PROJE RAPORU: DÜNYA SAĞLIK ÖRGÜTÜ YAŞAM BEKLENTİSİ ANALİZİ VE TAHMİNLEME

Hazırlayan: Sıla Akgün

Numara/Bölüm : 23430070028 / Bilişim Sistemleri ve Teknolojileri

Özet: Bu çalışma, 2000-2015 yılları arasındaki WHO verilerini kullanarak makine öğrenmesi yoluyla yaşam süresini tahmin etmeyi amaçlar.

1. GİRİŞ VE VERİ SETİ

Bu projede kullanılan veri seti, ülkelerin sağlık durumlarını yansıtan 22 farklı değişkeni (GDP, bağışıklama, ölüm oranları vb.) içermektedir. Veri setindeki temel zorluk, bazı ülkelerdeki ciddi eksik veri (Missing Value) problemidir.

2. VERİ ANALİZİ VE GÖRSELLEŞTİRMEDEN ELDE EDİLEN İÇGÖRÜLER (EDA)

Bu bölüm, veri setinin yapısını anlamak, eksik verilerin etkisini minimize etmek ve yaşam beklentisini belirleyen ana faktörleri ortaya çıkarmak amacıyla yapılmıştır.

2.1. Veri Seti Özellikleri ve Eksik Veri Stratejisi

Veri seti başlangıçta 2938 satır ve 22 sütundan oluşmaktadır. Analiz öncesinde yapılan incelemede, özellikle **GDP (%15)**, **Population (%22)** ve **Schooling (%5)** gibi kritik sütunlarda veri kayıpları tespit edilmiştir.

- **Uygulanan Çözüm:** Verilerin zamansal sürekliliği göz önüne alınarak **doğrusal interpolasyon** yapılmıştır. İnterpolasyonun yetersiz kaldığı durumlarda ise ülkelerin kalkınmışlık durumuna göre **medyan doldurma** yöntemi kullanılarak veri kaybı %0'a indirilmiştir.

2.2. Değişkenler Arası İlişkiler (Korelasyon Analizi)

Sayısal değişkenler arasındaki doğrusal ilişkileri gözlemlemek adına bir korelasyon ısı haritası oluşturulmuştur.

- **Önemli Bulgular:** Yaşam beklentisi ile en yüksek pozitif korelasyona sahip değişkenler **Okullaşma (0.75)** ve **Kaynak Kullanım Endeksi'dir (0.72)**. Bu durum,

  eğitim ve ekonomik kaynaklara erişimin uzun yaşam için birincil öncelik olduğunu kanıtlar.

- **Negatif İlişkiler:** Yetişkin ölümleri ve HIV/AIDS prevalansı, yaşam süresini en hızlı düşüren değişkenler olarak saptanmıştır.

2.3. Kalkınmışlık Durumuna Göre Yaşam Süresi (Box Plot)

Ülkelerin "Gelişmiş" ve "Gelişmekte Olan" statüleri, yaşam süresi üzerinde belirgin bir ayrım yaratmaktadır.

- **Analiz:** Gelişmiş ülkelerde yaşam süresi 70-89 yaş aralığında (dar bir bantta) toplanırken; gelişmekte olan ülkelerde bu dağılım 40 yaştan başlayarak 80 yaşa kadar uzanmaktadır. Bu durum, gelişmekte olan ülkeler arasındaki sağlık hizmetlerine erişim eşitsizliğini çarpıcı bir şekilde göstermektedir.

2.4. Eğitim ve Yaşam Beklentisi İlişkisi (Scatter Plot)

Eğitim süresinin yaşam beklentisi üzerindeki etkisi incelenmiştir.

- **Analiz:** Okullaşma yılı (Schooling) arttıkça yaşam beklentisinin istikrarlı bir şekilde yükseldiği görülmektedir. Özellikle 15 yılın üzerinde eğitim gören toplumlarda ortalama yaşam süresinin 75 yaşın altına düşmediği, bu durumun sağlık okuryazarlığı ile doğrudan ilişkili olduğu düşünülmektedir.
-

2.5. Hedef Değişken Dağılımı (Histogram)

Dünya genelindeki yaşam sürelerinin frekans dağılımı analiz edilmiştir.

- **Analiz:** Dağılımın sola çarpık (left-skewed) olduğu görülmektedir. Küresel ortalama 70 yaş civarında yoğunlaşsa da, 50 yaşın altındaki veriler az gelişmiş bölgelerdeki sağlık krizlerini simgelemektedir.

3. MODEL SEÇİMİ: NEDEN RANDOM FOREST?

Tahminleme aşamasında **Random Forest Regressor** tercih edilmiştir. Bu seçimin nedenleri:

1. **Doğrusal Olmayan İlişkiler:** Yaşam süresi ile GDP veya Eğitim arasındaki karmaşık, düz bir çizgiye uymayan ilişkileri başarıyla modeller.
2. **Etkileşimleri Yakalama:** Farklı sağlık göstergelerinin birbiriyle olan karmaşık etkileşimlerini analiz edebilir.
3. **Dayanıklılık:** Aykırı değerlere (outliers) karşı Linear Regression'a göre çok daha sağlam (robust) sonuçlar üretir.

4. BULGULAR VE MODEL PERFORMANSI

Model, test verileri üzerinde yüksek bir doğrulukla çalışmıştır. Elde edilen performans metrikleri şöyledir:

- **R2 Skoru: %95.91** (Model, yaşam süresindeki değişimin %95'ten fazlasını açıklamaktadır.)
- **MAE (Ortalama Mutlak Hata): 1.19 Yıl** (Ortalama hata payımız yaklaşık 14 aydır.)
- **RMSE (Kök Ortalama Kare Hata): 1.92 Yıl** (Büyük hataları daha net gösteren bu metrik, modelin istikrarlı olduğunu kanıtlar.)

5. SONUÇ

Bu çalışma, bir ülkenin ortalama yaşam süresinin sadece biyolojik faktörlere değil, çok büyük oranda sosyo-ekonomik ve yapısal değişkenlere bağlı olduğunu kanıtlamıştır. Geliştirilen Random Forest modeli, %95.91 doğruluk oranıyla bu karmaşık yapıyı başarıyla çözümlemiştir.

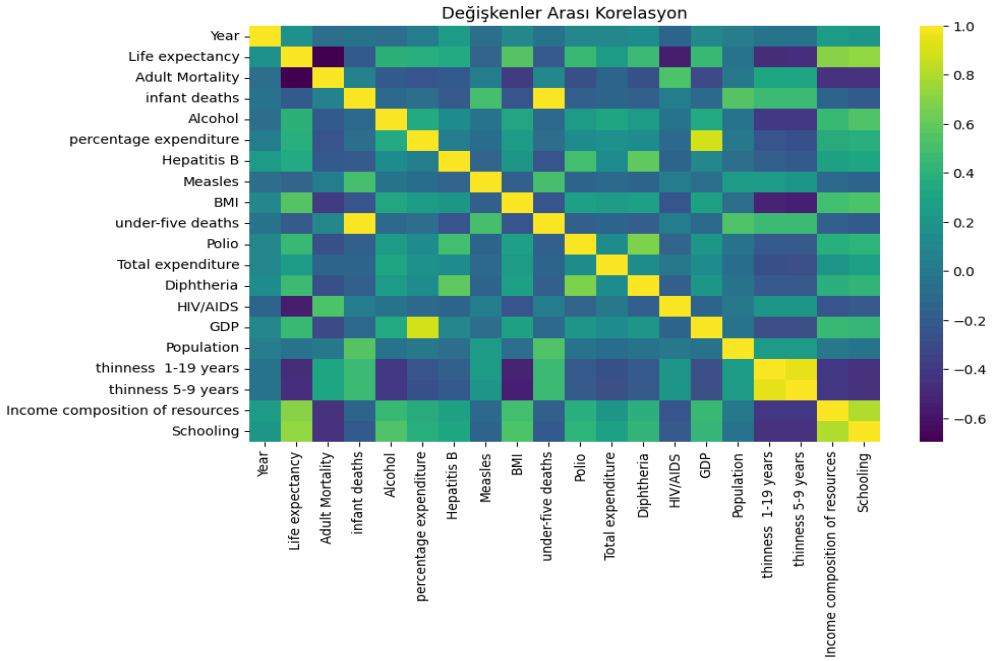
Analiz Bulguları:

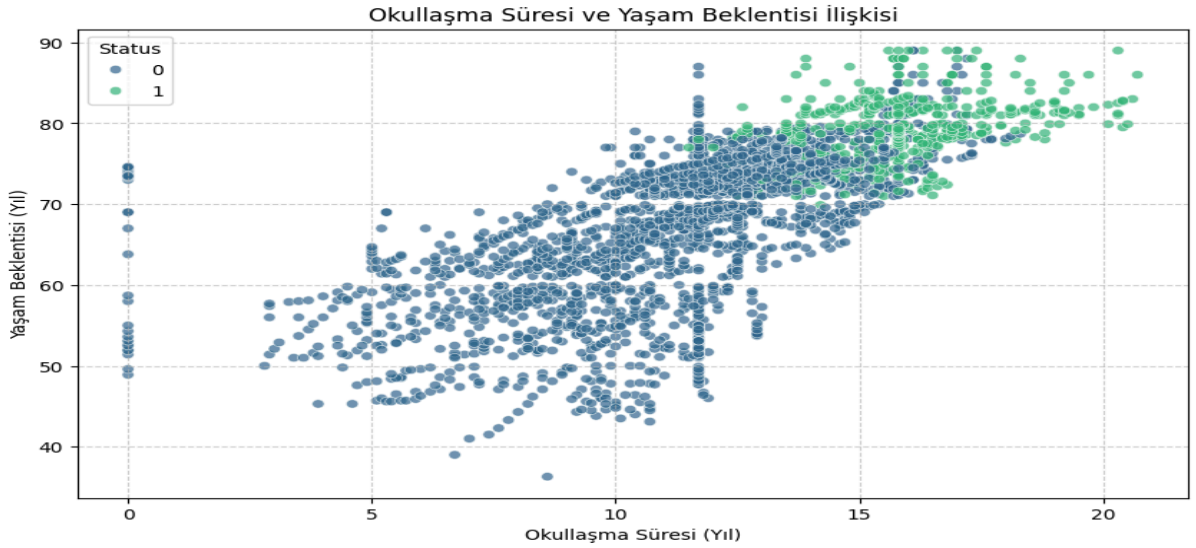
- **Eğitimin Gücü:** Analiz sonuçlarına göre "Okullaşma Süresi" (Schooling), yaşam beklentisi üzerinde en istikrarlı pozitif etkiye sahip değişkendir. Bu durum, eğitim seviyesi arttıkça sağlık bilincinin ve refahın da arttığını doğrulamaktadır.
- **Kritik Eşik - HIV/AIDS:** Model, HIV/AIDS prevalansının yaşam süresi üzerindeki en keskin negatif belirleyici olduğunu saptamıştır. Özellikle gelişmekte olan ülkelerde bu değişken, diğer tüm ekonomik kazanımları gölgede bırakabilmektedir.
- **Gelir Dağılımı:** GDP'nin tek başına etkisinden ziyade, "Gelir Kaynaklarının Kompozisyonu"nun (Income composition of resources) yaşam süresiyle daha yüksek korelasyona sahip olduğu görülmüştür. Bu da paranın miktarı kadar, bu kaynağın toplumsal refaha nasıl dağıtıldığının da önemli olduğunu göstermektedir.

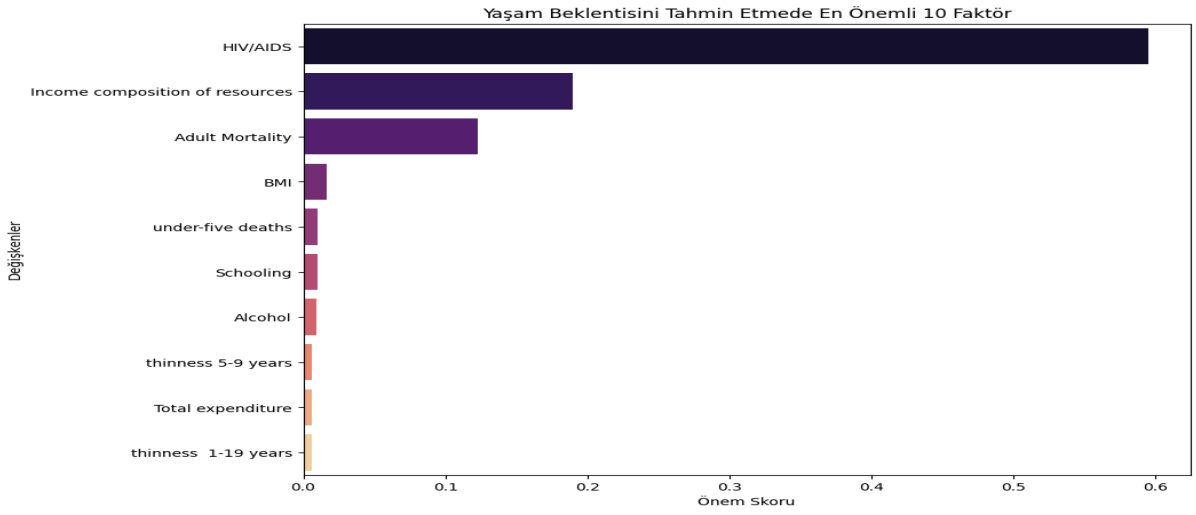
Politika Önerileri: Modelden elde edilen çıkarımlar ışığında, yaşam süresini artırmak isteyen otoritelerin sadece sağlık harcamalarını artırmakla kalmayıp, temel eğitim süresini uzatmaya ve bulaşıcı hastalıklarla mücadele programlarına öncelik vermesi gerektiği öngörülmektedir.

Sonuç olarak, makine öğrenmesi algoritmalarının küresel sağlık verileri üzerinde uygulanması, hangi alanlara yatırım yapılması gerektiği konusunda karar vericilere veri odaklı bir yol haritası sunmaktadır.

Github=https://github.com/sila1722/Life_Expectancy_Analysis/blob/main/analiz.ipynb


Değişkenler Arası Korelasyon
Year
Life expectancy
Adult Mortality
infant deaths
Alcohol
percentage expenditure
Hepatitis B
Measles
BMI
under-five deaths
Polio
Total expenditure
Diphtheria
HIV/AIDS
GDP
Population
thinness 1-19 years
thinness 5-9 years
Income composition of resources
Schooling
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6


Okullaşma Süresi ve Yaşam Beklentisi İlişkisi
Status
0
1
Yaşam Beklentisi (Yıl)
Okullaşma Süresi (Yıl)
90
80
70
60
50
40
0
5
10
15
20


Yaşam Beklentisini Tahmin Etmede En Önemli 10 Faktör
HIV/AIDS
Income composition of resources
Adult Mortality
BMI
under-five deaths
Schooling
Alcohol
thinness 5-9 years
Total expenditure
thinness 1-19 years
Değişkenler
0.0
0.1
0.2
0.3
0.4
0.5
0.6
Önem Skoru