

HISTORY OF MENTAL ILLNESS TAHMİNİ

BİR SINIFLANDIRMA PROBLEMİ
ÜZERİNE MAKİNE ÖĞRENMESİ
UYGULAMASI

KONU: GEÇMİŞTE PSİKOLOJİK RAHATSIZLIK
YAŞANIP YAŞANMADIĞININ TAHMINİ

HEDEF DEĞİŞKEN: HISTORY OF MENTAL ILLNESS

KULLANILAN MODELLER: LOGISTIC REGRESSION,
KNN, RANDOM FOREST

Grup No: 5

Hazırlayanlar:

Sude Nur Bulut

Yağmur Umutlu

Pınar Güner

İÇİNDEKİLER (CONTENTS)

1. Giriş ve Problem Tanımı
2. Literatür Özeti
3. Projenin Amacı
4. Veri Seti ve Yöntem
5. Kullanılan Modeller
6. Model Karşılaştırmaları
7. Sonuçlar
8. Tartışma ve Öneriler
9. Kaynaklar

GİRİŞ VE PROBLEM TANIMI

Ruh sağlığı bozuklukları, bireylerin günlük yaşam işlevselliğini, üretkenliğini ve sosyal ilişkilerini doğrudan etkileyen küresel bir halk sağlığı problemidir. Dünya Sağlık Örgütü (WHO), ruhsal bozuklukları dünya genelinde engelliliğin önde gelen nedenleri arasında göstermektedir.

Bu çalışma, geçmişte ruhsal bir rahatsızlık yaşayıp yaşanmadığını tahmin etmeye odaklanarak, hem bireysel hem de kurumsal düzeyde önleyici sağlık politikalarına katkı sağlamayı amaçlamaktadır.

LİTERATÜR ÖZETİ

Literatürde ruh sağlığı tahminine yönelik çalışmalar, istatistiksel yöntemler ile makine öğrenmesi algoritmalarının birlikte kullanımının güçlü sonuçlar ürettiğini göstermektedir.

1

WHO raporları, geçmiş ruhsal hastalık öyküsünün; uyku bozuklukları, kronik fiziksel hastalıklar ve işsizlik gibi faktörlerle birleştiğinde gelecekteki psikolojik riskleri anlamlı düzeyde artırdığını ortaya koymaktadır.

2

APA çalışmaları, geçmişte yaşanan ruhsal bozuklukların bireyin stres toleransını düşürdüğünü ve tekrar eden psikolojik sorunlara zemin hazırladığını vurgulamaktadır.

LİTERATÜR ÖZETİ

3

Makine öğrenmesi literatüründe Logistic Regression, yorumlanabilirliği nedeniyle sağlık alanında sık tercih edilmektedir.

4

KNN, benzer birey profilleri üzerinden tahmin yaparak özellikle davranışsal verilerde etkili sonuçlar sunmaktadır.

5

Random Forest, değişkenler arası karmaşık ve doğrusal olmayan ilişkileri yakalama gücü sayesinde ruh sağlığı verilerinde yüksek performans sağlamaktadır.

Bu çalışmalar doğrultusunda, farklı model türlerinin karşılaştırılması literatürle uyumlu ve bilimsel açıdan anlamlıdır.

PROJENİN AMACI

Bu projenin amacı:

- Bireyin (geçmişte psikolojik bir rahatsızlık yaşayıp yaşamadığını) tahmin eden bir sınıflandırma modeli geliştirmek,
- Farklı makine öğrenmesi yöntemlerini karşılaştırarak en iyi performansı gösteren modeli belirlemek,
- Tıbbi kayıtlara doğrudan erişim olmadan psikolojik risk tahmini yapılabilmesini sağlamaktır.

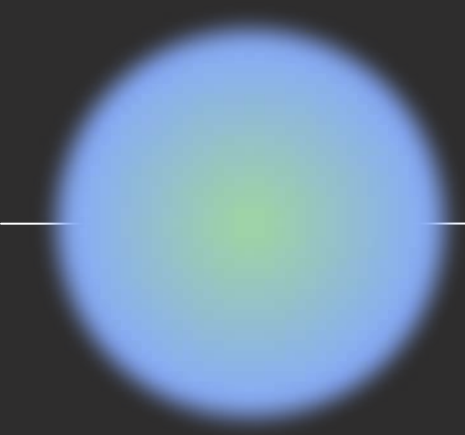
Geliştirilen model; sigorta şirketleri, kamu sağlık programları ve erken uyarı sistemleri için karar destek aracı olarak kullanılabilir.

VERİ SETİ VE YÖNTEM

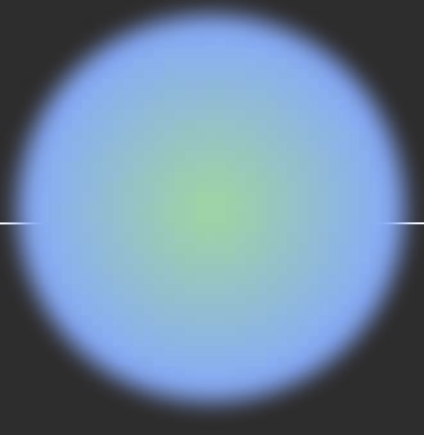
Veri Seti:

- Bu çalışmada kullanılan veri seti, bireylerin demografik özellikleri, sağlık durumları ve yaşam koşullarına ilişkin değişkenlerden oluşmaktadır.

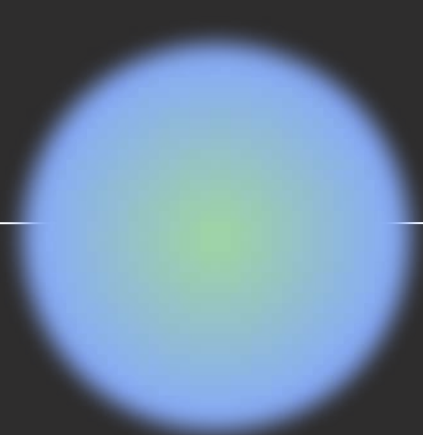
BAĞIMSIZ DEĞİŞKENLER



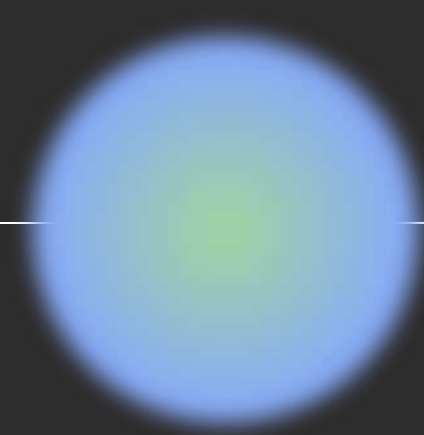
Uyku
Düzeni



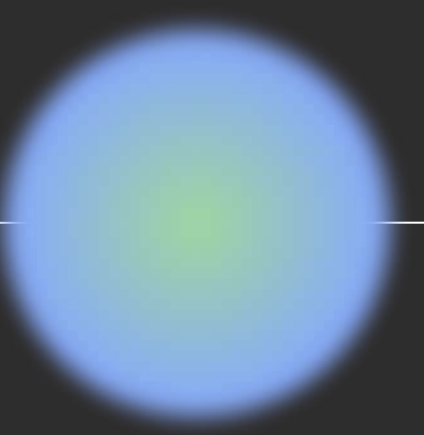
Kronik
Hastalık
Varlığı



İstihdam
Durumu



Günlük Stres
Seviyesi

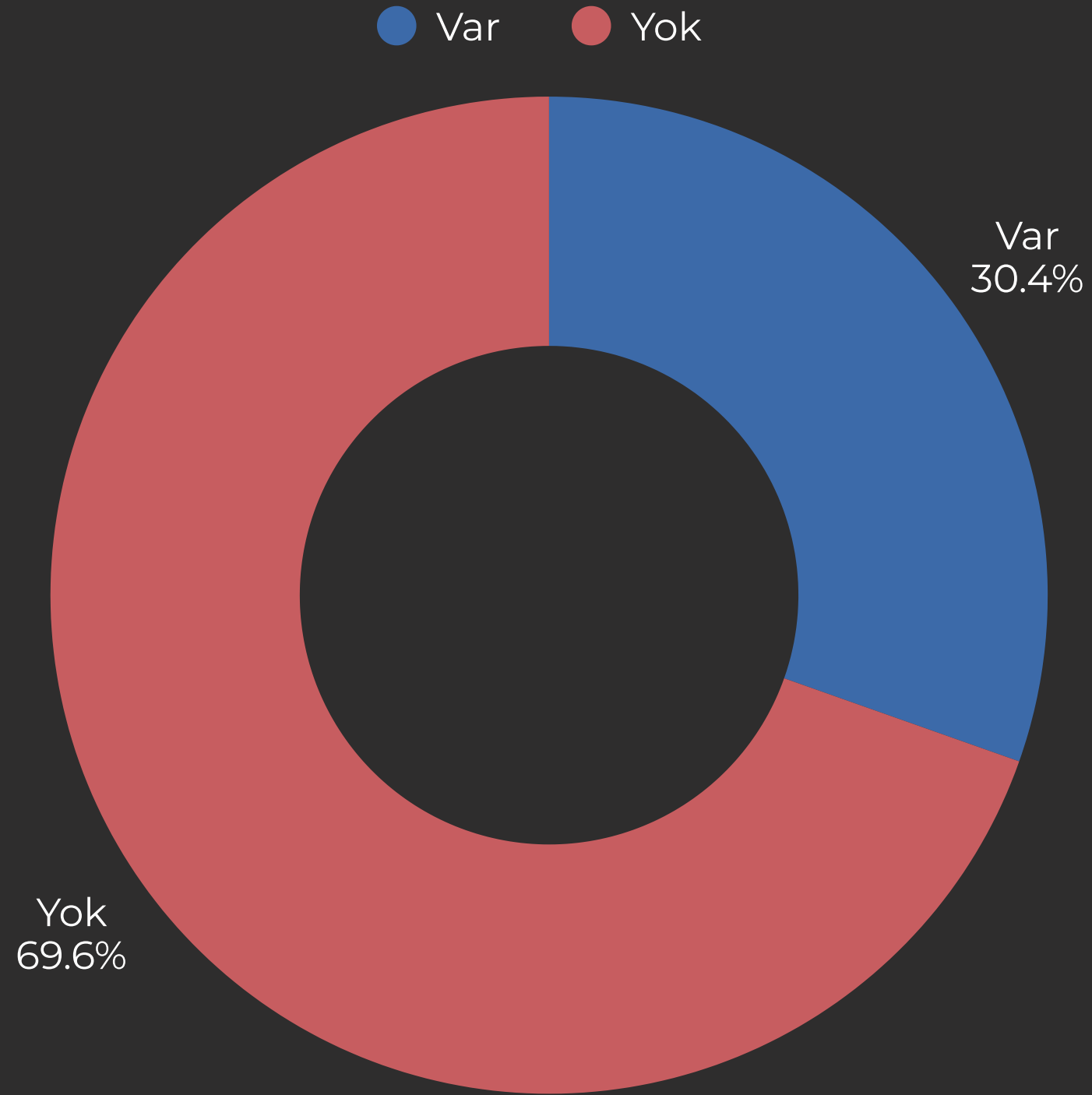


Sosyal ve
Çevresel
Faktörler

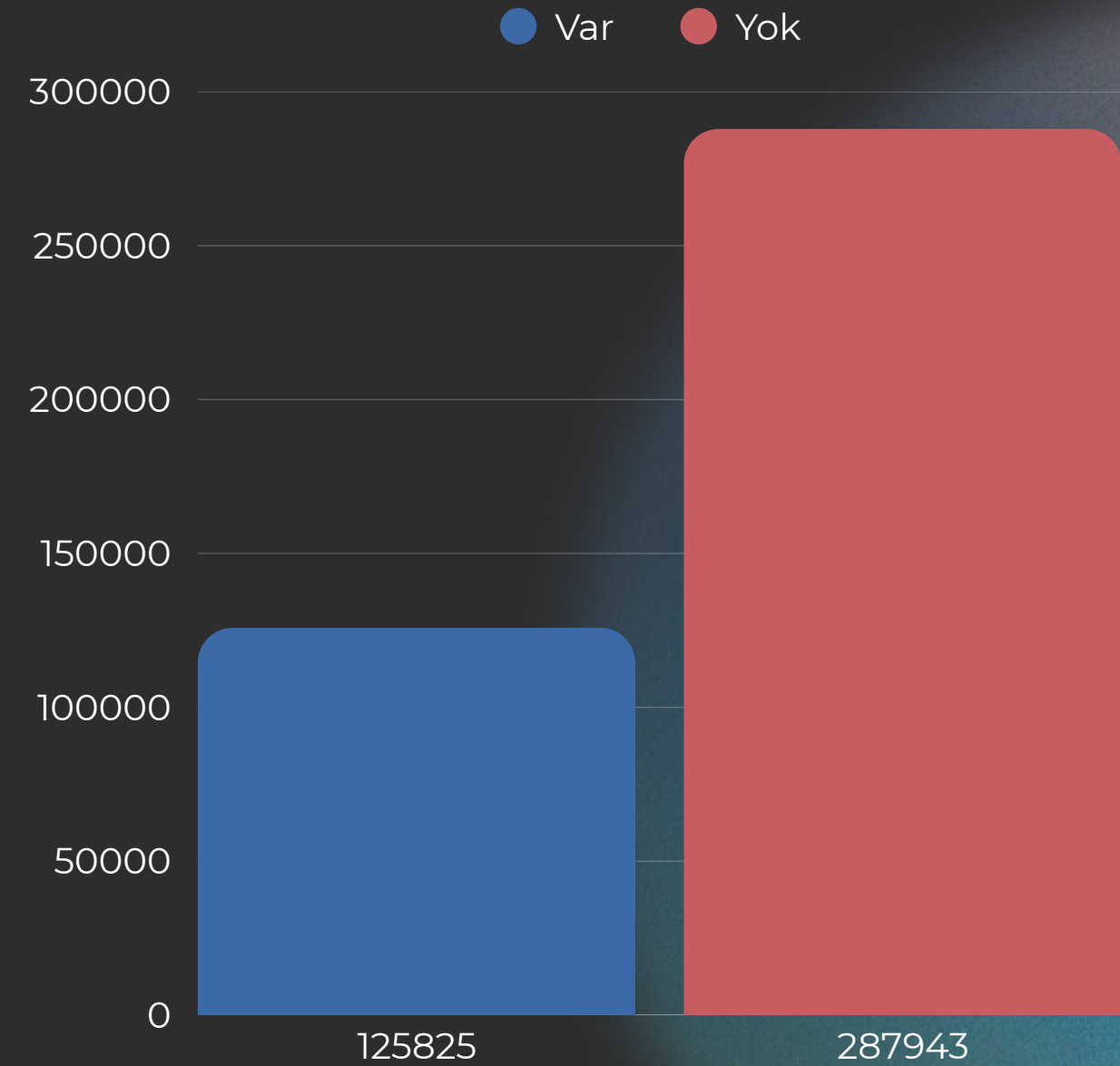
Hedef değişken: History of Mental Illness (İkili sınıflandırma)

HEDEF DEĞİŞKEN ANALİZİ

Mental Sağlık Geçmişi Dağılımı

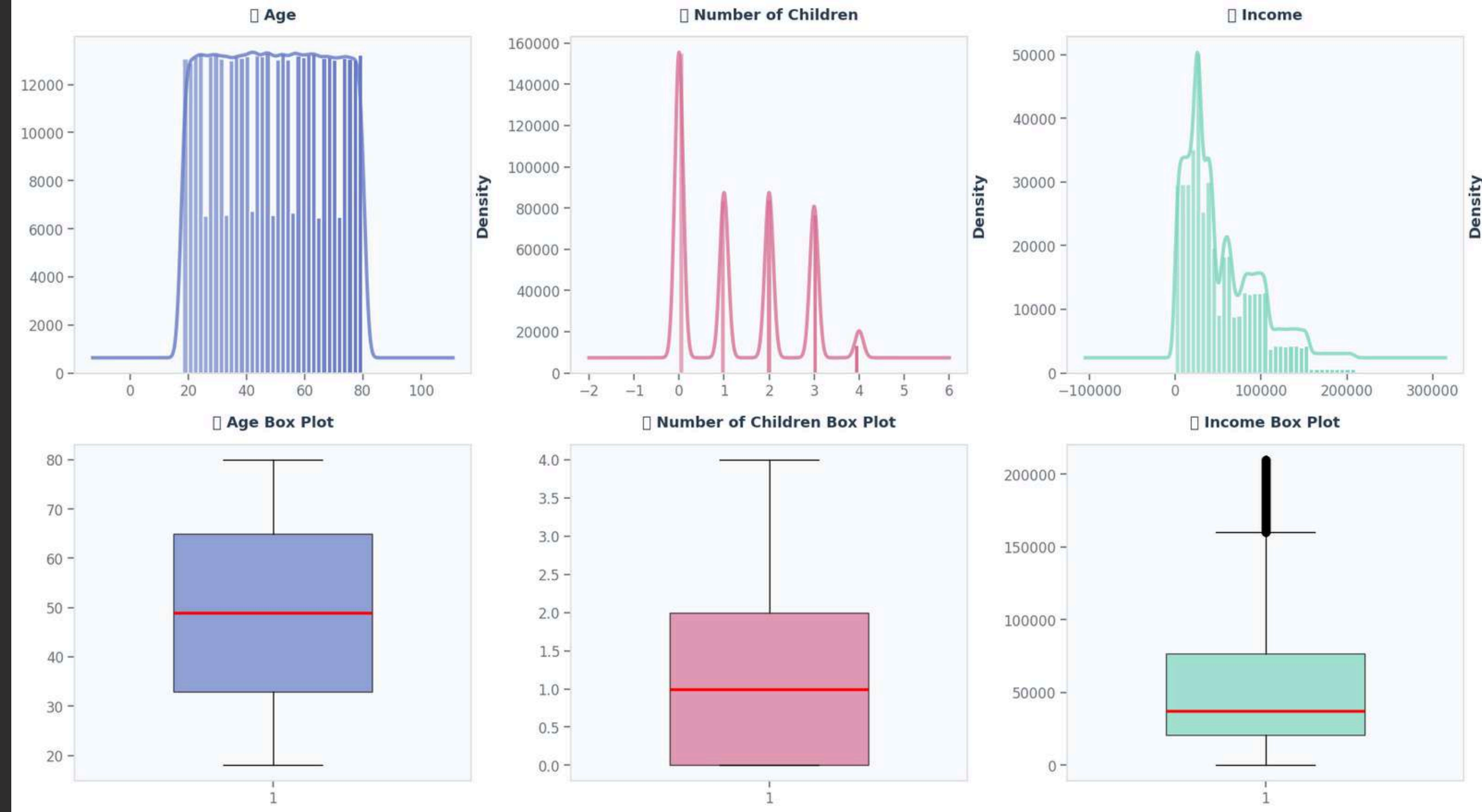


Sınıf Dağılımı



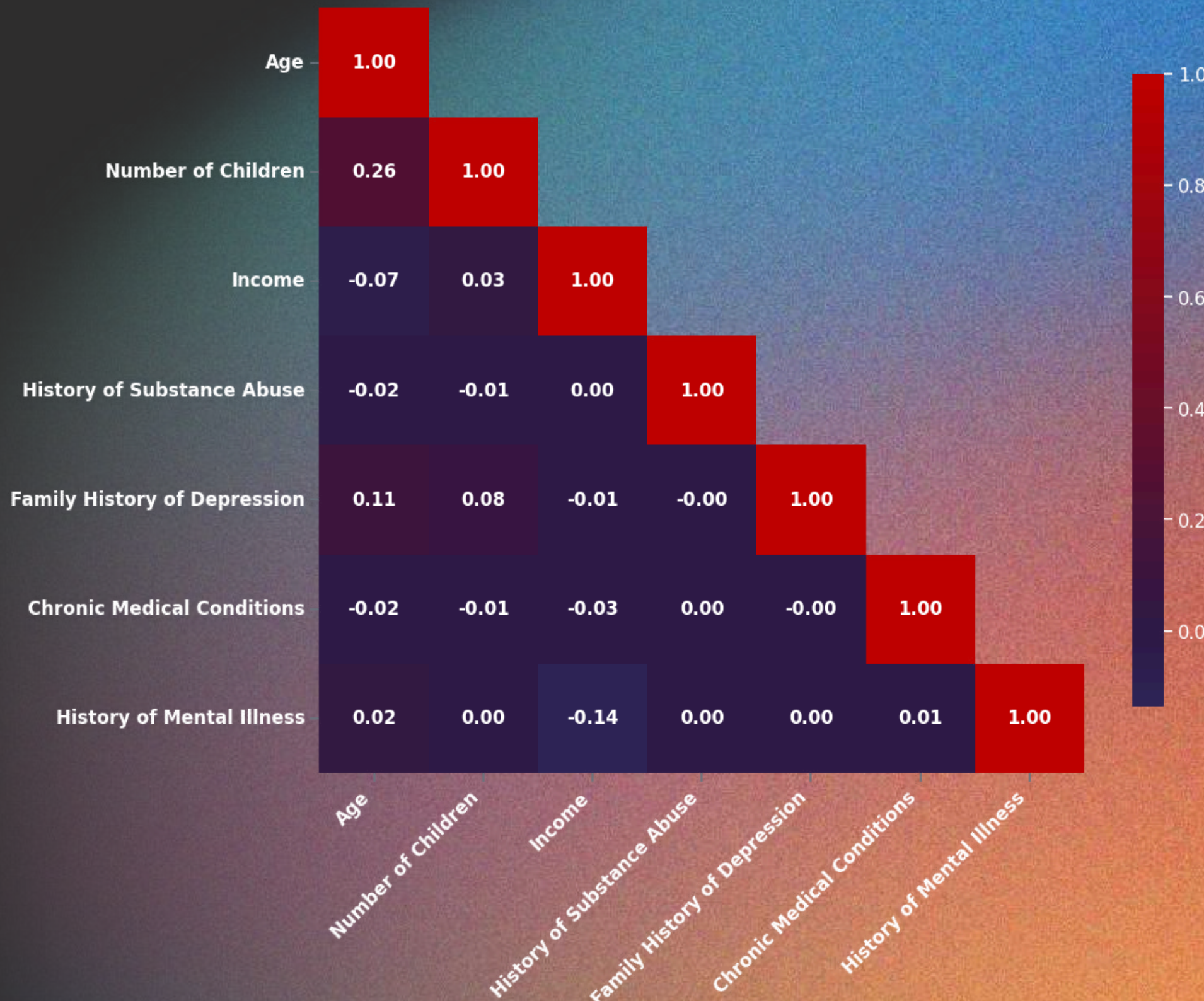
- **Veri Dağılımı:** Yaş, Çocuk Sayısı ve Gelir değişkenlerinin frekans dağılımları (Histogram) incelendi.
- **Aykırı Değer Tespiti:** Kutu Grafikleri (Box Plot) ile veri setindeki uç değerler (Outliers) analiz edildi; özellikle gelir seviyesinde sapmalar gözlemlendi.
- **Ölçeklendirme Kararı:** Değişkenler arasındaki büyük ölçek farkları (Örn: Yaş 0-80, Gelir 0-100k+) nedeniyle, model performansını artırmak için StandardScaler uygulanmasına karar verildi.

SAYISAL DEĞİŞKENLER DAĞILIM ANALİZİ



DEĞİŞKENLER ARASI İLİŞKİLER: KORELASYON ANALİZİ

Korelasyon Matrisi



Amaç: Özelliklerin birbirleriyle ve hedef değişken (History of Mental Illness) ile olan ilişkisini incelemek.

Çoklu Bağlantı (Multicollinearity) Kontrolü: Lojistik Regresyon modelinin istikrarını bozacak "aşırı benzer" değişkenler taranmıştır. Değişkenler arasında kritik bir çakışma (korelasyon > 0.70) gözlenmemiştir.

Temel Bulgu:

- Hedef değişken ile en belirgin ilişki Gelir (Income) seviyesinde görülmüştür (Katsayı: -0.14).
- Bu negatif korelasyon, gelir düzeyi arttıkça mental hastalık geçmişi riskinin düştüğüne işaret etmektedir (Koruyucu faktör).

VERİ ÖN İŞLEME

DATA PREPROCESSING

1

VERİ TEMİZLİĞİ VE KONTROLÜ:

Veri setindeki eksik ve tutarsız gözlemlerin analizi yapılmış, veri bütünlüğü sağlanmıştır.

3

ÖZELLİK MÜHENDİSLİĞİ (FEATURE ENGINEERING):

Modelin öğrenme kapasitesini artırmak için mevcut verilerden "Bileşik Risk Skoru" ve "Yaş Grupları" gibi yeni öznitelikler türetilmiştir.

2

SAYISALLAŞTIRMA (ENCODING):

Makine öğrenmesi modellerinin işleyebilmesi için kategorik değişkenler (Örn: Cinsiyet, Aile Geçmişi), Label Encoding yöntemiyle sayısal değerlere dönüştürülmüştür.

4

VERİ ÖLÇEKLENDİRME (SCALING):

Yaş ve Gelir gibi farklı birimlerdeki verilerin modeli yanıltmasını önlemek için StandardScaler ile standardizasyon uygulanmıştır.

MODELLEME VE DEĞERLENDİRME STRATEJİSİ

1

VERİ BÖLME (TRAIN-TEST SPLIT):

Veri seti, hedef değişkenin dağılım oranı korunarak (Stratified Split) %80 Eğitim ve %20 Test kümesi olarak ayrılmıştır.

2

ADIL KARŞILAŞTIRMA ORTAMI:

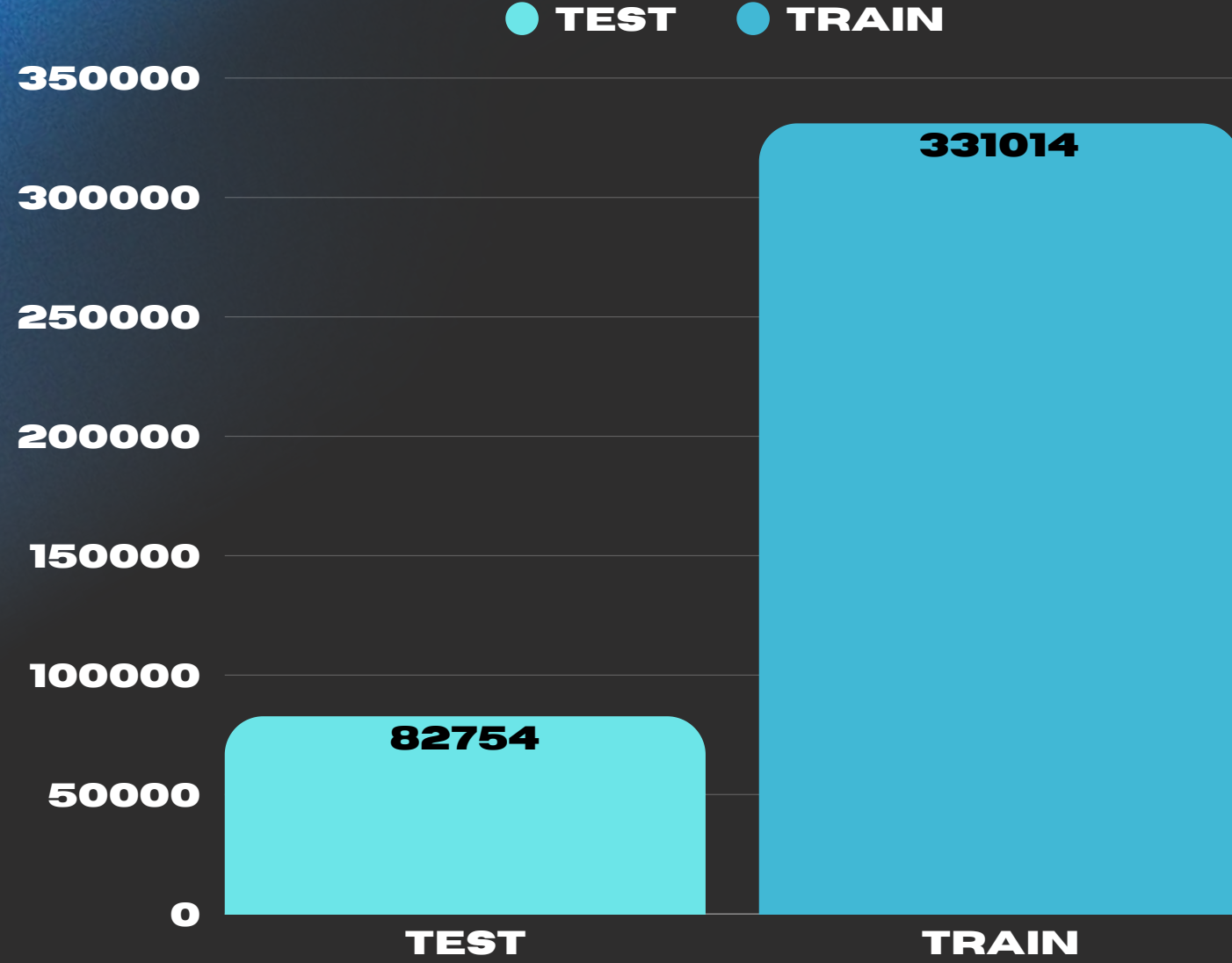
Lojistik Regresyon, KNN ve Random Forest modellerinin her biri, sonuçların tutarlı olması için aynı işlenmiş veri seti üzerinde eğitilmiştir.

3

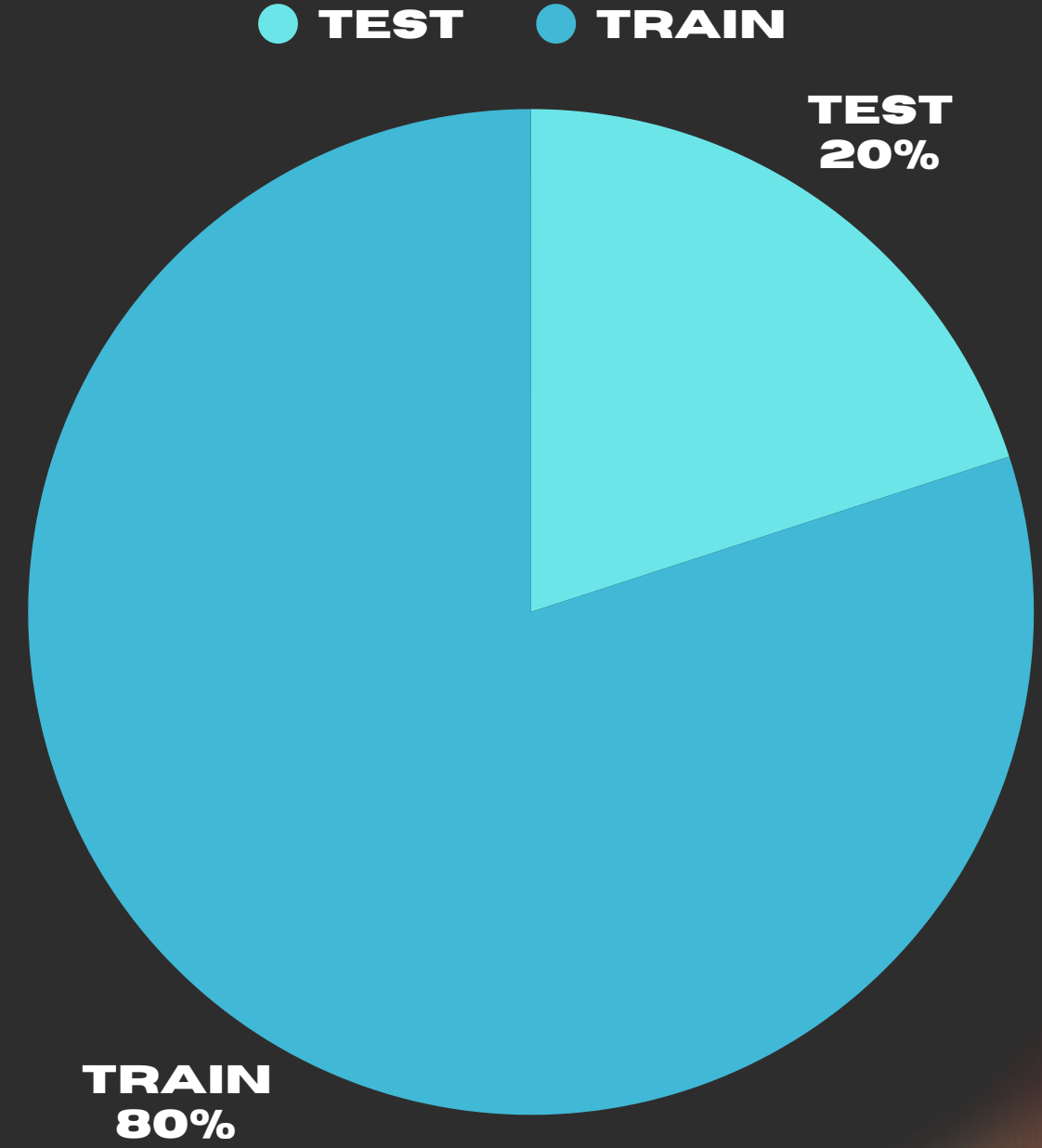
ÇOK YÖNLÜ DEĞERLENDİRME:

Modellerin başarısı sadece Doğruluk (Accuracy) ile değil; sağlık verilerinde kritik olan Recall (Duyarlılık), F1-Score ve AUC metrikleri ile ölçülmüştür.

TRAIN-TEST SPLIT



VERİ BÖLME ORANI



- **Bölme Oranı:** Veri seti %80 Eğitim (Train) ve %20 Test olarak ikiye ayrılmıştır.
- **Amaç:** Modelin daha önce hiç görmediği veriler üzerindeki gerçek performansını ölçmek.
- **Overfitting (Ezberleme) Önlemi:** Modelin veriyi ezberlemesini engelleyip, kuralları öğrenmesini sağlamak.
- **Stratified Split:** Veri集中的 hedef değışken (Mental Illness) dağılımını korumak için katmanlı bölme uygulanmıştır.

**Logistic
Regression**



**K-Nearest
Neighbors (KNN)**



Random Forest



**KULLANILAN
MODELLER**

Regresyon Değil, Sınıflandırma

İsminde "Regresyon" geçse de, verileri "Riskli" (1) veya "Risksiz" (0) olarak ayıran temel bir sınıflandırma algoritmasıdır.

Sigmoid Fonksiyonu (S Eğrisi)

Doğrusal modellerin aksine, verileri Sigmoid fonksiyonundan geçirerek sonuçları 0 ile 1 arasına sıkıştırır.

Karar Mekanizması (Threshold)

Model bir "olasılık" değeri üretir (Örn: 0.85).
Genel kabul gören 0.50 eşik değerini aşan olasılıklar "Pozitif (Riskli)" sınıfına atanır.

LOJİSTİK REGRESYON

LOJİSTİK

REGRESYON

NEDEN LOJİSTİK REGRESYON?



Tıbbi Altın Standart (Gold Standard):

Sağlık literatüründe hastalık risk faktörlerini belirlemede en güvenilir ve kabul görmüş yöntemdir.



Açıklanabilirlik (White Box):

"Kara Kutu" (Black Box) modellerin aksine şeffaftır. Sadece kimin hasta olduğunu değil, neden hasta olduğunu da açıklar.



Odds Ratio (Risk Çarpanı):

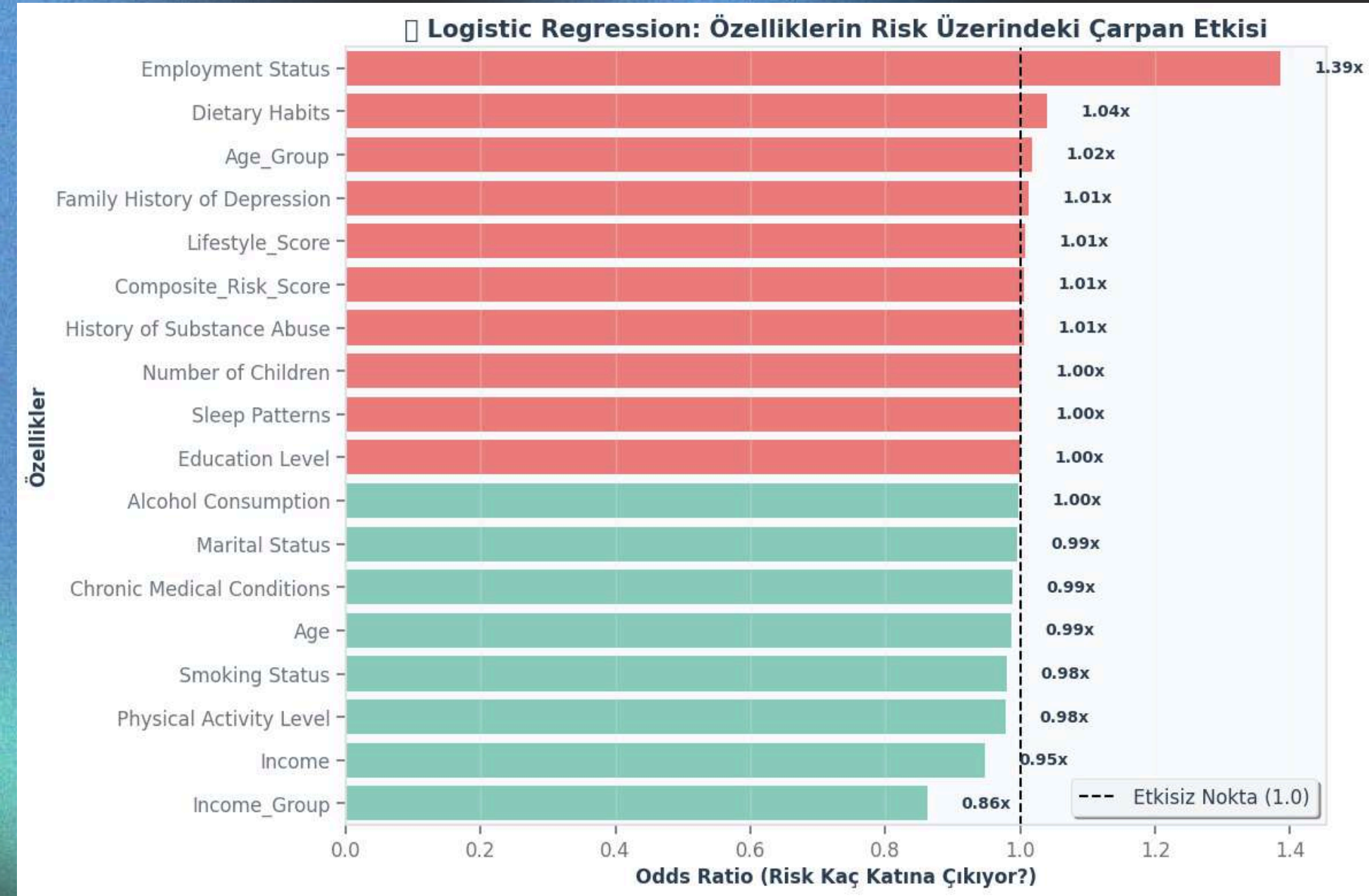
Hangi özelliğin riski matematiksel olarak kaç kat artırdığını (Örn: "Uyku bozukluğu riski 2 kat artırıyor") hesaplama imkanı sunar.

MODEL BULGULARI: RISK FAKTÖRLERİNİN ETKİSİ

Bu analiz, özelliklerin ruhsal hastalık geçmişi riski üzerindeki çarpan etkisini (Odds Ratio) göstermektedir; 1.0 değeri etkisiz referans noktasıdır.

- **En Güçlü Risk Faktörü:** İstihdam Durumu (Employment Status), 1.39x ile riski en belirgin şekilde artıran faktördür.
- ➡ **Diğer Risk Artırıcılar:** Beslenme Alışkanlıkları (1.04x) ve Yaş Grubu (1.02x) riski sınırlı düzeyde artırmaktadır.
- **Etkisiz Faktörler:** Çocuk Sayısı, Uyku Düzeni ve Eğitim Seviyesi gibi faktörlerin risk üzerinde bir çarpan etkisi (1.00x) görülmemiştir.
- **En Güçlü Koruyucu Faktör:** Yüksek Gelir Grubu (Income_Group), 0.86x çarpanı ile riski en fazla düşüren faktör olarak tespit edilmiştir.

LOJİSTİK REGRESYON



RANDOM FOREST

Rastgele Orman (Random Forest), birden fazla karar ağacını büyütüp birleştirerek bir "orman" oluşturan güçlü ve çok yönlü bir denetimli makine öğrenmesi algoritmasıdır

Nasıl Çalışır?

1. Veri setinden rastgele örnekler seçilir
2. Her örnek için farklı bir karar ağacı eğitilir
3. Her ağaç kendi tahminini yapar
4. Sonuçlar birleştirilir
 - Sınıflandırma → çoğunluk oyu
 - Regresyon → ortalama değer

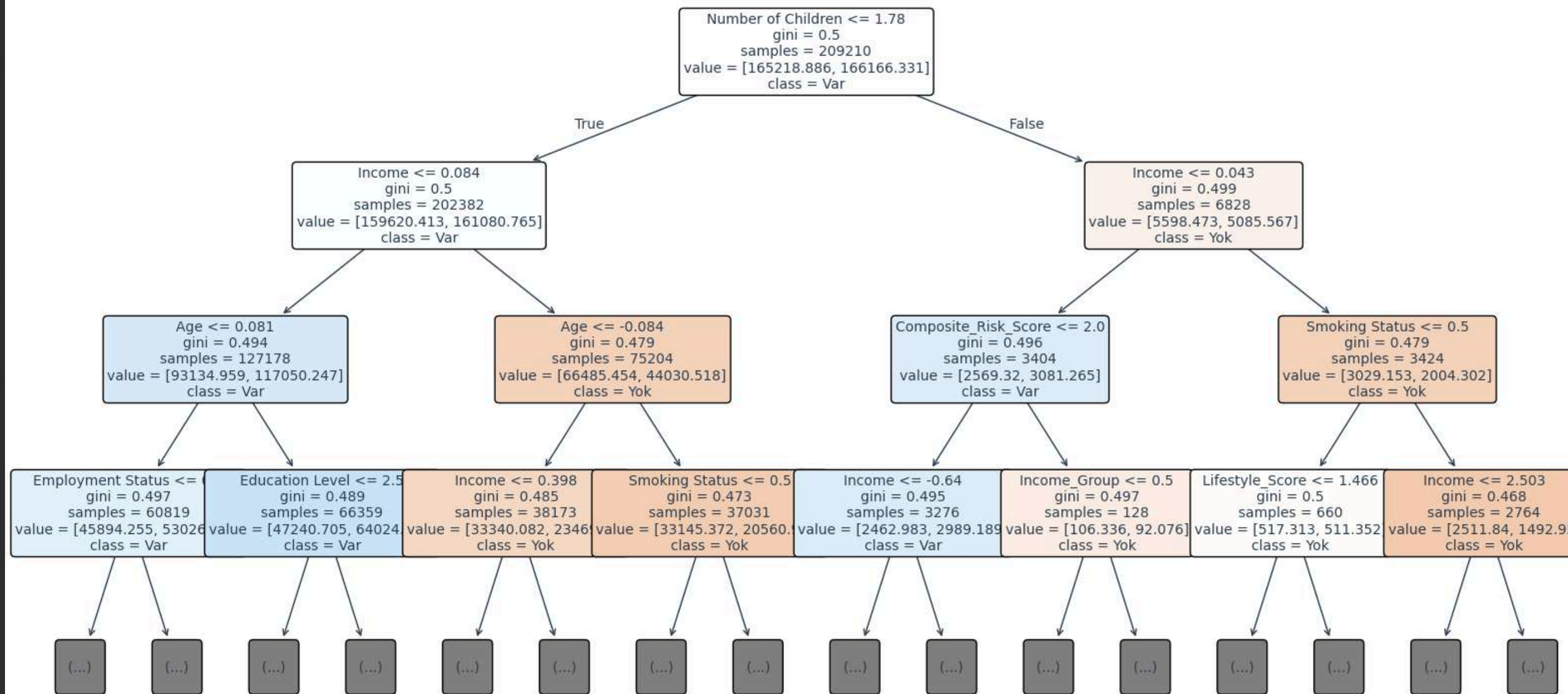
RASTGELE ORMAN ALGORİTMASININ TEMEL ÖZELLİKLERİ

1. **Çeşitlilik** - Bireysel bir ağaç oluşturulurken tüm nitelikler, değişkenler veya özellikler dikkate alınmaz; her ağaç benzersizdir.
2. **Boyutluluk kısıtlamasına karşı bağışıklıdır** - Her ağaç tüm özellikleri dikkate almadığı için özellik alanı en aza indirilir.
3. **Paralleleştirme** - Her ağaç, farklı veriler ve özellikler kullanılarak sıfırdan oluşturulur. Bu, rastgele ormanlar oluşturmak için CPU'yu tam olarak kullanabileceğimiz anlamına gelir.
4. **Eğitim-Test ayrımı yapmıyor** - Rastgele orman algoritmasında, karar ağacı verilerin %30'unu her zaman kaçıracağı için eğitim ve test verilerini ayırmaya gerek yoktur.
5. **İstikrar** - Sonuç, çoğunluk oylaması/ortalama alma yöntemine dayandığı için istikrarlıdır

RASTGELE ORMAN ALGORİTMASININ ÇALIŞMA PRENSİBİ

(Bu görsel, HISTORY OF MENTAL ILLNESS değişkeninin Random Forest modeli içinde yer alan bir karar ağacının yapısını göstermektedir.).

□ Random Forest İçindeki Bir Karar Ağacının Mantığı (İlk 3 Seviye)



- En üstte tüm veri setini temsil eden kök düğüm bulunur.
- Ağaç, her seviyede bir değişkene göre ikiye ayrılarak aşağı doğru dallanır.
- Her kutu bir karar noktasını, dallar ise izlenen karar yollarını temsil eder.
- En alt seviyedeki kutular, modelin ulaştığı nihai tahminleri göstermektedir.

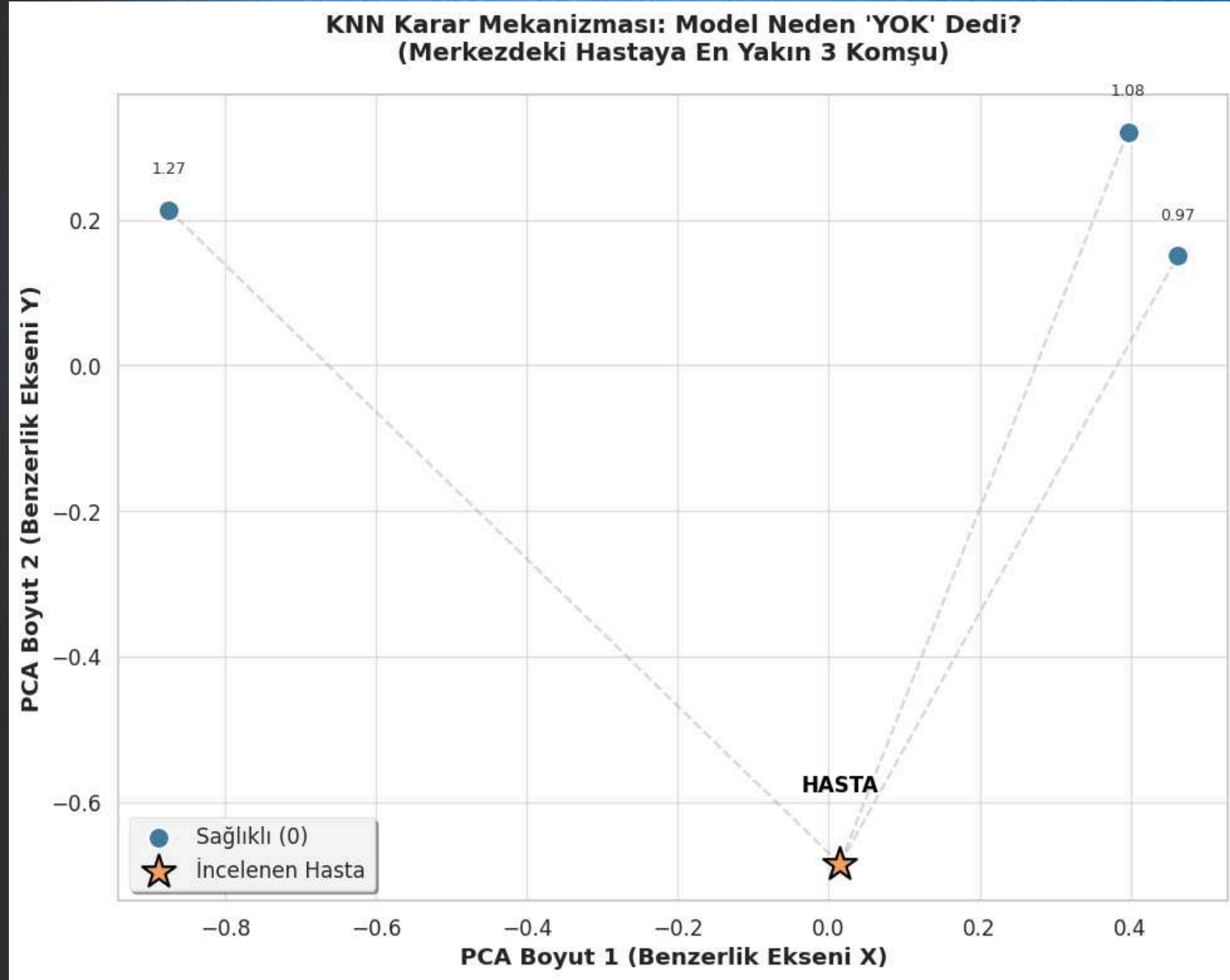
K-NEAREST NEIGHBORS (KNN)

- K-Nearest Neighbors (KNN), Türkçe adıyla K-En Yakın Komşu algoritması, makine öğrenmesi dünyasının en anlaşılır ve en popüler algoritmalarından biridir.
 - Bunu en basit şu atasözüyle özetleyebiliriz: "Bana arkadaşını söyle, sana kim olduğunu söyleyeyim."
 - KNN algoritması tam olarak bu mantıkla çalışır. Yeni bir verinin ne olduğunu anlamak için, ona en yakın olan diğer verilere (komşularına) bakar.
-
- KNN gibi benzerliğe dayalı algoritmalar, insan davranışlarının karmaşık ve yerel örüntülerini yakalamada, diğer formül tabanlı modellere göre daha iyi performans gösterir, çünkü insanlar genellikle kendilerine en çok benzeyenlerin izlediği yolu takip eder.

KNN NASIL KARAR VERIR?

- **K Değerini Seç:** Kaç komşuya bakılacağına karar verilir (Örn: $K=3$ veya $K=5$).
- **Mesafeyi Hesapla:** Yeni nokta ile eğitim setindeki tüm noktalar arasındaki uzaklık bulunur.
- **K Komşuyu Bul:** En kısa mesafeye sahip ilk K komşu seçilir.
- **Tahmin Et (Çoğunluk Oyu):** K komşu arasında en çok hangi sınıf varsa, yeni nokta o sınıfa atanır.

K-EN YAKIN KOMŞU



Bu grafik ve tablo, k-NN modelinin merkezdeki HASTA'nın durumunu neden 'YOK' olarak tahmin ettiğini net bir şekilde gösterir.

KNN birimin kuralı olan çoğunluk oylaması sonucunda, model de bu yeni hastanın durumunu en yakın komşularının sınıfı olan 'YOK' olarak etiketlemiştir. Modelin tahmini, gerçek durumu da ● YOK olduğu için, bu durumda modelin doğru bir tahmin yaptığı görülmektedir.

İNCelenen HASTA GERÇEK DURUM: ● YOK

Komşu #	Durum	Uzaklık
---------	-------	---------

1	● YOK (0)	0.9684
---	-----------	--------

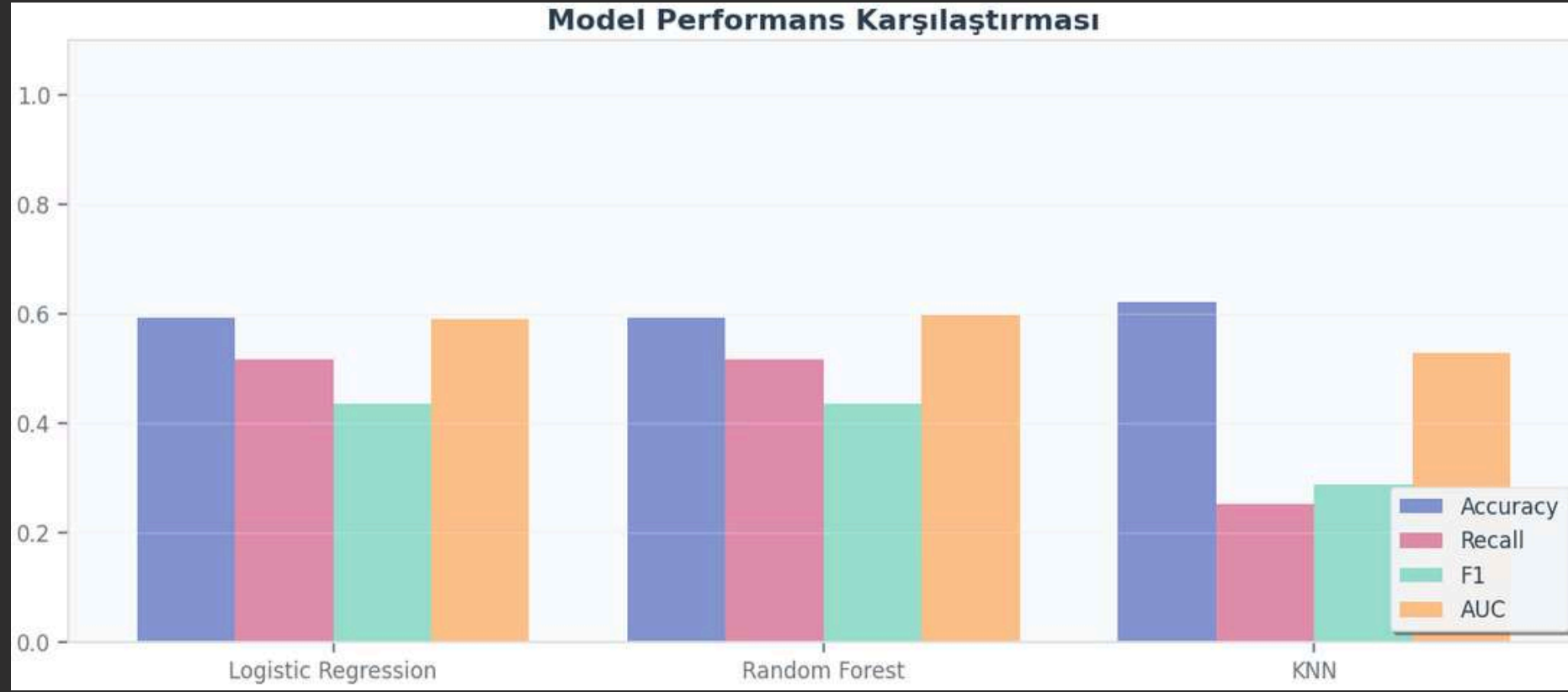
2	● YOK (0)	1.0842
---	-----------	--------

3	● YOK (0)	1.2658
---	-----------	--------

KARAR: 0 Kişi 'VAR', 3 Kişi 'YOK' dedi.

SONUÇ: Model 'YOK' tahmininde bulundu.

MODEL PERFORMANS KARŞILAŞTIRMASI

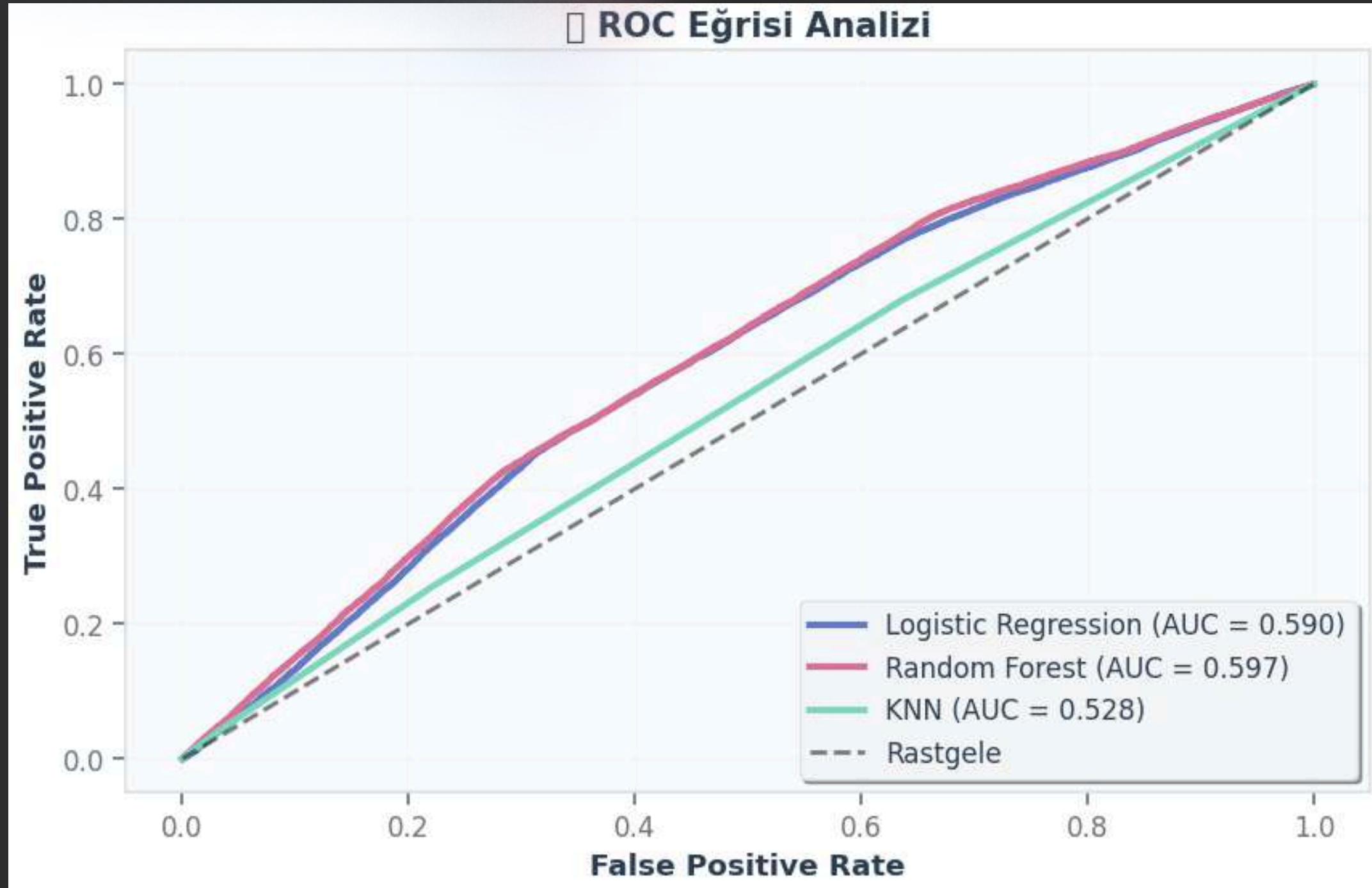


- Dengeli Performans:** Logistic Regression ve Random Forest modelleri, %59 Doğruluk ve %51 Recall oranlarıyla birbirine çok yakın ve istikrarlı sonuçlar vermiştir.
- KNN Yanılgısı (Accuracy Paradox):** KNN modeli %62 ile en yüksek doğruluğa sahip görünse de, %25'e düşen Recall değeri modelin pozitif vakaları yakalamada başarısız olduğunu gösterir.

	ACCURACY	RECALL	F1	AUC
LOGISTIC REGRESSION	0.5915	0.5152	0.4341	0.5899
RANDOM FOREST	0.5913	0.5155	0.4341	0.5974
KNN	0.6204	0.2529	0.2884	0.5283

- Sonuç:** Yüksek doğruluk tek başına yeterli değildir; veri setindeki dengesizlik göz önüne alınarak, daha güvenilir tahmin yapan Logistic Regression veya Random Forest modelleri tercih edilmelidir.

ROC EĞRİLERİ ANALİZİ



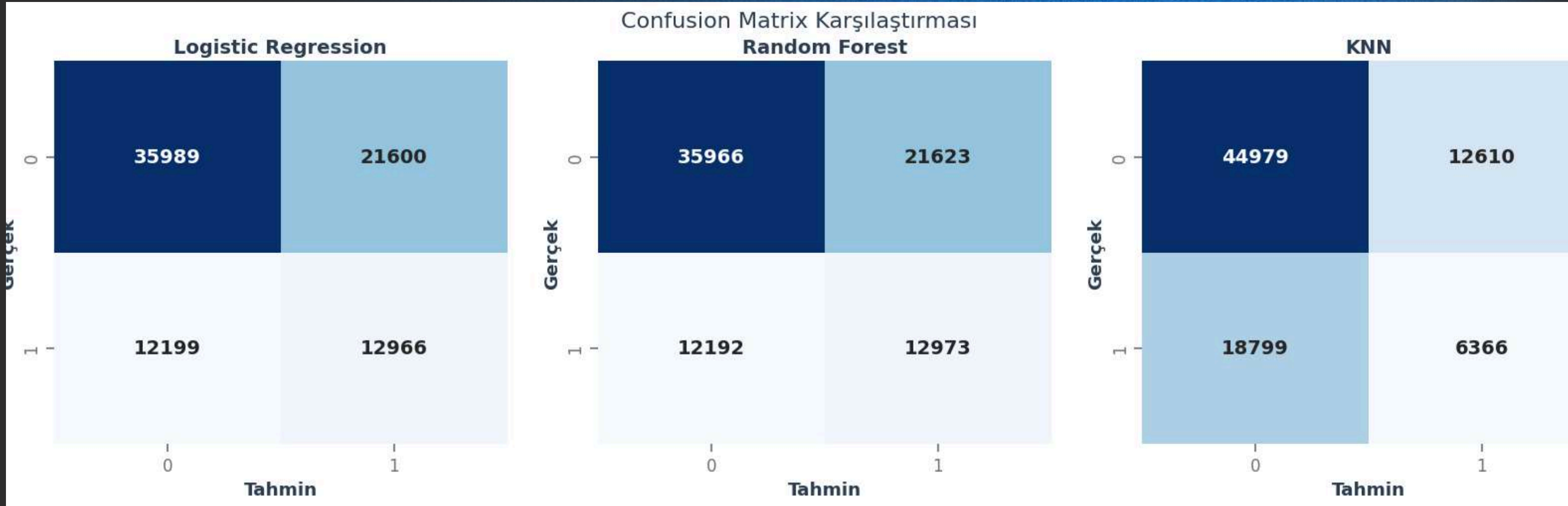
ROC Eğrisi Analizi

ROC Eğrisi Analizi grafiğine baktığımızda, modellerin genel başarısını gösteren AUC (Area Under Curve) skorları şöyledir:

- **Random Forest: 0.597 (En Yüksek)**
- Logistic Regression: 0.590
- KNN: 0.528

Yorum: AUC skoru ne kadar yüksekse, model hasta olanla olmayanı o kadar iyi ayırt eder. Random Forest, küçük bir farkla da olsa en yüksek skora ulaşarak liderliği almıştır.

CONFUSION MATRIX



Sağlık projelerinde en önemli soru şudur: "**Model kaç hastayı gözden kaçırdı?**" (False Negative). Confusion Matrix Karşılaştırması grafiğini inceleyelim:

- **KNN:** 18,799 hastayı kaçırmış (False Negative). Bu çok kötü bir skor, KNN elendi.
- **Logistic Regression:** 12,199 hastayı kaçırmış.
- **Random Forest:** 12,192 hastayı kaçırmış.

Yorum: Random Forest, Lojistik Regresyon'a göre 7 kişiyi daha fazla tespit etmiş (True Positive: 12,973 vs 12,966) ve daha az hata yapmıştır. Sağlıkta 1 hayat bile önemli olduğu için Random Forest tercih edilir.

MODEL KARŞILAŞTIRMALARI VE SONUÇLAR

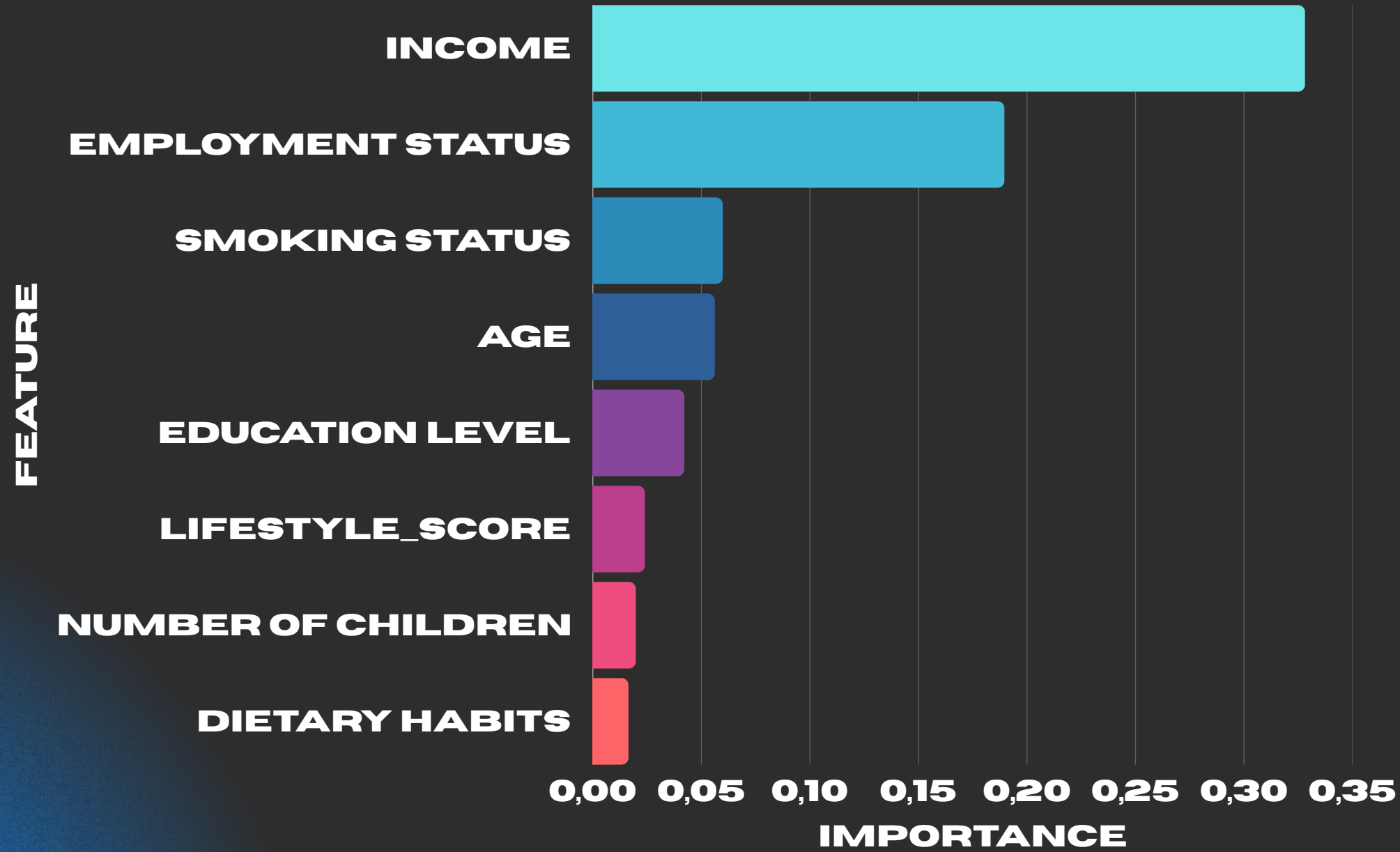
Aşağıdaki tabloda modellerin genel karşılaştırması sunulmuştur:

MODEL	GÜÇLÜ YÖN	ZAYIF YÖN	KULLANIM ALANI
LOGISTIC REGRESSION	Yorumlanabilir	Karmaşık ilişkileri yakalayamaz	Sağlık politikaları
KNN	Benzerlik temelli güçlü tahmin	Ölçeklendirmeye duyarlı	Davranışsal analiz
RANDOM FOREST	Yüksek doğruluk	Yorumlanabilirlik düşük	Risk tahmin sistemleri

- Tüm modeller geçmiş ruhsal rahatsızlık öyküsünü tahmin etmede anlamlı performans göstermiştir.

ÖZELLİK ÖNEMLERİ

EN ÖNEMLİ 10 RISK FAKTÖRÜ RANDOM FOREST



Modelimiz, bir kişinin depresyon veya ruh sağlığı geçmişini tahmin ederken, %67 oranında bu kişinin finansal güvencesine ve çalışma hayatına bakıyor. Yani model istatistiksel olarak diyor ki; 'Ekonomik belirsizlik veya işsizlik, ruh sağlığı üzerinde, sigara kullanımı veya yaştan çok daha büyük bir risk faktörüdür.'

ANALİZİN SONUÇLARI

TEMEL BULGULAR



En İyi Model Seçimi:

Yapılan kıyaslamalarda **Random Forest**, 0.597 AUC Skoru ve 12.973 doğru vaka tespiti (True Positive) ile en başarılı model olmuştur.



Koruyucu Faktörler:

Analizin en çarpıcı bulgusu; **Gelir (Income) ve Gelir Grubu** değişkenleridir. Gelir düzeyi arttıkça riskin 0.86 katına düştüğü (Odds Ratio) ve hedef değişkenle negatif korelasyon (-0.14) gösterdiği kanıtlanmıştır.



Kritik Risk Faktörleri:

İstihdam Durumu (Employment Status):

Mental hastalık geçmişi riskini 1.39 kat artıran en baskın faktör olarak tespit edilmiştir.

Beslenme ve Yaş:

Beslenme alışkanlıkları (1.04x) ve yaş grupları (1.02x) riski marjinal düzeyde artıran diğer değişkenlerdir.

ÖZET

BULGULAR

1

Ekonomik Belirleyicilik:

Mental hastalık geçmişi üzerinde biyolojik faktörlerden ziyade, sosyo-ekonomik faktörlerin (Gelir ve İş Durumu) çok daha baskın bir etkisi olduğu matematiksel olarak kanıtlanmıştır.

2

Model Stratejisi: Tek bir modele bağlı kalmak yerine hibrit bir yaklaşım benimsenmiştir. Risk faktörlerini açıklamak için

Lojistik Regresyon'un şeffaflığından, nihai tahminleme başarısı için ise **Random Forest'in** gücünden yararlanılmıştır.

3

Başarı Kriteri: Proje, 13.000'e yakın riskli vakayı doğru tespit ederek, klinik ön tarama (pre-screening) süreçlerinde kullanılabilecek potansiyel bir model ortaya koymuştur.

Klinik Karar Destek Sistemi:

Hastanelerde yoğunluk anında, hastaların risk skorunu (Random Forest ile) hesaplayarak doktora "Yüksek Riskli Vaka" uyarısı veren bir ön tarama (triyaj) aracı olarak kullanılabilir.

Kamu Sağlığı Politikaları:

Analizimiz "Düşük Gelir = Yüksek Risk" ilişkisini kanıtlamıştır. Sağlık Bakanlığı, psikolojik destek fonlarını işsizliğin yüksek olduğu bölgelere yönlendirmek için bu modeli kullanabilir.

Kurumsal Sağlık Yönetimi:

"İstihdam Durumu" en büyük risk faktörü olduğundan; İnsan Kaynakları departmanları, riskli çalışan gruplarına yönelik önleyici "Mental Wellness" programları geliştirebilir.

GERÇEK HAYAT UYGULAMALARI

Model Canlı Test Simülasyonu: Rastgele Örneklem Analizi

 Hasta ID	 Gerçek Durum	 Model Tahmini	 Sonuç
 #3859	 Risk Yok (Sağlıklı)	 Risk Yok (Sağlıklı)	 Doğru Tahmin
 #78918	 Risk Yok (Sağlıklı)	 Risk Yok (Sağlıklı)	 Doğru Tahmin
 #64685	 Risk Yok (Sağlıklı)	 Risk Yok (Sağlıklı)	 Doğru Tahmin
 #25167	 Risk Yok (Sağlıklı)	 Risk Yok (Sağlıklı)	 Doğru Tahmin
 #46084	 Risk Yok (Sağlıklı)	 Risk Yok (Sağlıklı)	 Doğru Tahmin

 **TEST ÖZETİ:** Rastgele seçilen 5 hastanın tamamı doğru sınıflandırılmıştır. (Başarı Oranı: 5/5 - %100)

TARTIŞMALAR VE ÖNERİLER

- Analizlerimiz gösterdi ki, insan psikolojisi haritadaki noktalar gibi (KNN mantığı) değil, karmaşık kurallar zinciri gibi çalışıyor. O yüzden 'Mesafe ölçen' değil, 'Soru soran' algoritmalar (yani Random Forest) bu projede kazandı.
1. Giyilebilir Teknoloji Entegrasyonu (IoT):
 - Analizimizde Uyku ve Fiziksel Aktivite verileri anket tabanlı olduğu için düşük etki göstermiştir.
 - Öneri: Modelin akıllı saatlere (Apple Watch, Fitbit) entegre edilmesi. Böylece uyku/hareket verileri gerçek zamanlı çekilerek hata payı sıfıra indirilebilir.
 2. Mobil Sağlık Asistanı (App):
 - Öneri: Random Forest modelinin bir mobil uygulamaya gömülmesi.
 - Kullanıcılar her gün "Bugün nasıl hissediyorsun?" sorusuna cevap verirken, model arka planda İş Durumu ve Gelir verilerini de hesaba katarak kişiye özel "Mental Hava Durumu" tahmini yapabilir.

KAYNAKLAR

- World Health Organization. (2023). Mental Health and Disorders.
- American Psychiatric Association. (2022). Mental Illness Overview.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning. Springer.
- Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32.
- slam, M. R., et al. (2025). Predicting Student Depression using Machine Learning: A Comparative Study of Logistic Regression and Random Forest. ResearchGate.
- Sridevi, S., et al. (2024). A Comparative Study of Mental Health Prediction using Machine Learning Algorithms. JETIR, 11(3).
- Hasan, M. A., & Nasution, N. (2025). Comparison of Logistic Regression, Random Forest, SVM and K-Nearest Neighbor (KNN) Algorithms in Diabetes Prediction. Journal of Applied Intelligent Computing.
- Colab
- Canva

**DİNLEDİĞİNİZ İÇİN
TEŞEKKÜRLER**