

# Kafein Tüketiminin Bireysel Farklılıklara Göre Bilişsel ve Sağlık Etkilerinin İncelenmesi ve Kişisel Öneri Sistemi

1<sup>st</sup> Ece Derya

Yapay Zeka Mühendisliği Bölümü  
TOBB Ekonomi ve Teknoloji Üniversitesi  
Ankara, Türkiye  
ederya@etu.edu.tr

**Abstract**—Bu çalışmada, kafein tüketiminin bireylerin bilişsel performansı ve uyku kalitesi üzerindeki etkileri farklı veri setleri üzerinden çok boyutlu olarak analiz edilmiştir. Kahve, çay ve enerji içeceği gibi farklı kafein kaynaklarının etkileri karşılaştırılmış; bireylerin odaklanma düzeyleri, uyku kalitesi ve tepki süreleri gibi bilişsel ölçütler üzerinden çeşitli ilişkiler incelenmiştir. Üç ayrı veri seti üzerinde gerçekleştirilen keşifsel veri analizi, istatistiksel hipotez testleri ve regresyon tabanlı makine öğrenmesi modelleri sonucunda, bireysel farklılıkların (yaş, cinsiyet, uyku düzeni, stres, egzersiz vb.) kafein etkileriyle olan ilişkileri değerlendirilmiştir. Eğitilen modellerin çıktıları, bireysel profil ve yaşam tarzı özelliklerine göre kişiselleştirilmiş kafein tüketim önerileri sunan bir sistem tasarımı kullanılmıştır.

**Index Terms**—kafein, bilişsel performans, sağlık, öneri sistemi

## I. GİRİŞ

Kafein, dünya genelinde en yaygın şekilde tüketilen psikoaktif bileşen olup kahve, çay, enerji içecekleri gibi birçok günlük üründe bulunmaktadır. Merkezi sinir sistemi üzerindeki uyarıcı etkileri nedeniyle dikkat, uyanıklık ve tepki süresi gibi bilişsel fonksiyonlar üzerinde kısa vadeli faydalar sağladığı bilinmektedir [1]. Ancak kafeinin etkileri bireyler arasında önemli ölçüde değişkenlik göstermektedir; yaş, cinsiyet, genetik yapı, uyku düzeni, stres seviyesi ve alışkanlık gibi birçok faktör bu farklılıkta rol oynamaktadır [2].

Bu çalışma, üç farklı veri seti kullanarak kafein tüketiminin bireysel faktörlere göre bilişsel performans ve sağlık göstergeleri üzerindeki etkilerini incelemeyi amaçlamaktadır. Projenin nihai hedefi, bu analizler doğrultusunda kişiselleştirilmiş bir kafein tüketim öneri sistemi geliştirmek olup, bireylerin yaşam tarzlarına ve ihtiyaçlarına en uygun kafein alım miktarını bilimsel verilere dayalı olarak belirlemelerine yardımcı olmaktadır.

### A. Motivasyon

Kafeinin dikkat, hafıza ve tepki süresi gibi bilişsel süreçler üzerindeki etkileri uzun süredir araştırılmakta olup bu konuda geniş bir literatür oluşmuştur. Bununla birlikte, aynı miktarda kafein tüketiminin farklı bireylerde farklı etkilere yol açtığı gözlemlenmektedir [3]. Özellikle son yıllarda kişiselleştirilmiş

sağlık uygulamalarına olan ilginin artması, kafein gibi yaygın maddelerin de bireysel düzeyde analiz edilmesini önemli hale getirmiştir.

Bu projede, açık veri setleri ve kontrollü deney verileri kullanılarak bireylerin yaş, uyku düzeni, stres seviyesi, cinsiyet ve yaşam tarzı gibi parametreler doğrultusunda kafein tüketiminin etkileri detaylı şekilde analiz edilmiştir. Çalışmanın sonunda, bireysel farklılıkları dikkate alarak kafein tüketimi için öneriler sunan bir sistem geliştirilmiştir. Böylece kullanıcıların daha sağlıklı ve bilinçli tüketim alışkanlıkları geliştirmeleri hedeflenmiştir.

### B. Literatür Taraması

Literatürde, kahve ve kafein tüketiminin sağlık ve bilişsel performans üzerindeki etkileri geniş kapsamda araştırılmıştır. Wang ve ark. (2025), ABD’de kahve tüketim zamanlamasının mortalite üzerine etkisini incelemiş ve sabah kahvesi tüketiminin ölüm riskini azalttığını göstermiştir [4]. Stevens ve ark. (2021), büyük epidemiyolojik veri setlerinde makine öğrenmesi teknikleriyle orta düzey kahve tüketiminin kalp yetmezliği riskini düşürdüğünü ortaya koymuştur [5]. McLellan ve ark. (2016) ise kafeinin bilişsel, fiziksel ve mesleki performans üzerindeki olumlu etkilerini derlemiş, ancak yüksek dozların yan etkilerine dikkat çekmiştir [6]. Ma ve ark. (2025) kahve tüketiminin kalp yapısı ve fonksiyonları üzerindeki çok boyutlu etkilerini incelemiş, bireysel metabolik farklılıkların önemini vurgulamıştır [7]. Bu çalışma, mevcut literatürü farklı veri setleriyle zenginleştirerek, kafeinin bireysel profillere bağlı etkilerini derinlemesine analiz etmektedir.

Uyku üzerine etkiler bağlamında Clark ve ark. (2017) sistematik incelemelerinde kafeinin uykuya giriş süresini uzattığını, total uyku süresini azalttığını ve uyku etkinliğini düşürdüğünü öne sürmüştür [8]. Daha yakın dönemde, Gardiner (2023–2025) deneysel çalışmalarla, akşam saatlerinde alınan yüksek doz kafeinin uyku yapısını bozduğunu göstermiştir [9].

Wu ve ark. (2024) 3 mg/kg kafein takviyesi ile Stroop ve görsel arama testlerinde tepki sürelerinde anlamlı iyileşme olduğunu bildirmiştir [10]. Rase-Abdullahi (2024) ise orta

düzy kahve tüketiminin dikkat, işlem hızı, karar verme ve bazı yürütücü işlevlerde fayda sağladığını vurgulamıştır [11].

Bu çalışma, literatürde sıkça ele alınan kafein ve bilişsel performans ilişkisini bireysel özellikler bağlamında değerlendirerek özgün bir katkı sunmaktadır. Farklı kaynaklardan elde edilen çok boyutlu veri setleri ile gerçekleştirilen analizlerde, hem istatistiksel yöntemler hem de makine öğrenimi tabanlı regresyon modelleri kullanılmış; elde edilen bulgular, bireylere özel kafein önerileri sunabilen bir sistem geliştirilerek pratik bir çıktıya dönüştürülmüştür. Geliştirilen bu kişiselleştirilmiş öneri sistemi, yaş, uyku süresi, stres düzeyi, egzersiz sıklığı, diyet tipi gibi bireysel değişkenleri dikkate alarak günlük alınması gereken optimal kafein miktarını ve önerilen içecek türünü belirlemekte; böylece kullanıcıların bilinçli ve sağlıklı kafein tüketimi gerçekleştirmelerine olanak sağlamaktadır.

### C. Genel Metodoloji

Projede uygulanan metodoloji, veri madenciliği sürecinin temel adımlarını kapsamaktadır. İlk olarak, veri temizliği adımında eksik ve aykırı değerler analiz edilerek uygun şekilde işlenmiş (eksik değerlere rastlanmamıştır), kategorik değişkenler etiketlenerek sayısal formata dönüştürülmüştür. Sayısal değişkenler için standartlaştırma ve normalize etme işlemleri gerçekleştirilmiştir. Keşifsel veri analizi (EDA) kapsamında dağılımlar, korelasyon matrisleri ve görselleştirmeler aracılığıyla veri yapısı detaylı şekilde incelenmiştir. Ardından anlamlı fark ve ilişkilerin belirlenmesi amacıyla t-testi, ANOVA, Pearson ve Spearman korelasyon testleri gibi istatistiksel yöntemler uygulanmıştır. Gruplar arası karşılaştırmalar, içecek türü, günün zamanı, cinsiyet gibi kategorilere göre gerçekleştirilmiş ve kafein tüketimini etkileyen faktörler ortaya konmuştur. Son aşamada ise çeşitli makine öğrenmesi algoritmaları (Random Forest, Gradient Boosting, SVR, MLP vb.) kullanılarak bireylere özel tahmin modelleri geliştirilmiş; bazı modellerde feature engineering, etkileşim terimleri ve stacking gibi yöntemlerle performans artırılmıştır.

### D. Amaçlar

- Kafein tüketiminin bilişsel işlevler (dikkat, hafıza, odaklanma) üzerindeki etkilerini incelemek.
- Yaş, cinsiyet, uyku düzeni, stres ve egzersiz gibi bireysel faktörlerin kafeinin etkisini nasıl değiştirdiğini değerlendirmek.
- Uyku kalitesi ile kafein tüketimi arasındaki ilişkiyi analiz ederek, kafeinin uykuya etkisini değerlendirmek.
- Farklı kafein kaynaklarının (kahve, enerji içeceği, çay) bilişsel işlevler ve uyku kalitesi üzerindeki etkilerini karşılaştırmalı olarak incelemek.
- Keşifsel veri analizi, korelasyon, regresyon ve varyans analizi gibi yöntemlerle kafeinin etkilerini ortaya koymak.
- Bireysel profillere göre kişiselleştirilmiş günlük kafein tüketimi öneri sistemi geliştirmek.
- Kafein tüketiminin bilişsel ve sağlık etkileri üzerine pratik ve uygulanabilir rehberlik sunmak.

## II. VERİ SETİ, VERİ ÖZELLİKLERİ, ÖZNİTELİKLER

### A. Veri Kaynağı

Bu çalışmada, bireylerin yaşam tarzı, kafein tüketimi ve bilişsel performans ilişkilerini analiz etmek amacıyla üç farklı veri seti kullanılmıştır. Veri setleri farklı kaynaklardan sağlanmış, yapılandırılmış ve çeşitli ön işleme adımlarıyla analizlere uygun hale getirilmiştir.

**1) Caffeine Intake Tracker Veri Seti:** Bu veri seti, bireylerin kafein tüketim miktarları ile birlikte yaşam alışkanlıkları ve algılanan etkiler üzerine bilgiler içeren, açık kaynaklı bir sağlık analiz setidir.<sup>1</sup> Veride toplam 13 sütun ve 500 kayıt bulunmaktadır. Veriler normalize edilmiş (0-1 arası) değerler içermektedir.

Caffeine Intake Tracker veri seti, bireylerin kafein tüketim miktarlarının yanı sıra, odaklanma düzeyi, uyku kalitesi, içecek tercihleri, tüketim zamanı ve cinsiyet gibi çeşitli yaşam tarzı ve davranışsal faktörleri içermektedir. *caffeine\_mg*, *age*, *focus\_level* ve *sleep\_quality* gibi sürekli değişkenler, bireylerin beyanlarına dayalı olarak 0-1 aralığında ölçeklendirilmiştir. *sleep\_impacted* değişkeni, kafein tüketiminin uyku üzerindeki etkisini ikili biçimde (0: etkilenmedi, 1: etkilendi) temsil etmektedir.

İçecek tercihini belirten sütunlar (*beverage\_coffee*, *beverage\_tea*, *beverage\_energy\_drink*) ve tüketim zamanı ile ilgili sütunlar (*time\_of\_day\_morning*, *time\_of\_day\_afternoon*, *time\_of\_day\_evening*) boolean formatında düzenlenmiştir. Cinsiyet bilgisi, *gender\_male* ve *gender\_female* sütunları ile yine ikili biçimde kodlanmıştır. Bu yapılandırma, hem istatistiksel analizlerde hem de makine öğrenimi modellerinde esnek bir kullanım sağlamaktadır.

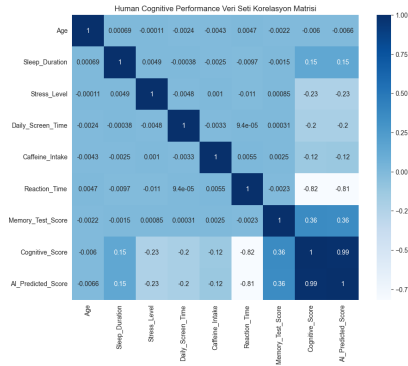
Bu veri seti; özellikle içecek tercihi, zamanlama, yaş ve cinsiyet gibi bireysel faktörlerin kafein tüketimi ile ilişkisini incelemek için kullanılmıştır. Veride yer alan tüm sütunlar yapısal ve analiz edilebilir formdadır.

**2) Human Cognitive Performance Veri Seti:** Bu veri seti, yaşam tarzı faktörlerinin insan bilişsel performansı üzerindeki etkisini incelemeye yönelik hazırlanmıştır.<sup>2</sup> Toplam 80.000 satır ve 13 sütun içermektedir. Veride hem kullanıcı beyanları hem de bilişsel test sonuçları yer almaktadır.

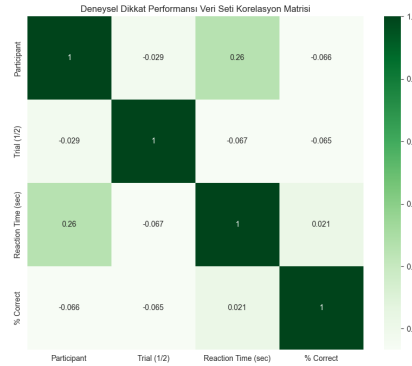
İncelenen veri setinde öne çıkan değişkenler arasında bireylerin yaşam tarzı özelliklerini yansıtan *Sleep\_Duration*, *Stress\_Level*, *Diet\_Type*, *Daily\_Screen\_Time* ve *Exercise\_Frequency* gibi faktörler; bilişsel performanslarını ölçen *Reaction\_Time*, *Memory\_Test\_Score* ve *Cognitive\_Score* gibi ölçümler; ayrıca günlük kafein tüketim miktarını ifade eden *Caffeine\_Intake* yer almaktadır. Bunun yanında, yapay zeka tarafından hesaplanan tahmini performans skoru olan *AI\_Predicted\_Score*, analizlerde yalnızca referans amaçlı tutulmuş; eğitim, test veya modelleme süreçlerine dahil edilmemiştir.

<sup>1</sup><https://www.kaggle.com/datasets/prekshad2166/caffeine-intake-tracker-csv>

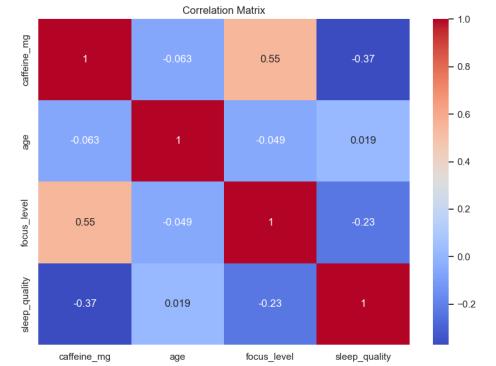
<sup>2</sup><https://www.kaggle.com/datasets/samxam/human-cognitive-performance-analysis>



(a) Human Cognitive Performance Veri Seti Korelasyon Matrisi



(b) Caffeine-Fatigue Veri Seti Korelasyon Matrisi



(c) Caffeine Intake Tracker Veri Seti Korelasyon Matrisi

Fig. 1: Korelasyon Matrisleri

Bu veri seti, hem klasik istatistiksel testler hem de regresyon modelleri için kullanılmıştır. Normalize edilmemiş orijinal değerler içerdiğinden sayısal sütunlar z-score normalizasyon ile dönüştürülmüştür. Kategorik sütunlar one-hot-encoding kullanılarak ile ikili sütunlara ayrılmıştır.

3) **Exercise-Induced Fatigue and Caffeine Supplementation Effect Veri Seti:** Bu veri seti, kontrollü bir deneysel çalışmadan elde edilen verileri içermektedir.<sup>3</sup> Toplam 576 gözlem, iki farklı deney grubu (S1 ve S2), müdahale (kafein/placebo), fiziksel egzersiz ve zaman (pre/post) gibi deneysel faktörlere sahiptir. Gözlemcilerin yarısı placebo, yarısı da kafein grubuna aittir.

Bu veri setinde, katılımcılara uygulanan kafein (*Intervention*, 1) veya plasebo (0) müdahalesi, ölçüm zamanı (öncesi/sonrası), deneysel koşul tipi (endogenous veya exogenous), uyarıcı türü (valid, invalid, neutral) gibi deneysel değişkenler yer almaktadır. Ayrıca, tepki süresi (saniye cinsinden) ve doğruluk yüzdesi gibi performans ölçümleri de veri setinde bulunmaktadır.

Feature Engineering kapsamında aşağıdaki etkileşim değişkenleri türetilmiştir:

- *Intervention \* Time*, *Time \* Condition*, *All\_combo*
- *Trial\_squared*: Deneme sayısının karesi.
- *Participant\_Avg\_RT*: Katılımcı bazlı ortalama tepki süresi.

Veri seti üzerinden regresyon modelleri eğitilmiş ve ön işleme sürecinde sayısal değişkenler z-score normalizasyonu ile standartlaştırılırken, kategorik değişkenler one-hot-encoding yöntemiyle sayısal formata dönüştürülmüştür. Bu dönüşümler, her bir değişken tipine uygun olarak özelleştirilmiş veri işleme adımları içeren bir sütun bazlı yapı üzerinden uygulanmıştır.

Bu veri seti, kafein müdahalesinin egzersiz sonrası bilişsel performansa etkisini modellemek için kullanılmıştır. Bulgular, kafeinin özellikle tepki süresinde anlamlı iyileşme sağladığını ortaya koymuştur.

## B. Özellikler ve Ölçüm Seviyeleri

Kullanılan veri setleri farklı yapı ve içeriklere sahip olup her biri sayısal ve kategorik değişkenlerden oluşmaktadır. Tüm verisetleri için değişkenlerin ölçüm seviyeleri aşağıdaki tablolarda özetlenmiştir:

TABLE I: Caffeine Intake Tracker – Ölçüm Seviyeleri

Değişken	Tür	Ölçüm Seviyesi
caffeine_mg	float64	Ratio
age	float64	Ratio
focus_level	float64	Ratio
sleep_quality	float64	Ratio
sleep_impacted	int64	Nominal
beverage_coffee	bool	Nominal
beverage_energy_drink	bool	Nominal
beverage_tea	bool	Nominal
time_of_day_afternoon	bool	Nominal
time_of_day_evening	bool	Nominal
time_of_day_morning	bool	Nominal
gender_female	bool	Nominal
gender_male	bool	Nominal

TABLE II: Human Cognitive Performance – Ölçüm Seviyeleri

Değişken	Tür	Ölçüm Seviyesi
User_ID	object	Nominal
Age	int64	Ratio
Gender	object	Nominal
Sleep_Duration	float64	Ratio
Stress_Level	int64	Ordinal
Diet_Type	object	Nominal
Daily_Screen_Time	float64	Ratio
Exercise_Frequency	object	Nominal
Caffeine_Intake	int64	Ratio
Reaction_Time	float64	Ratio
Memory_Test_Score	int64	Ratio
Cognitive_Score	float64	Ratio
AI_Predicted_Score	float64	Ratio

<sup>3</sup>[https://figshare.com/collections/\\_/3521103](https://figshare.com/collections/_/3521103)

TABLE III: Caffeine-Fatigue Veri Seti – Ölçüm Seviyeleri

Değişken	Tür	Ölçüm Seviyesi
Participant	int64	Ordinal
Intervention	object	Nominal
Trial (1/2)	int64	Ordinal
Time Point (pre/post)	object	Nominal
Condition (Endo/Exo)	object	Nominal
Cue Type	object	Nominal
Reaction Time (sec)	float64	Ratio
% Correct	float64	Ratio

Şekil 1’de üç farklı veri seti için hesaplanan korelasyon matrisleri sunulmuştur. Bu matrisler değişkenler arasındaki korelasyon ilişkilerini göstermektedir. Human Cognitive Performance veri setinde (1a) özellikle *Cognitive\_Score* ile *Memory\_Test\_Score* arasında pozitif anlamlı korelasyon gözlemlenmiştir. Caffeine Intake Tracker veri setinde (1c), *Caffeine\_mg* ile *Focus\_Level* arasında güçlü pozitif korelasyon, *Sleep\_Quality* ile ise negatif yönde anlamlı korelasyon tespit edilmiştir. Bu matrisler, değişkenler arasındaki ilişkileri önceden belirlemek ve modelleme sürecine yön vermek açısından kritik öneme sahiptir.

### C. Veri Ön İşleme Süreci

Veri temizleme, dönüştürme ve hazırlama süreçleri veri setine özel olarak yürütülmüştür:

- **Eksik veri temizliği:** Yapılan kontroller sonucu üç verisetinde de eksik verilere rastlanmamıştır.
- **Kategorik verilerin dönüştürülmesi:** Tüm veri setlerinde kategorik sütunlar one-hot-encoding kullanılarak ikili sütunlara ayrılmıştır.
- **Sayısal verilerin ölçeklenmesi:** Z-score normalizasyonu ile tüm sayısal değişkenler ortalama 0 ve standart sapma 1 olacak şekilde normalize edilmiştir.

### D. Öznitelik (Feature) Analizi ve Seçimi

- **Korelasyon Tabanlı Seçim:** Caffeine ile yüksek korelasyon gösteren değişkenler tespit edilerek “bilgilendirici örnekler” (top 1000 sample) üzerinden ek analizler yapılmıştır.
- **Feature Engineering:** Deneysel veri setinde çeşitli etkileşim terimleri türetilmiştir: *Intervention \* Time*, *Time \* Condition*, *All\_combo*, *Participant\_Avg\_RT* gibi değişkenler oluşturulmuştur.
- **Polynomial Features:** GB modeli ile birlikte **PolynomialFeatures (degree=2)** kullanılmıştır.

### E. Sınıf Dağılımı ve Normalizasyon

Regresyon problemine odaklanıldığından sınıf dengesizliği sorunu bulunmamaktadır. Ancak hedef değişkenin dağılımı analiz edilerek bazı durumlarda aykırı değerlerin etkisini azaltmak için örnekler filtrelenmiştir.

### F. Görselleştirme Teknikleri

Korelasyon analizleri için ısı haritası (heatmap) ve sayısal değişkenler arası ilişkileri incelemek amacıyla dağılım grafikleri (scatter plot ve regplot) kullanılmıştır. Kategorik değişkenlerle sayısal değişkenler arasındaki ilişkiler *box plot*

ve *violin plot* ile görselleştirilmiştir. Grup karşılaştırmalarının yapılabilmesi için bazı sürekli değişkenler, örneğin yüzde doğru cevap (% Correct), *qcut* yöntemiyle kategorilere ayrılmış ve ANOVA ile Kruskal-Wallis testlerine uygun hale getirilmiştir. Model performansını değerlendirmek üzere *confusion matrix* görselleştirmeleri de kullanılmıştır.

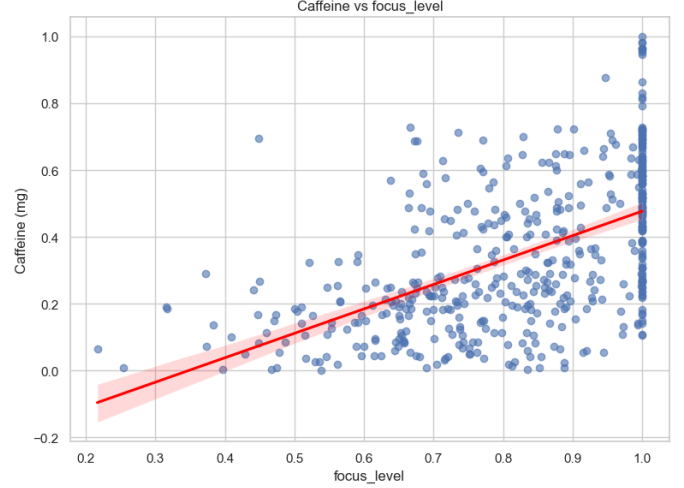


Fig. 2: Kafein tüketimi ile odaklanma seviyesi arasındaki ilişki.

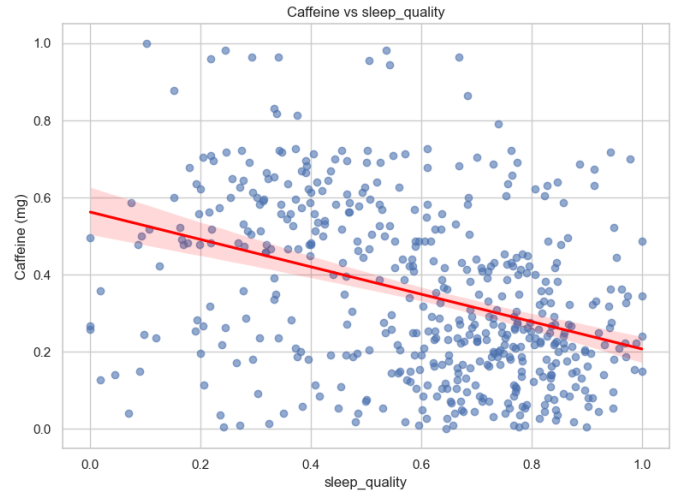


Fig. 3: Kafein tüketimi ile uyku kalitesi arasındaki ilişki.

Kafein tüketiminin bireylerin bilişsel performansı ve uyku kalitesi üzerindeki etkilerini görselleştirmek amacıyla iki temel korelasyon grafiği hazırlanmıştır. Şekil 2’te görüldüğü üzere, odaklanma seviyesi ile kafein tüketimi arasında pozitif yönlü bir ilişki gözlemlenmiştir; bu durum, bireylerin daha fazla kafein tükettiklerinde kendilerini daha odaklanmış hissettiklerini göstermektedir. Öte yandan, Şekil 3’te kafein ile uyku kalitesi arasında negatif bir korelasyon bulunmuş, bu da kafein tüketiminin uyku kalitesini olumsuz yönde etkileyebileceğini ortaya koymuştur.

Focus düzeyine ilişkin sayısal değişkenin daha anlamlı görsel analizini gerçekleştirebilmek için, veri maden-

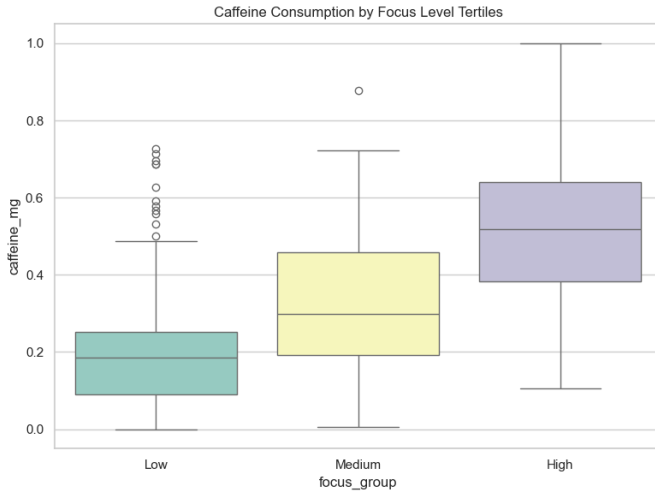


Fig. 4: Odaklanma düzeyine göre gruplandırılmış bireylerde kafein tüketimi.

ciliğinde *discretization* olarak bilinen yöntem uygulanmış ve katılımcılar odaklanma skorlarına göre üç gruba (Low, Medium, High) ayrılmıştır. Şekil 4'te bu gruplara ait kafein tüketimlerinin dağılımı gözlemlenebilir. Özellikle “High” grubunda tüketimin daha yoğun olduğu, “Low” grubunda ise düşük kaldığı görülmektedir.

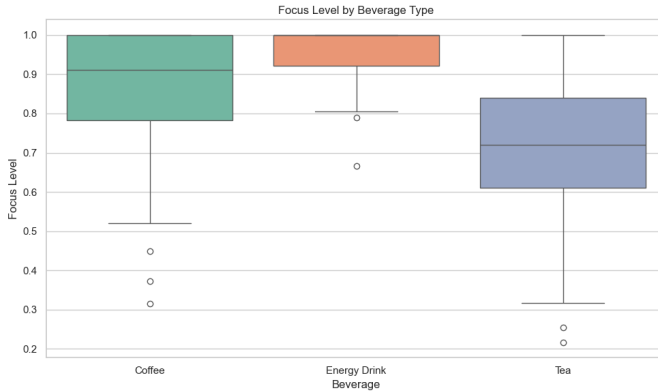


Fig. 5: İçecek türlerine göre odaklanma seviyesi dağılımı.

Farklı içecek türlerinin bilişsel ve fizyolojik etkilerini incelemek amacıyla yapılan analizlerde, Şekil 5'te görüldüğü üzere enerji içeceği ve kahve tüketen bireylerin odaklanma seviyeleri, çay tüketenlere kıyasla daha yüksek dağılım göstermektedir. Bu durum, kafein içeriği daha yüksek olan içeceklerin kısa vadede uyarıcı etki oluşturduğunu desteklemektedir. Öte yandan, Şekil 6 çay tüketicilerinin ortalama uyku kalitesinin, enerji içeceği ve kahve tüketen bireylere göre daha yüksek olduğunu ortaya koymaktadır. Bu bulgular, çayın nispeten daha düşük kafein içeriği nedeniyle uyku düzeni üzerinde daha az bozucu etkisi olabileceğini göstermektedir.

Şekil 7 incelendiğinde, kadın ve erkek bireyler arasında günlük kafein tüketim düzeylerinde belirgin bir fark

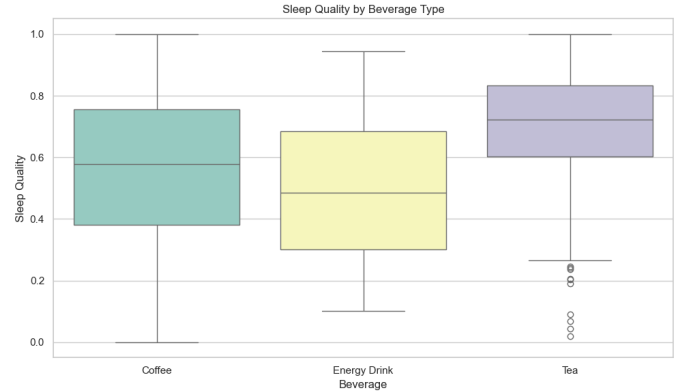


Fig. 6: İçecek türlerine göre uyku kalitesi dağılımı.

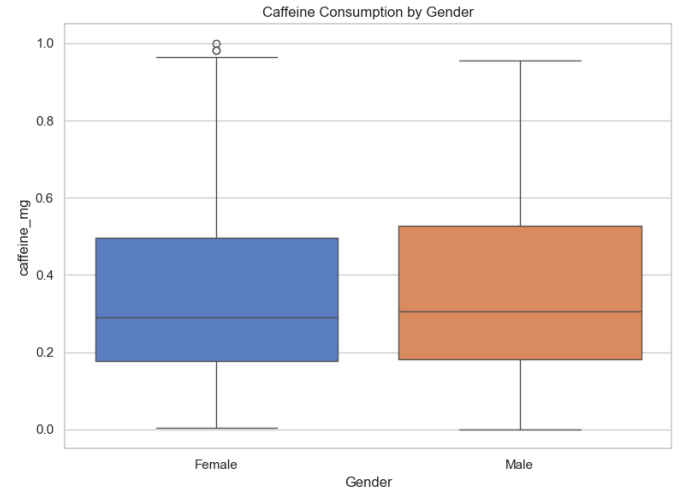


Fig. 7: Cinsiyete göre kafein tüketiminin dağılımı.

gözlemlenmemektedir. Yapılan varyans analizine göre bu farklılık istatistiksel olarak da anlamlı bulunmamıştır ( $p = 0.7097$ ). Bu sonuç, cinsiyetin bireylerin kafein tüketim miktarlarını belirlemede anlamlı bir rol oynamadığını göstermektedir.

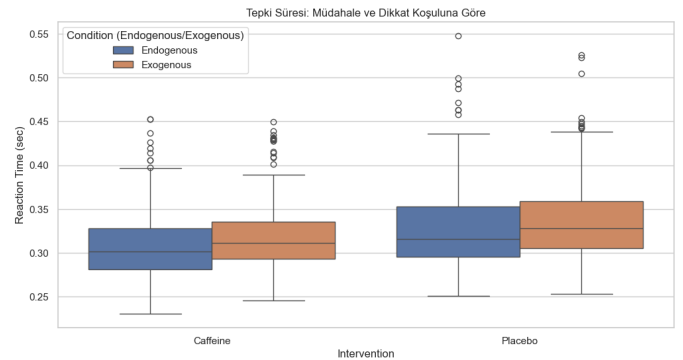


Fig. 8: Tepki süresinin müdahale türüne (kafein/plasebo) ve dikkat koşuluna (endogenous/exogenous) göre dağılımı.

Şekil 8'de, tepki süresinin müdahale türüne (kafein veya plasebo) ve dikkat koşuluna (içsel/exogenous ve

dışsal/exogenous) göre dağılımı gösterilmektedir. Endogenous (içsel) dikkat, bireyin kendi isteğiyle odaklandığı durumları; exogenous (dışsal) dikkat ise çevresel uyarıcılara otomatik tepkileri ifade etmektedir. Kafein uygulanan grupta hem içsel (endogenous) hem de dışsal (exogenous) dikkat koşullarında tepki süresi daha düşük seviyelerde gözlemlenmiştir. Buna karşılık, plasebo grubunda ortalama tepki süreleri daha yüksektir ve varyans daha geniştir.

Özellikle kafein grubunda dikkat koşulundan bağımsız olarak daha düşük tepki süresi değerleri öne çıkmakta, bu da kafeinin dikkatsel işlem hızını artırıcı potansiyel etkisine işaret etmektedir.

### III. METODOLOJİ

Bu çalışmada, bireysel özellikler doğrultusunda kafein tüketiminin bilişsel ve sağlık üzerindeki etkilerini incelemek amacıyla dört temel analiz yaklaşımı benimsenmiştir: (i) keşifsel veri analizi (EDA) ile veri dağılımlarının, eksik/aykırı değerlerin ve değişkenler arası ilişkilerin görselleştirilmesi, (ii) istatistiksel hipotez testleri ile anlamlı ilişki ve farkların araştırılması, (iii) makine öğrenimi modelleri ile regresyon tahminlerinin gerçekleştirilmesi, (iv) elde edilen modellerin çıktıları doğrultusunda kişiselleştirilmiş bir kafein öneri sisteminin geliştirilmesi.

#### A. Keşifsel Veri Analizi (EDA):

Veri setlerindeki dağılımlar, eksik ve aykırı değerler tespit edilmiş (eksik değere rastlanmamıştır), korelasyon analizleri ile değişkenler arası ilişkiler ortaya konmuştur. Bu adım, veri kalitesini ve yapısını anlamak için temel bir aşamadır.

#### B. İstatistiksel Yöntemler

Veri analiz sürecinin ilk aşamasında, değişkenler arasındaki ilişkileri ve gruplar arası farkları değerlendirmek amacıyla çok sayıda istatistiksel test uygulanmıştır. Kullanılan testler ve uygulama amaçları aşağıda özetlenmiştir:

- **Kruskal-Wallis Testi:** Üç ve üzeri grup arasında sayısal değişkenlerin farkını değerlendirmek için kullanılmıştır. Örneğin, *Focus Level* değişkeni kahve saatine göre karşılaştırılmıştır.
- **Ki-Kare ( $\chi^2$ ) Testi:** İki kategorik değişkenin ilişkisini test etmek amacıyla kullanılmıştır. Örneğin, *Sleep Impacted* ile *Beverage Type* arasındaki ilişki test edilmiştir.
- **Pearson Korelasyon Testi:** Sayısal değişkenler arası doğrusal ilişkiler araştırılmıştır. Örneğin, *Caffeine Intake* ile *Focus Level* arasında güçlü pozitif korelasyon elde edilmiştir.
- **ANOVA:** Sayısal bir değişkenin kategorik gruplara göre ortalama farklarını test etmek için uygulanmıştır (örneğin, *Caffeine Intake* ~ Gender).
- **Mann-Whitney U Testi:** Normal dağılım varsayımını sağlamayan durumlarda iki bağımsız grup arası farklar değerlendirilmiştir (örneğin, *Caffeine/Placebo* ~ Reaction Time).

- **Eşleştirilmiş t-testi:** Müdahale öncesi ve sonrası ölçümler (*pre* vs *post*) arasındaki farkları karşılaştırmak için uygulanmıştır.

Bu testler sayesinde, hangi bireysel faktörlerin kafein etkisini modüle ettiği istatistiksel olarak anlamlı biçimde belirlenmiştir.

#### C. Makine Öğrenimi Yöntemleri

İkinci aşamada, üç farklı veri setinde regresyon problemleri çözülmüş ve kafein tüketimi veya reaksiyon süresi gibi sayısal hedefler tahmin edilmiştir:

- 1) **Coffee Intake Tracker:** Bireylerin içecek alışkanlıkları ve sağlık verileri ile kafein miktarının tahmini,
- 2) **Human Cognitive Performance:** Yaşam tarzı, sağlık, stres ve uyku bilgileri ile kafein alımının tahmini,
- 3) **Caffeine-Fatigue:** Müdahale (kafein/plasebo), zaman ve koşullar üzerinden tepki süresinin tahmini.

Tüm veri setlerinde ortak olarak veri temizliği ve ön işleme adımları uygulanmıştır. İlk olarak hedef değişken dışındaki ilgisiz sütunlar çıkarılmış, eksik veya hatalı değerler veri setinden ayıklanmıştır. Sayısal değişkenler, ortalaması sıfır ve standart sapması bir olacak şekilde *z-puanı normalizasyonu* ile ölçeklendirilmiş; kategorik değişkenler ise her kategori için ayrı sütunlar oluşturularak ikili (0-1) gösterime dönüştürülmüştür (one-hot encoding yöntemiyle). Ardından veri setleri %80 eğitim ve %20 test olacak şekilde ayrılmıştır. Bu hazırlık süreci, hem istatistiksel testler hem de makine öğrenimi modelleri için tutarlı ve sağlıklı bir analiz altyapısı oluşturmuştur.

1) **Kullanılan Modeller:** Aşağıdaki regresyon modelleri hem temel hem de ensemble yaklaşımlar üzerinden değerlendirilmiştir:

- **Linear Modeller:** Linear Regression, Ridge, Lasso, ElasticNet
- **Ağaç Tabanlı Modeller:** Decision Tree, Random Forest, Extra Trees, Gradient Boosting, HistGradientBoosting
- **Robust Linear Modeller:** HuberRegressor, Theil-Sen, PassiveAggressive, SGDRegressor
- **Ensemble Yöntemleri:** Bagging, AdaBoost, Stacking
- **Diğer:** KNeighborsRegressor, SVR, MLPRegressor

2) **Model Kombinasyonları ve İyileştirmeler:** Bazı modellerin performansını artırmak için aşağıdaki özel kombinasyonlar denenmiştir:

- **Gradient Boosting + Polynomial Features + Feature Engineering (GB + Poly + FE):** Deneysel veri setinde 2. dereceden polinomlar ve etkileşim terimleri oluşturularak  $R^2$  skoru iyileştirilmiştir.
- **Stacking:** Ridge + SVR tahminleri, Gradient Boosting ile birleştirilerek son çıktı alınmıştır.
- **MLP Regressor:** Human Cognitive Performance veri setinde yüksek doğruluk elde etmiştir.

Her model MAE, RMSE ve  $R^2$  metrikleriyle değerlendirilmiş; sıralamalar doğruluk skorlarına göre yapılmıştır.



#### D. Kişiselleştirilmiş Kafein Öneri Sistemi

Bu proje kapsamında geliştirilen kafein öneri sistemi, bireylerin yaş, uyku süresi, günlük ekran süresi, stres seviyesi, reaksiyon süresi, hafıza durumu, bilişsel performansı, cinsiyet, diyet tipi ve egzersiz sıklığı gibi kişisel özelliklerine dayalı olarak günlük alabilecekleri optimum kafein miktarını tahmin etmektedir. Bu amaçla, yapılan model karşılaştırmaları sonucunda en başarılı performansı gösteren **MLP Regressor** modeli tercih edilmiştir. Sistem, bu regresyon modelini kullanarak bireysel girdilerden elde edilen tahmini değeri hesaplamakta ve ardından bu miktarı kahve, çay ve enerji içeceği cinsinden pratik önerilere dönüştürmektedir.

İçecek türüne özgü dönüşümler yapılırken, her içeceğin ortalama kafein içeriğine göre hesaplama yapılmış, bu kapsamda *Caffeine Content of Drinks*<sup>4</sup> veri setinden faydalanılmıştır. Böylece sistem, "yaklaşık 2 fincan kahve" gibi daha anlaşılır öneriler de sunmaktadır.

Tahmin edilen değer doğrudan kullanıcıya görsel arayüz üzerinden iletilmekte ve uygun içecek alternatifleri ile birlikte sade ve anlaşılır bir biçimde sunulmaktadır. Bu sistem, kullanıcıların günlük yaşamda daha bilinçli ve kontrollü kafein tüketimi yapmalarına katkı sağlamayı hedeflemektedir.

Öte yandan, bazı bireyler için kafein tüketimi uygun olmayabilir; sistem, belirli sağlık riskleri veya özel durumları göz önünde bulundurarak bu bireylerde kafein tüketimini önermemektedir ve kullanıcıya bu durum açıkça bildirilmektedir. Ayrıca, modelin tahmin ettiği değerler, güvenli tüketim sınırlarını aşmayacak şekilde üst limitlerle sınırlandırılmıştır. Bu sayede, kafein tüketimi önerilen bireylerde dahi aşırı alımın önüne geçilmesi sağlanmıştır.

Fig. 9: Kafein öneri sistemine ait kullanıcı girdi ekranı. Bireyler yaş, uyku süresi, stres seviyesi gibi kişisel özelliklerini girerek kafein tahmini alabilmektedir.

<sup>4</sup><https://www.kaggle.com/datasets/heitorunes/caffeine-content-of-drinks/data>

Fig. 10: Kafein öneri sisteminin çıktı ekranı. Tahmini günlük kafein miktarı ve karşılık gelen içecek önerileri kullanıcıya sunulmaktadır.

Şekil 9 ve Şekil 10, geliştirilen kişiselleştirilmiş kafein öneri sisteminin arayüzünü göstermektedir. Kullanıcıdan yaş, uyku süresi, ekran süresi, stres seviyesi, cinsiyet, diyet tipi ve bilişsel durum gibi bireysel faktörleri girmesi istenmekte; bu girdiler daha önce eğitilen regresyon modellerine aktarılmaktadır. Sistem, tahmini günlük kafein ihtiyacını hesaplayarak kullanıcıya hem mg cinsinden miktarı hem de içecek türlerine göre sadeleştirilmiş bir öneri sunmaktadır.

#### IV. BULGULAR

##### A. İstatistiksel Analiz Bulguları

Kafein tüketiminin bilişsel performans, uyku kalitesi ve odaklanma üzerindeki etkilerini değerlendirmek amacıyla çeşitli istatistiksel testler uygulanmıştır. Korelasyon analizleri sonucunda, kafein tüketimi ile odaklanma seviyesi arasında pozitif ve anlamlı bir ilişki saptanmıştır (Pearson  $r = 0.547$ ,  $p < 0.0001$ ; Spearman  $r = 0.585$ ,  $p < 0.0001$ ). Uyku kalitesi ile ise negatif yönde anlamlı ilişkiler elde edilmiştir (Pearson  $r = -0.369$ ,  $p < 0.0001$ ; Spearman  $r = -0.364$ ,  $p < 0.0001$ ). Buna karşılık, yaş ile kafein tüketimi arasında zayıf ve anlamsız bir ilişki bulunmuştur (Pearson  $r = -0.063$ ,  $p = 0.1585$ ).

Farklı içecek türlerinin etkisini değerlendirmek için Mann-Whitney U testi uygulanmış ve kahve, çay ve enerji içeceği içen bireylerle içmeyenler arasında hem odaklanma seviyesi hem de uyku kalitesi açısından anlamlı farklar tespit edilmiştir ( $p < 0.001$ ). Özellikle enerji içeceği tüketen bireylerde uyku kalitesinin anlamlı biçimde daha düşük olduğu belirlenmiştir ( $p = 0.0018$ ).

Kahve tüketim saatine göre odaklanma seviyesi (Kruskal-Wallis  $H = 1.778$ ,  $p = 0.4111$ ) ve uyku kalitesi ( $H = 4.930$ ,  $p = 0.0850$ ) açısından anlamlı bir fark gözlenmemiştir. Uyku etkilenmesi değişkeni ile içecek türleri arasında anlamlı ilişkiler bulunmuştur (Chi-Square: Kahve  $p < 0.0001$ , Çay  $p < 0.0001$ , Enerji İçeceği  $p = 0.0012$ ), ancak tüketim saatine göre anlamlı bir fark bulunmamıştır (Sabah  $p = 0.9267$ , Öğleden sonra  $p = 0.0582$ , Akşam  $p = 0.0520$ ).

Cinsiyet ( $p = 0.7097$ ) ve diyet tipi ( $p = 0.2953$ ) gibi değişkenlere göre kafein alımında anlamlı farklar gözlenmemiştir (ANOVA).

Deneysel veride ise, kafein ve plasebo grupları arasında tepki süresi bakımından anlamlı farklar saptanmıştır (Mann-Whitney  $U = 32158.0$ ,  $p < 0.0001$ ). Ayrıca, caffeine level grupları arasında hem odaklanma seviyesi ( $H = 135.437$ ,  $p < 0.0001$ ) hem de uyku kalitesi ( $H = 43.910$ ,  $p < 0.0001$ ) açısından anlamlı farklılıklar bulunmuştur.

Bu bulgular, bireysel değişkenlerin kafein etkilerinin biçimlenmesinde nasıl rol oynadığını ortaya koymakta ve kişiselleştirilmiş öneri sistemi için güçlü bir temel sağlamaktadır.

### B. Makine Öğrenimi Bulguları

Farklı veri setleri üzerinde regresyon modelleri eğitilerek, kafein tüketimi ve reaksiyon süresi gibi hedef değişkenlerin tahmini gerçekleştirilmiştir. Modellerin başarısı Mean Squared Error (MAE), Root Mean Square Error (RMSE) ve Doğruluk Katsayısı ( $R^2$ ) gibi metriklerle değerlendirilmiştir.

TABLE IV: Reaction Time Tahmini (Caffeine/Placebo Verisi)

Model	MAE	RMSE	$R^2$
Random Forest	0.0232	0.0305	0.683
GB + Poly + FE	0.0237	0.0310	0.672
Bagging	0.0245	0.0316	0.660
HistGradientBoosting	0.0251	0.0317	0.657
Gradient Boosting	0.0249	0.0337	0.612
AdaBoost	0.0309	0.0364	0.549
Extra Trees	0.0285	0.0369	0.536
Stacking	0.0308	0.0392	0.478
Ridge Regression	0.0331	0.0414	0.416
Linear Regression	0.0331	0.0414	0.415
Huber Regressor	0.0328	0.0411	0.401
KNN	0.0352	0.0441	0.338

TABLE V: Caffeine Consumption Tahmini (Human Cognitive Performance Verisi)

Model	MAE	RMSE	$R^2$
MLP Regressor	6.17	27.54	0.963
Linear Regression	36.53	69.28	0.768
Ridge Regression	36.66	69.27	0.768
Lasso Regression	52.77	73.85	0.737
Huber Regressor	11.96	81.90	0.676
Random Forest	95.15	112.57	0.388
Extra Trees	96.02	113.52	0.378
Bagging	99.10	118.98	0.316
HistGradientBoosting	103.42	120.54	0.298
SVR	109.76	127.22	0.219
Gradient Boosting	116.42	135.00	0.120
ElasticNet	122.82	141.92	0.028

TABLE VI: Caffeine Consumption Tahmini (Coffee Intake Tracker Verisi)

Model	MAE	RMSE	$R^2$
Linear Regression	0.106	0.134	0.649
Ridge Regression	0.106	0.134	0.649
KNN	0.116	0.147	0.582
Gradient Boosting	0.114	0.148	0.575
Random Forest	0.114	0.149	0.565
SVR	0.121	0.151	0.556
ElasticNet	0.124	0.152	0.547
Lasso Regression	0.154	0.184	0.341
Decision Tree	0.136	0.184	0.338
MLP Regressor	0.152	0.190	0.294

En yüksek doğruluk, Human Cognitive Performance veri seti üzerinde eğitilen MLP Regressor modeliyle elde edilmiştir ( $R^2 = 0.963$ ). Coffee Intake Tracker veri setinde ise Linear ve Ridge Regression modelleri benzer doğrulukla ( $R^2 = 0.649$ ) öne çıkmıştır. Reaction Time tahmininde Random Forest modeli en başarılı sonuçları vermiştir ( $MAE = 0.0232$ ,  $R^2 = 0.683$ ). Bu bulgular, farklı veri setlerinin ve hedef değişkenlerin model performansı üzerindeki etkilerini ortaya koymaktadır.

### V. SONUÇLAR VE TARTIŞMA

Bu projede, kafein tüketiminin bireysel özelliklere bağlı olarak bilişsel performans ve sağlık üzerindeki etkileri incelenmiştir. Keşifsel veri analizi, istatistiksel testler ve makine öğrenimi yöntemleri kullanılarak üç farklı veri seti üzerinde detaylı analizler gerçekleştirilmiştir.

İstatistiksel testler sonucunda kafein tüketimi ile odaklanma düzeyi ve uyku kalitesi arasında anlamlı ilişkiler saptanmış; bu bulgular modelleme aşamasında değişken seçiminde dikkate alınmıştır. Regresyon analizleri sonucunda özellikle *MLP Regressor* ve *Gradient Boosting + Polynomial Features* gibi modellerin yüksek başarı sağladığı görülmüştür. Bu model kombinasyonları, farklı veri setlerinin yapısına göre optimize edilerek değerlendirilmiştir.

Proje sonunda elde edilen en somut katkı, bireysel özelliklere dayalı olarak kafein tüketim düzeyini tahmin eden ve buna uygun önerilerde bulunmayı amaçlayan bir öneri sistemi altyapısının geliştirilmesidir. Bu sistem, yaş, cinsiyet, uyku düzeni, odaklanma düzeyi, ekran süresi gibi bireysel faktörleri girdi olarak alıp kişiye özel kafein önerileri sunmayı hedeflemektedir.

Proje başında hedeflenen bir diğer amaç ise kafein tüketiminin sağlık durumu üzerindeki etkilerini hastalık verileri üzerinden incelemektir. Bu kapsamda açık kaynaklardan *NHANES* gibi büyük veri tabanlarından kalp hastalıkları, depresyon ve benzeri sağlık durumlarıyla ilgili veriler incelenmiştir. Ancak bu veri setlerinde kafein tüketimiyle doğrudan eşleşebilecek örneklerin azlığı ve yapılan analizlerde anlamlı ilişkilerin çıkmaması nedeniyle bu boyut öneri sistemine entegre edilememiştir.

Gelecek çalışmalarda bu eksikliklerin giderilmesi amaçlanmaktadır. Özellikle sağlıkla ilgili kafein etkilerini daha iyi değerlendirmek için kalp rahatsızlıkları, migren, anksiyete gibi durumları içeren veri setleri ile analizler yapılması



planlanmaktadır. Bu tür bireylere yönelik özelleştirilmiş öneriler, sistemin pratik uygulanabilirliğini ve kişiselleştirme düzeyini artıracaktır.

Sonuç olarak bu proje, hem metodolojik hem de uygulamalı açıdan bireye özgü kafein tüketim rehberliği oluşturma yönünde sağlam bir temel sunmakta ve gelecekteki çok-disiplinli sağlık odaklı kişiselleştirilmiş sistemlere örnek teşkil etmektedir.

#### REFERENCES

- [1] A. Smith, "Effects of caffeine on human behavior," *Food and chemical toxicology*, vol. 40, no. 9, pp. 1243–1255, 2002.
- [2] P. Nawrot, S. Jordan, J. Eastwood, J. Rotstein, A. Hugenholtz, and M. Feeley, "Effects of caffeine on human health," *Food Additives & Contaminants*, vol. 20, no. 1, pp. 1–30, 2003.
- [3] M. J. Glade, "Caffeine—not just a stimulant," *Nutrition*, vol. 26, no. 10, pp. 932–938, 2010.
- [4] X. Wang, H. Ma, Q. Sun, J. Li, Y. Heianza, R. M. Van Dam, F. B. Hu, E. Rimm, J. E. Manson, and L. Qi, "Coffee drinking timing and mortality in us adults," *European Heart Journal*, p. ehae871, 2025.
- [5] L. M. Stevens, E. Linstead, J. L. Hall, and D. P. Kao, "Association between coffee intake and incident heart failure risk: a machine learning analysis of the fhs, the aric study, and the chs," *Circulation: Heart Failure*, vol. 14, no. 2, p. e006799, 2021.
- [6] T. M. McLellan, J. A. Caldwell, and H. R. Lieberman, "A review of caffeine's effects on cognitive, physical and occupational performance," *Neuroscience & Biobehavioral Reviews*, vol. 71, pp. 294–312, 2016.
- [7] X.-B. Ma, Y.-L. Lv, L. Qian, J.-F. Yang, Q. Song, and Y.-M. Liu, "Evaluating the effects of coffee consumption on the structure and function of the heart from multiple perspectives," *Frontiers in Cardiovascular Medicine*, vol. 12, p. 1453106, 2025.
- [8] I. Clark and H.-P. Landolt, "Coffee, caffeine, and sleep: A systematic review," *Journal of Clinical Sleep Medicine*, 2017.
- [9] C. Gardiner, "Dose and timing effects of caffeine on subsequent sleep," *Sleep*, 2025.
- [10] S. Wu *et al.*, "Caffeine supplementation improves the cognitive abilities....," *Scientific Reports*, 2024.
- [11] P. Rase-Abdullahi, "Effects of coffee on cognitive function," *Journal of Psychopharmacology*, 2024.