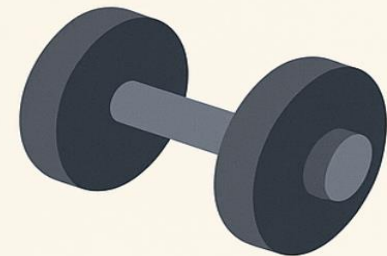




HEALTH AND FITNESS



UNSUPERVISED LEARNING
Özden Gültekin

K-MEANS
<https://www.kaggle.com/code/zdengltekin/health-and-fitness-unsupervised-3>

Gözetimsiz Öğrenmede Kullanılan Metrikler

Aykırı Değer Analizi: IQR yöntemi kullanılmıştır. K-means kümeleme de uzaklık temelli bir algoritma(öklidiyen) kullandığından mutlaka aykırı değerler elimine edilmelidir.

Standartlaştırma: MinMaxScaler Veri setimizin her değişkeninin aralıkları birbirinden çok farklı aralıklardadır. MinMaxScaler bu değişken aralıklarını 0-1 aralığında sınırlandırarak kümeleme performansını arttırmıştır.

Model: K-Means

Model Optimizasyonu: Kümeleme optimizasyonu için elbow(dirsek) yöntemi kullanılmıştır. Bu yöntemle elde ettiğimiz küme sayısı veriyi iyi gruplandırmamıştır.

PCA: Kümelediğimiz veri setini daha düzgün bir şekilde görüntülemek adına kullanıldı. Çünkü Temel bileşenler analizi(PCA) bilgiyi en fazla taşıyan değişkenlerle grafiği temsil eder.
silhouette_score: Elbow yöntemiyle optimize edilmiş kümeleme sayısı ile kümeleme başarısı değerlendirilmiştir.

Problem Tanımı

- HAM VERİ İLE YAPILAN İLK KÜMELEME DENEMELERİNDE ANLAMLI GRUPLAR ELDE EDİLEMEDİ.

- VERİ TEKRARLARI VE YÜKSEK BOYUT, MODEL EĞİTİMİNİN UZUN SÜRMESİNE NEDEN OLDU.

- KÜMELER ARASI FARKLAR YETERİNCE BELİRGİN DEĞİLDİ.

Veri Dönüştürme Yaklaşımı

- Her bireyin günlük verilerinin ortalamaları alınarak yeni bir veri seti oluşturuldu.

- Bu dönüşüm ile:

- Temsili ve sade gözlemler elde edildi.

- Model eğitimi hızlandı.

- Kümeler arası farklar daha net hale geldi.

Uygun Küme Sayısının Belirlenmesi

- İlk incelemeler: fitness_level 2 gruba ayrılıyor gibi.

- Elbow yöntemi ile optimum küme sayısı belirlendi.



- Silhouette skoru ile kümenin kalitesi ölçümlendi.

- Sonuç: 2 küme veri yapısına en uygun yapı olarak seçildi. Bu seçim veriye dayanarak yapıldı.

Elde Edilen Segmentler

Cluster 1 –   Kondisyonlu Yetişkinler:

- 40 yaş üstü, düşük HR, yüksek kondisyon.

Cluster 2 –   Genç ve Aktif Bireyler:

- 20–40 yaş, yüksek HR, kondisyon değişken.

Segment Bazında Öneriler

Cluster 1 – Kondisyonlu Yetişkinler

(40+ yaş)

Özellikler:

Düşük nabız, yüksekondisyon



Hafif tempolu kardiyo
(yürüyüş, yuzme)
→ haftada 3–4 gün



Yoga & denge
egzersizleri
→ düşme riskine karşı



Hafif direnç
antrenmanları
→ kas kaybını önler



Akdeniz tipi
beslenme



Stres yönetimi

Cluster 2 – Genç ve Aktif Bireyler

(20–40 yaş)



Özellikler: Yüksek nabız
değişken kondisyon



Kombine egzersiz
(kardiyo + kuvvet)



HIIT / interval →
kondisyon artırır



Nabız takibi →
aşırı efordan kaçın



Dinlenme günleri
→ toparlanma şartı



Dengeli beslenme
(protein + komple

Arayüz Geliştirme ve Gelecek Plan

- Kullanıcıların veri girişi yapabileceği bir arayüz geliştirilecektir. (Streamlit).
- Veri seti, farklı yaş ve sağlık geçmişine sahip bireylerle zenginleştirilmeli.
- Daha fazla çeşitlilik, modelin genelleme gücünü artıracaktır.

Sonuç

- Ortalama verilere geçiş, kümeler arası farkları netleştirdi.

- 2 küme kararı, hem görsel hem istatistiksel olarak desteklendi.

- Arayüz ve chatbot eklentisi ile arayüz geliştirilebilir. Veri zenginleştirme ile segmentasyon daha verimli bir şekilde yapılabilir.