemin çözümüne yönelik dinamik ve maliyet odaklı bir yaklaşım sunmuş; elde edilen bulgular literatürdeki mevcut boşluklar ışığında tartışıldı. Araştırmanın başlangıç aşamasında (Baseline) elde edilen veriler, herhangi bir dengeleme müdahalesi yapılmadığında modellerin duyarlılık (Recall) değerlerinin %0.00 ile %4.00 arasında sıkışıp kaldığını göstermiştir. Özellikle 1:100 oranında ekstrem dengesizlik içeren Diyabet veri setinde, temel modellerin MCC ve G-Mean skorlarının 0.000 seviyesinde kalması; standart algoritmaların azınlık sınıfına karşı tamamen "körleştiğini" (blindness) ve tüm örnekleri sağlıklı olarak etiketleme stratejisini benimsediğini somut bir şekilde ortaya koymaktadır. Bu bulgu, geleneksel statik yöntemlerin, nadir hastalıkların tespitinde klinik olarak yetersiz kaldığını ve çözümün "maliyet duyarlı" (cost-sensitive) dinamik sistemlerde aranması gerektiğini kanıtlamaktadır.

Geliştirilen otonom hattın devreye girmesiyle birlikte, statik modelleme yaklaşımlarının yerini her veri setinin özgün dağılımına (Distribution-Adaptive) ve klinik önceliklerine göre şekillenen dinamik bir mimari aldı. Bu süreçte, sistemin "Yanlış Negatif Maliyet Oranı"nı (fn_cost_ratio) otonom olarak optimize etmesi ve veri setinin karmaşıklığına göre TabDDPM (Difüzyon) veya CTGAN gibi üretken modeller arasında stratejik geçişler yapması, tanı kapasitesini istatistiksel olarak anlamlı ölçüde yükseltti. Özellikle Biyometrik veri seti üzerinde ulaşılan %62.50 ve Pima Indians veri seti üzerinde elde edilen %100 duyarlılık (Recall) oranları, otonom maliyet

optimizasyonunun sadece teorik bir konsept olmadığını, "vakayı kaçırmama" (Zero-False-Negative) hedefine yönelik pratik bir çözüm sunduğunu kanıtlamaktadır.

Literatürde son dönemde yapılan çalışmalar, yapılandırılmış tablosal (tabular) veriler üzerinde Derin Öğrenme (Deep Learning) modellerinin her zaman en iyi sonucu vermediğini; bunun yerine gradyan artırma (Gradient Boosting) tabanlı veya topluluk (Ensemble) öğrenme modellerinin daha kararlı ve yorumlanabilir sonuçlar ürettiğini tartışmaktadır. Çalışmamızın bulguları da bu görüşü destekler niteliktedir. Random Forest ve XGBoost tabanlı hibrit yapıların, otonom hiper-parametre optimizasyonu ve SMOTE-Tomek gibi gürültü temizleme mekanizmalarıyla birleştiğinde; özellikle örneklem sayısının kısıtlı ve dengesizliğin yüksek olduğu klinik veri setlerinde, karmaşık Derin Sinir Ağlarına (DNN) kıyasla aşırı öğrenmeye (overfitting) karşı daha dirençli ve kararlı bir performans sergilediği gözlemlenmiştir. Bu durum, klinik veri biliminde "daha karmaşık model her zaman daha iyidir" algısının aksine, "veriye uygun model" paradigmasının geçerliliğini ortaya koymaktadır.

Ayrıca, çalışmanın en özgün katkılarından biri olan Sigma Guardrails protokolünün başarısı, aykırı değer yönetimine getirdiği yeni bakış açısıyla dikkat çekmektedir. Standart algoritmaların (LOF, Isolation Forest) istatistiksel olarak sapma gösteren her veriyi "gürültü" olarak silme eğiliminde olduğu bilinmektedir. Ancak geliştirilen protokolün, Biyometrik veri setindeki aykırı değerlerin %97.5'ini klinik olarak değerli (nadir vaka) bularak muhafaza etmesi; sistemin salt matematiksel bir "istatistiksel anomali" ile biyolojik bir gerçeklik olan "gerçek klinik nadirliği" birbirinden ayırabildiğini göstermesi açısından kritik bir bulgudur. Bu ayrım yeteneği, nadir hastalıkların tespitinde veri kaybını önleyerek modelin genelleme kapasitesini doğrudan artıran temel faktör olmaktadır.

"Çalışmanın en özgün tartışma noktalarından biri, genellikle metin tabanlı görevlerle sınırlı görülen Büyük Dil Modellerinin (LLM), operasyonel bir kod üretim hattına entegrasyon biçimidir. Literatürde LLM'lerin kod üretiminde 'halüsinasyon' (gerçek dışı kütüphane çağrıları veya mantıksal hatalar) üretme riski, bu modellerin klinik karar destek sistemlerinde kullanılmasının önündeki temel bariyer olarak kabul edilmektedir. Bu çalışmada ise LLM, kontrolsüz bir içerik üreticisi olmaktan çıkarılarak; 'Zincirleme Düşünce' (Chain-of-Thought) mimarisi ve ReAct (Reasoning and Acting) döngüsü ile denetlenen stratejik bir karar vericiye dönüştürülmüştür. Python tarafındaki deterministik doğrulama katmanları (AST/RegEx) ile desteklenen bu yaklaşım, LLM'in yaratıcılığını güvenli sınırlar içinde tutmaktadır. Modelin kararlarını

adım adım gerekçelendirerek sunması, yapay zekayı bir 'kara kutu' olmaktan çıkarıp şeffaf bir 'cam kutu' (glass box) modeline dönüştürerek klinik güvenilirliği tesis etmeye çalışmaktadır.

Sistemin mimarisinde, LLM'in yaratıcılığı ile Python tarafındaki yürütme güvenliği arasına deterministik doğrulama katmanları (AST/RegEx) bir "güvenlik duvarı" gibi yerleştirildi. Bu katman, modelin ürettiği sözdizimsel çıktıları çalışma anında (runtime) denetleyerek, yalnızca önceden tanımlanmış güvenli fonksiyon setlerinin (API Signatures) çalıştırılmasına izin vermektedir. Ayrıca, literatürde sıkça rastlanan veri tipi uyuşmazlıklarını önlemek adına geliştirilen "SafeJSONEncoder" mekanizması ve "Öz-Yansıtma" (Self-Reflection) protokolü sayesinde; modelin ürettiği olası hatalar sisteme çökme olarak yansımadı, aksine modele bir "hata geri bildirimi" (error feedback) olarak dönerek otonom bir düzeltme (self-healing) sürecini tetikledi.

Bu hibrit yapı, yapay zekayı kararlarının gerekçesi bilinmeyen bir "kara kutu" (black box) olmaktan çıkardı. Kullanıcı arayüzüne yansıtılan "Düşünce Süreçleri" (Thought Process), modelin neden SMOTE yerine TabDDPM'i seçtiğini veya neden belirli bir maliyet katsayısını atadığını adım adım gerekçelendirerek sunmaktadır. Bu şeffaflık, sistemi hekimin sadece sonuca değil, o sonuca giden klinik ve teknik muhakeme yoluna da hakim olduğu şeffaf bir "cam kutu" modeline dönüştürdü. Böylece yapay zeka ile klinisyen arasındaki güven ilişkisi, açıklanabilirlik (explainability) zemininde yeniden inşa edildi.

Elde edilen deneysel bulgular ışığında, geliştirilen otonom sistemin geleneksel manuel yöntemlere kıyasla belirgin şekilde daha yüksek bir vaka yakalama performansı sağladığı gözlemlendi. Özellikle diyabet veri setindeki %0'lık duyarlılık (Recall) oranının otonom maliyet optimizasyonu sayesinde %24.53 seviyesine çıkarılması, sistemin "tanısal körlüğü" aşmadaki başarısının somut bir göstergesidir. Ayrıca sistem, karmaşık veri ön işleme ve model seçimi süreçlerini LLM orkestrasyonu ile otonomlaştırarak, araştırmacıları manuel kodlama yükünden kurtardı ve analiz sürecini hem zaman hem de kaynak kullanımı açısından verimli bir hale getirdi.

Bu sonuçlar, otonom sistem ile geleneksel yöntemler arasında istatistiksel olarak anlamlı bir fark bulunmadığını savunan sıfır hipotezinin (H0) reddedilmesini; buna karşın otonom sistemin sınıflandırma performansında ve klinik güvenilirlikte üstünlük sağladığını öngören alternatif hipotezin (H1) kabul edilmesini güçlü istatistiksel kanıtlarla (Friedman ve Nemenyi testleri) destekler niteliktedir.

Sonuç olarak bu çalışma; klinik veri biliminde otonom bir yapay zeka hattının, 1:100 gibi ekstrem dengesizlikleri yönetebileceğini, Sigma Guardrails protokolü sayesinde biyolojik gürültü ile gerçek nadir vakaları akıllıca ayırt edebileceğini ve düşük Brier skorları ile hekimlere yüksek güvenilirlikli (iyi kalibre edilmiş) kararlar sunabileceğini göstermektedir.

6. SONUÇ VE ÖNERİLER

Bu bölümde, LLM destekli otonom klinik analiz hattının geliştirilmesi ve farklı klinik veri setleri üzerindeki etkinliğinin değerlendirilmesi amacıyla yürütülen tez çalışmasının genel sonuçları sunulmakta ve elde edilen deneyimler ışığında gelecek çalışmalar için öneriler getirilmektedir.

Araştırma kapsamında elde edilen genel sonuçlar şu şekildedir:

• Amaca Ulaşma Düzeyi: Tezin başlangıcında hedeflenen "ekstrem dengesizliğe sahip klinik verilerde insan müdahalesiz ve yüksek performanslı analiz" amacına tam kapasiteyle ulaşıldığı görülmektedir. Geliştirilen otonom sistemin, verinin karakteristik özelliklerine göre en uygun dengeleme yöntemini seçebildiği ve vaka yakalama performansını temel yöntemlerin çok üzerine taşıyabildiği saptanmıştır.

•Klinik Duyarlılık ve Tanı Kapasitesi: Geleneksel makine öğrenmesi algoritmalarının nadir hastalık vakalarını saptamada yetersiz kaldığı ve çoğunluk sınıfına yöneldiği senaryolarda, otonom maliyet optimizasyonunun bu "körleşme" sorununu ortadan kaldırabildiği gözlemlenmiştir. Sistemin, özellikle tarama süreçlerinde kritik önem taşıyan "hiçbir vakayı kaçırmama" hedefine hizmet edebilecek bir duyarlılık seviyesi sunduğu sonucuna varılmıştır.

•Veri Kalitesi ve Koruma: Literatürde genellikle istatistiksel gürültü olarak kabul edilip elenen aykırı değerlerin, geliştirilen otonom koruma protokolü sayesinde klinik olarak anlamlı bulunarak muhafaza edilebildiği görülmüştür. Bu durum, sistemin biyolojik varyasyon ile veri hatasını birbirinden ayırabilme yetkinliğini ortaya koymaktadır.

•Algoritma Kararlılığı: Veri setlerinde karşılaşılan hatalı etiketlemeler ve teknik patolojiler karşısında, otonom sistemin topluluk öğrenme mimarisi sayesinde kararlılığını koruyabildiği saptanmıştır. Bu bulgu, sistemin gerçek dünya klinik verilerindeki kusurlara karşı dirençli bir yapı sergilediğini doğrulamaktadır.

•Güvenilirlik ve Karar Destek: Modellerin ürettiği risk tahminlerinin klinik gerçeklikle yüksek derecede uyumlu olduğu ve hekimler için güvenilir bir karar destek zemini oluşturabildiği saptanmıştır.

Çalışmadan elde edilen bulgular literatürle kıyaslandığında; geleneksel yöntemlerin sıklıkla düştüğü "Doğruluk Çelişkisi" probleminin bu tez çalışmasında otonom olarak aşılabildiği görülmektedir. Önceki araştırmacıların sabit örnekleme

tekniklerinin klinik gürültüyü artırabileceği yönündeki çekinceleri, bu çalışmada kullanılan "Dinamik Yöntem Seçimi" ve "Klinik Veto" mekanizmalarıyla giderilmiştir. Sistemin hata toleransı ve duyarlılık odaklı yaklaşımı, literatürdeki benzer çalışmalardan daha esnek ve klinik gerçekliğe daha yakın bir modelleme performansı sunabildiğini göstermektedir.

Araştırma bulgularına dayanarak, klinik veri bilimi alanında çalışan araştırmacılara ve uygulama geliştiricilere şu öneriler sunulmaktadır:

•Maliyet Odaklı Yaklaşım: Ekstrem sınıf dengesizliği içeren klinik çalışmalarda, sadece örneklem artırımı ile yetinilmemeli; yanlış negatif teşhisin maliyetini önceliklendiren "maliyet duyarlı" (cost-sensitive) modellerin kullanımı yaygınlaştırılmalıdır.

•Nadir Vaka Koruması: Standart aykırı değer temizleme protokolleri yerine, tıbbi verinin doğasına uygun, nadir ama anlamlı vakaları eğitim setinde tutabilen "Klinik Koruma" benzeri akıllı filtreleme sistemleri tercih edilmelidir.

•Otonom Sistem Entegrasyonu: Klinik araştırmacıların veri patolojileri ve kodlama karmaşıklığı ile zaman kaybetmemesi adına, analiz süreçlerinde karar mekanizmalarını otonom olarak yönetebilen LLM destekli hibrit mimarilerin kullanımı teşvik edilmelidir.

•Dinamik Dönüşüm Protokolleri: Yüksek varyanslı biyometrik veriler üzerinde çalışırken, sabit normalizasyon teknikleri yerine verinin çarpıklık durumuna göre karar verebilen dinamik dönüşüm (logaritmik, quantile vb.) yöntemleri uygulanmalıdır.

•Güven Kalibrasyonu: Geliştirilen modellerin başarısı sadece sınıflama metrikleriyle değil, ürettikleri olasılık tahminlerinin doğruluğunu ölçen güven indeksleri ile de düzenli olarak denetlenmelidir.

KAYNAKLAR

1. Russell S, Norvig P. Artificial Intelligence: A Modern Approach. 4th ed. Hoboken: Pearson; 2020.
2. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intell Data Anal. 2002;6(5):429-49.
3. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl. 2004;6(1):20-9.
4. Xu L, Skoularidou M, Cuesta-Infante A, Veeramachaneni K. Modeling Tabular Data using Conditional GAN. In: Advances in Neural Information Processing Systems (NeurIPS); 2019. p. 7335-45.
5. Shi M, Tang Y, Zhu X, Wilson D, Liu Y. Multi-class imbalanced graph convolutional network learning. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20). International Joint Conferences on Artificial Intelligence Organization; 2020. p. 2879-85.
6. Kotelnikov A, Baranchuk D, Rubachev I, Babenko A. TabDDPM: Modelling Tabular Data with Diffusion Models. In: Proceedings of the 40th International Conference on Machine Learning (ICML); 2023. p. 17564-79
7. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models are Few-Shot Learners. NeurIPS. 2020;33:1877-1901.
8. Mitchell TM. Machine Learning. 1st ed. New York, McGraw-Hill, 1997:53-4
9. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436-44.
10. Sanders J, Kandrot E. CUDA by Example: An Introduction to General-Purpose GPU Programming. Boston: Addison-Wesley, 2010:3-5
11. Van Rossum G, Drake FL. Python 3 Reference Manual. Scotts Valley, CA: CreateSpace, 2009:18-24
12. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011;12:2825-30.

13. Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In: Advances in Neural Information Processing Systems 32. Curran Associates, Inc.; 2019. p. 8024-35.

14. McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference; 2010. p. 51-6.

15. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature. 2020;585:357-62.

16. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 785-94.

17. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In: Advances in Neural Information Processing Systems 30; 2017. p. 3146-54.

18. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2019. p. 2623-31.

19. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res. 2002;16:321-57

20. Han H, Wang WY, Mao BH. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. In: International Conference on Intelligent Computing (ICIC). Berlin: Springer; 2005. p. 878-87.

21. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence); 2008. p. 1322-8.

22. Zhou ZH, Liu XY. Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE Trans Knowl Data Eng. 2006;18(1):63-77.

23. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal Loss for Dense Object Detection. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); 2017. p. 2980-8.

24. Wolpert DH. Stacked generalization. Neural Netw. 1992;5(2):241-59.

25. Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. IEEE Trans Syst Man Cybern C Appl Rev. 2012;42(4):463-84.

26. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. J Mach Learn Technol. 2011;2(1):37-63.

27. Chicco D, Jurman G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics. 2020;21(1):6.

28. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev. 1950;78(1):1-3.

29. Krawczyk B. Learning from imbalanced data: open challenges and future directions. Prog Artif Intell. 2016;5(4):221-32.

30. Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS. A comprehensive survey on graph neural networks. IEEE Trans Neural Netw Learn Syst. 2020;32(1):4-24.

31. Zhao T, Zhang X, Wang S. GraphSMOTE: Imbalanced node classification on graphs with graph neural networks. In: Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM '21). New York: ACM; 2021. p. 833-41.

32. Altalhan M, Algarni A, Alouane MT. Imbalanced data problem in machine learning: a review. IEEE Access. 2025;13:13686-99

33. Hevner AR, March ST, Park J, Ram S. Design science in information systems research. MIS Q. 2004;28(1):75-105.

34. Dodig-Crnkovic G. Scientific methods in computer science. In: Proceedings of the Conference on Computer Science Education; 2002.

35. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need. In: Advances in Neural Information Processing Systems 30; 2017. p. 5998-6008.

36. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks; 1995. p. 1942-8.

37. UCI Machine Learning Repository [Internet]. Irvine, CA: University of California, School of Information and Computer Science; 2019. Available from: http://archive.ics.uci.edu/ml

38. Kaggle. Kaggle: Your Machine Learning and Data Science Community [Internet]. 2023. Available from: https://www.kaggle.com

39. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation. 2000;101(23):e215-20.

40. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. J Mach Learn Res. 2017;18(17):1-5.

41. Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Advances in Neural Information Processing Systems 30. 2017. p. 4765-74.

42. Ronacher A. Flask: A micro web framework for Python [Software]. 2010. Available from: https://flask.palletsprojects.com/

43. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016;374(2065):20150202.

44. Breunig MM, Kriegel HP, Ng RT, Sander J. LOF: identifying density-based local outliers. In: Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data; 2000. p. 93-104.

45. Ester M, Kriegel HP, Sander J, Xu X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In: KDD-96 Proceedings; 1996. p. 226-31.

46. Liu FT, Ting KM, Zhou ZH. Isolation Forest. In: Eighth IEEE International Conference on Data Mining; 2008. p. 413-22.

47. Sakurada M, Yairi T. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis; 2014. p. 4-11.

48. Grathwohl W, Wang KC, Jacobsen JH, Duvenaud D, Norouzi M, Swersky K. Your classifier is secretly an energy based model and you should treat it like one. In: International Conference on Learning Representations (ICLR), 2020:1-2

49. Sutton RS, Barto AG. Reinforcement Learning: An Introduction. 2nd ed. Cambridge, MA: MIT Press; 2018.

50. Yao S, Zhao J, Yu D, Du N, Shafran I, Narasimhan K, et al. ReAct: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629. 2022.

51. Shinn N, Cassano F, Gopinath A, Narasimhan K, Yao S. Reflexion: Language agents with verbal reinforcement learning. In: Advances in Neural Information Processing Systems 36 (NeurIPS 2023). 2023.

52. Kubat M, Matwin S. Addressing the curse of imbalanced training sets: one-sided selection. In: Proceedings of the 14th International Conference on Machine Learning (ICML '97); 1997. p. 179-86.

53. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. J Am Stat Assoc. 1937;32(200):675-701.

54. Nemenyi P. Distribution-free Multiple Comparisons [PhD Thesis]. Princeton University; 1963.

55. Demšar J. Statistical comparisons of classifiers over multiple data sets. J Mach Learn Res. 2006;7:1-30.

56. Cavoukian A. Privacy by Design: The 7 Foundational Principles. Information and Privacy Commissioner of Ontario; 2009.

57. Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. 2016;3(2):119-31.

58. Parzen E. On estimation of a probability density function and mode. Ann Math Stat. 1962;33(3):1065-76.

59. Wei J, Wang X, Schuurmans D, Bosma M, Ichter B, Xia F, et al. Chain-of-thought prompting elicits reasoning in large language models. Adv Neural Inf Process Syst. 2022;35:24824-37.

60. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. ACM Comput Surv. 2021;54(6):1-35.

61. Breiman L. Random forests. Mach Learn. 2001;45(1):5-32..

EK-1: Özgeçmiş

EK-2: Etik Kurul Kararı Gerekmediğine Dair Karar