

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 ÖN İŞLEME VE METODOLOJİ

Veri kalitesini artırmak adına şu stratejiler izlenmiştir:

- İnterpolasyon:** Ülke bazlı zaman serisi verileri için doğrusal interpolasyon uygulanmıştır. Bu sayede bir ülkenin kayıp yılı, önceki ve sonraki yılların trendine göre doldurulmuştur.
- Status Bazlı Medyan:** Hiç verisi olmayan ülkeler için "Gelişmiş" veya "Gelişmekte Olan" (Status) gruplarının medyan değerleri kullanılarak veri bütünlüğü sağlanmıştır.

3. MODEL SEÇİMİ: NEDEN RANDOM FOREST?

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

- Doğrusal Olmayan İlişkiler:** Yaşam süresi ile GDP veya Eğitim arasındaki karmaşık, düz bir çizgiye uymayan ilişkileri başarıyla modeller.
- Etkileşimleri Yakalama:** Farklı sağlık göstergelerinin birbiriyle olan karmaşık etkileşimlerini analiz edebilir.
- 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 çözümlenmiştir.

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.