



4. DÖNEM 2. GRUP VERİ ANALİTİĞİ GRUP BİTİRME PROJESİ RAPORU

KÜRESEL ISINMA, AŞIRI HAVA OLAYLARI VE GÖÇ İLİŞKİSİ

Beyza Fatıma Çekerekli

Dilan Sazan

Esmâ Nur Beğbağâ

Feyza Nur Demirbaş

Hilal Üçüncü

Şevval Ok

SoftlTo Yazılım Bilişim Akademisi

Ağustos 2026

1. Giriş
2. Veri Kaynakları
3. İlgili Literatür
 - 3.1 Literatürden Ortak Bulgular
 - 3.2 Belirlenen Araştırma Boşlukları
 - 3.3 Projenin Literatürdeki Konumu
4. Bölge Seçimi
5. Veri Temizleme
 - 5.1 Göç Verisi
 - 5.2 İklim (Sıcaklık / Yağış) Verisi
 - 5.3 Afet Verisi
6. Veri Birleştirme
7. Keşifsel Veri Analizi (EDA)
8. Hipotezler ve Test Sonuçları
 - H1 - Kuraklık arttıkça göç artar
 - H7 - Sel, kuraklıktan daha güçlü ve ani bir etki yaratır
 - H5 - Kuraklığın etkisi Toplam Etkilenen üzerinden dolaylı olarak göçe yansır
 - H20 - Çok değişkenli iklim modeli, tek değişkenliden daha iyi tahmin eder
9. Modelleme Çalışması 1: CatBoost & Ridge
10. Modelleme Çalışması 2: Ağaç Tabanlı Model Karşılaştırması
11. Sonuç
12. Kısıtlar ve Gelecek Çalışmalar
 - Proje Ekibi
13. Kaynakça

1. Giriş

Bu proje, küresel ısınmayla ilişkili iklim değişkenlerinin ve aşırı hava olaylarının sınır ötesi göç hareketleriyle olan ilişkisini incelemektedir. Çalışma, gözlemsel iklim ve afet verilerini UNHCR'in resmi sınır ötesi göç istatistikleriyle birleştirerek, iklimsel baskıların göç üzerindeki etkisini istatistiksel olarak test etmeyi ve makine öğrenmesi modelleriyle tahmin edilebilirliğini ölçmeyi amaçlamaktadır.

Projede göç, sadece ülkeler arası hareket olarak tanımlanmıştır; bir ülke içinde yerinden edilme göç kapsamı dışında tutulmuştur. Bu tercih, IDMC gibi iç yerinden edilme verisi yerine UNHCR'in ülkeler arası göç verisinin esas alınmasına yol açmıştır.

2. Veri Kaynakları

Analizde iki temel veri tablosu kullanılmıştır:

- UNHCR'in yayınladığı, yıl - çıkış ülkesi (Country of Origin) - varış ülkesi (Country of Asylum) bazında kişi sayısını (Total) içeren ikili göç verisi.
- Ülke - yıl bazında ortalama sıcaklık (Avg_Temp), ortalama yağış (Avg_Precip), afet türü sayıları (Kuraklik_Sayisi, Sel_Sayisi, Firtina_Sayisi, Asiri_Sicaklik_Sayisi, toplam Afet_Sayisi) ve afetlerin etkisini gösteren Toplam_Olum, Toplam_Etkilenen, Toplam_Hasar_1000USD değişkenlerini içeren tablo.

Her iki dosya da 6 farklı UNHCR bölgesinden derlenen ham verilerin temizlenmesi ve birleştirilmesiyle oluşturulmuştur.

3. İlgili Literatür

Proje kapsamında, küresel ısınma - aşırı hava olayları - göç ilişkisini inceleyen 16 akademik/kurumsal çalışmayı kapsayan bir literatür taraması yapılmıştır. Taranan çalışmalar ekonometrik panel analizleri, sistematik derlemeler ve makine öğrenmesi temelli tahmin modelleri olmak üzere üç ana metodolojik kümeye ayrılmaktadır. Bu bölümde literatürden çıkan ortak bulgular özetlenmekte ve projenin bu literatürdeki konumu açıklanmaktadır.

3.1 Literatürden Ortak Bulgular

- İklim şoklarının göç üzerindeki etkisi doğrusal değildir; gelir düzeyi, yönetim kalitesi ve bölgesel bağlama göre koşulludur [1][2].
- Aşırı hava olaylarının sıklığı, çoğu zaman ortalama sıcaklık/yağış değişiminden daha güçlü bir göç itici güçtür; ancak ekonometrik literatürde bu boyut hâlâ yeterince temsil edilmemektedir [4].
- Ağaç tabanlı makine öğrenmesi modelleri (Random Forest, XGBoost), klasik istatistiksel yöntemlere (ARIMA, doğrusal regresyon) kıyasla artan biçimde tercih edilmekte ve genellikle daha yüksek tahmin doğruluğu sunmaktadır.

- Düşük gelirli grupların göç edecek kaynaktan yoksun kalması ("hapsolmuş nüfus" / trapped populations) nicel modellerde yeterince yansıtılmayan bir konudur (Sedova & Kalkuhl, 2024) - bu kavram, projenin H5 bulgusunu ve Ridge modelindeki negatif kuraklık katsayısını yorumlamak için doğrudan kullanılmıştır.
- Sahra-altı Afrika'da hava anomalilerinin 1960-2000 arası toplam 5 milyon net yerinden edilmeye yol açtığını göstermiştir; bu çalışma iklim-göç ilişkisinin Afrika bağlamındaki en sık atıf yapılan kaynaklarından biridir ve projenin bölge seçimini de destekleyen bir literatür temeli sunmaktadır.[6]

3.2 Belirlenen Araştırma Boşlukları

Literatür taramasında öne çıkan araştırma boşlukları şunlardır: nicel/ML tabanlı iklim-göç tahmin çalışmalarının literatürde hâlâ azınlıkta olması; iç göç ve uluslararası göçün aynı modelde birlikte ele alındığı çalışma sayısının sınırlı olması; aşırı hava olayı sıklığının (yalnızca ortalama sıcaklık/yağış değil) çoğu ekonometrik modelde yeterince temsil edilmemesi; ve standart, karşılaştırılabilir bir değerlendirme çerçevesinin eksikliği.

3.3 Projenin Literatürdeki Konumu

Bu proje, literatürde tanımlanan iki boşluğa doğrudan katkı sunmayı hedeflemektedir: (1) iklim, afet ve göç verilerinin tek bir veri analitiği çerçevesinde sistematik olarak birleştirilmesi, ve (2) aşırı hava olayı sıklığının (yalnızca ortalama iklim değil) göç üzerindeki etkisinin ayrı ayrı test edilmesi (bkz. H1 ve H7). Ayrıca projenin iki bağımsız modelleme çalışmasında da ağaç tabanlı modellerin doğrusal modelleri geçmesi, literatürdeki genel eğilimle birebir örtüşmektedir.

4. Bölge Seçimi

UNHCR verisinde 6 bölge (Asya-Pasifik, Amerika, Avrupa, Ortadoğu-Kuzey Afrika, Doğu-Güney Afrika, Batı-Orta Afrika) mevcuttur. Zaman ve kapsam kısıtları nedeniyle projede yalnızca iki bölge ile çalışılmıştır: Doğu-Güney Afrika ve Batı-Orta Afrika.

Bu iki bölgenin seçilme gerekçesi aşağıdaki kriterlere dayanmaktadır:

- İklim değişikliğine coğrafi ve ekonomik açıdan yüksek kırılganlık gösteren, kuraklık/sel gibi afetlerin sık görüldüğü bölgeler olmaları.
- Göç hareketliliğinin (hem hacim hem de yıllar arası değişkenlik açısından) analiz için yeterli varyasyonu sunması.
- Verinin belirli birkaç ülkeye aşırı yoğunlaşmamış olması; yani bulguların tek bir kriz ülkesinin sonucu olmaktan çok, bölgesel bir eğilimi yansıtmayı hedeflenmiştir.
- Veri tamlığı: İki bölgenin de sıcaklık/yağış ve afet kayıtlarının 2001-2025 aralığını kesintisiz kapsaması.

Diğer 4 bölgenin ham temizlik kodu proje deposunda referans amaçlı bırakılmış, ancak analiz ve modelleme yalnızca seçilen 2 bölge üzerinden yürütülmüştür.

5. Veri Temizleme

5.1 Göç Verisi

Doğu-Güney Afrika ve Batı-Orta Afrika bölgelerine ait ham göç dosyaları ayrı ayrı temizlenip birleştirilmiştir:

- Doğu-Güney Afrika ham veri boyutu: 19.991 satır x 20 sütun.
- İki bölgenin temizlenmiş dosyaları birleştirildiğinde toplam 43.555 satır elde edilmiş, mükerrer satır tespit edilmemiştir.
- UNHCR bilateral verisinde "Country of Origin" ile "Country of Asylum" aynı olan 393 satır tespit edilmiştir. Bu satırlar ülke içi yerinden edilmeyi temsil ettiği ve proje kapsamında göç sınır ötesi hareket olarak tanımlandığı için veri setinden çıkarılmıştır.
- Temizleme sonrası nihai göç tablosu: 43.162 satır.

5.2 İklim (Sıcaklık / Yağış) Verisi

Sıcaklık ve yağış verileri orijinalinde geniş formatta (her ay/yıl ayrı sütun) gelmiş, analiz için uzun formata dönüştürülmüştür. Dönüşüm sırasında ortaya çıkan eksik değerler ilgili ülkenin kendi zaman serisi içinde tahminleme ile doldurulmuştur.

5.3 Afet Verisi

Ham afet verisi 12.906 satır x 25 sütundan oluşmaktadır. Afet türü dağılımı: Sel 6.203, Fırtına 5.169, Kuraklık 804, Aşırı Sıcaklık 730 kayıttır.

- 188 satırda başlangıç/bitiş tarihi sırasının hatalı olduğu (bitiş, başlangıçtan önce) tespit edilip düzeltilmiştir.
- Düzeltme sonrası veriler ülke-yıl bazında özetlenerek afet_yil_ulke_ozet tablosu oluşturulmuştur: 2.825 satır x 11 sütun.

6. Veri Birleştirme

Göç verisi ikili yapıda, yani her satır bir (Yıl, Çıkış Ülkesi, Varış Ülkesi) üçlüsüne karşılık gelmektedir. İklim/afet verisi ise Yıl-Ülke bazındadır. Bu iki tabloyu doğrudan Yıl ve Çıkış Ülkesi üzerinden birleştirmek önemli bir hataya yol açmaktadır: her iklim/afet kaydı, o ülkenin o yıl kaç farklı varış ülkesine göç veriyse o kadar kez kopyalanır (pseudo-replication / sahte tekrar). Bu sorun ekip tarafından da bağımsız olarak tespit edilmiş; doğrudan birleştirmede Çıkış-Yıl başına ortalama tekrar sayısının 36,6 olduğu hesaplanmıştır.

Çözüm olarak iki tablo şu şekilde ayrı tutulmuştur:

- Bilateral düzeyde göç tablosu (goc_tablosu): kim nereye, ne zaman, kaç kişi.
- Ülke-Yıl düzeyinde tekilleştirilmiş iklim/afet tablosu (iklim_afet_tablosu).

Nihai birleşik tablo oluşturulurken göç verisi önce Yıl + Çıkış Ülkesi bazında toplanmış (Toplam_Goc), ardından iklim/afet tablosuna eklenmiştir.

Sonuç:

- birlesik_tablo: 1.190 satır x 14 sütun.
- 48 farklı ülke, 2001-2025 yılları.
- 0 mükerrer satır, 0 eksik değer.

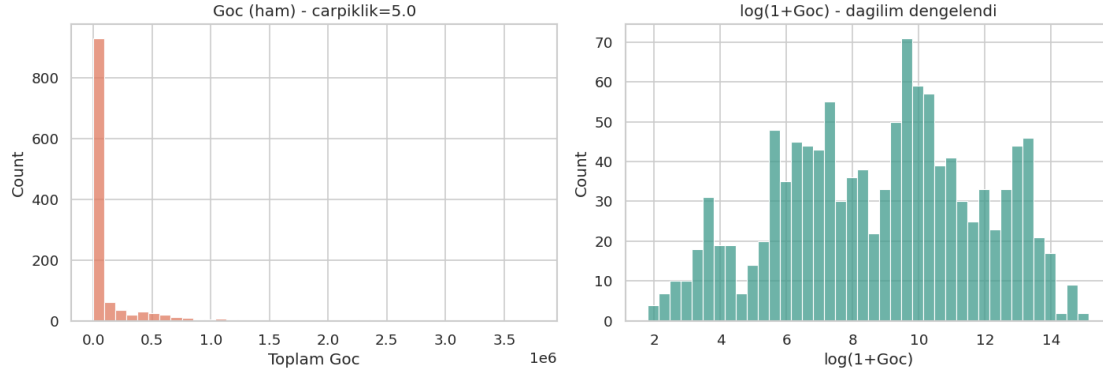
Önemli bir noktayı vurgulamak gerekir: satır sayısının 43.162'den 1.190'a düşmesi veri kaybı değildir; bilateral kayıtların ülke-yıl bazında toplanmasının doğal sonucudur (detay seviyesi değişir, toplam kişi sayısı korunur). Bu birleşik tablo, hem hipotez testlerinde hem de her iki modelleme çalışmasında ortak veri kaynağı olarak kullanılmıştır.

7. Keşifsel Veri Analizi (EDA)

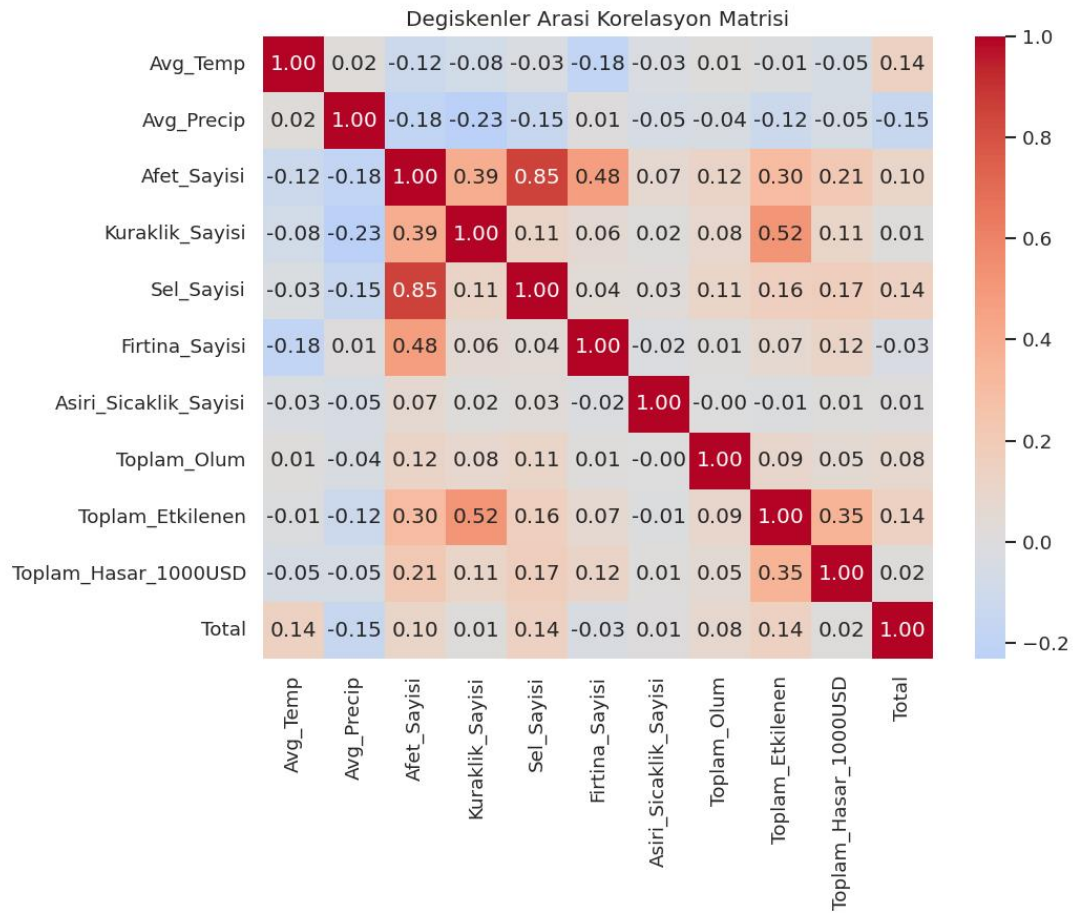
Birleşik tablo üzerinde yapılan keşifsel analizde, hedef değişken olan göç belirgin şekilde sağa çarpık bir dağılım göstermektedir. Çarpıklık katsayısı, verinin hangi düzeyde ölçüldüğüne göre değişmektedir: ham bilateral göç kayıtlarında (43.162 satır, ülke çiftleri bazında) çarpıklık $\approx 19,6$ iken, modelleme ve hipotez testlerinde fiilen kullanılan ülke-yıl bazında toplanmış Total değişkeninde (1.190 satır) çarpıklık $\approx 4,99$ 'dur. İkisi de aynı yönde (aşırı sağa çarpık) bir dağılımı gösterir; farkın sebebi toplama sonucu uç değerlerin kısmen yumuşamasıdır. Modelleme ve hipotez testleri için geçerli olan referans değer 4,99'dur.

Bu çarpıklığın temel sebebi, kriz/savaş yaşayan bazı ülkelerin göç sayılarının milyonlara ulaşp dağılımın geri kalanından çok uzaklaşmasıdır.

Bu nedenle hem hipotez testlerinde hem de modelleme çalışmalarında hedef değişken $\log(1+x)$ dönüşümü ($\log(1+x)$) ile analiz edilmiş, böylece dağılım normale yakın bir hale getirilerek hem korelasyon hem de regresyon analizlerinin daha güvenilir sonuç vermesi sağlanmıştır. Aynı çarpıklık Toplam_Olum, Toplam_Etkilenen ve Toplam_Hasar_1000USD değişkenlerinde de gözlemlenmiş, bu değişkenler de modelleme aşamasında log dönüşümüne tabi tutulmuştur.



Şekil 1. Göç değişkeninin ham ve $\log(1+p)$ dönüşümlü dağılımı (ülke-yıl bazında, $n=1.190$).



Şekil 2. İklim/afet değişkenleri ve göç arasındaki korelasyon matrisi.

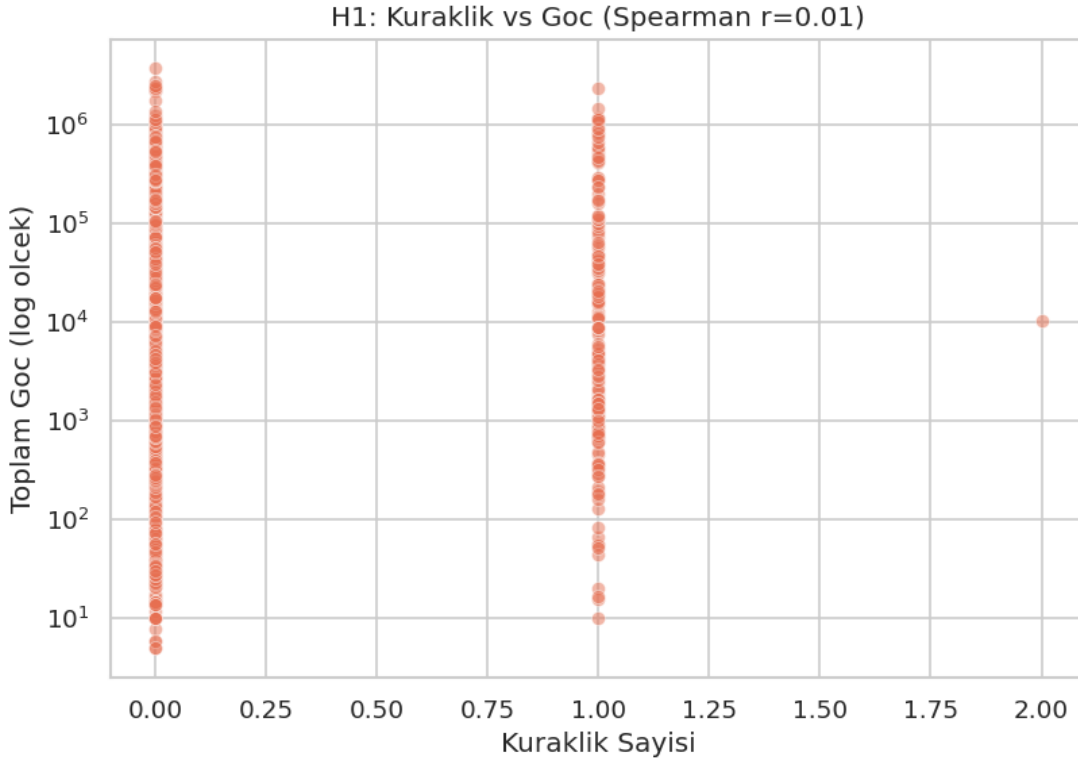
8. Hipotezler ve Test Sonuçları

Ekibin hazırladığı geniş hipotez listesinden, birleşik tablo (goc_tablosu + iklim_afet_tablosu) ile doğrudan test edilebilecek ve birbiriyle anlamlı bir bütün oluşturan 3 ana hipotez (H1, H7, H5) ve bunlara model karşılaştırması açısından köprü görevi gören 1 ek hipotez (H20) seçilmiş ve test edilmiştir.

H1 - Kuraklık arttıkça göç artar

Kuraklik_Sayisi ile Total Göç arasındaki ilişki Pearson ve Spearman korelasyonu ile ölçülmüştür: Pearson $r = 0,013$.

Sonuç: İlişki çok zayıf. Kuraklık tek başına göç miktarını açıklıyor gibi görünmüyor; hipotez desteklenmemiştir.

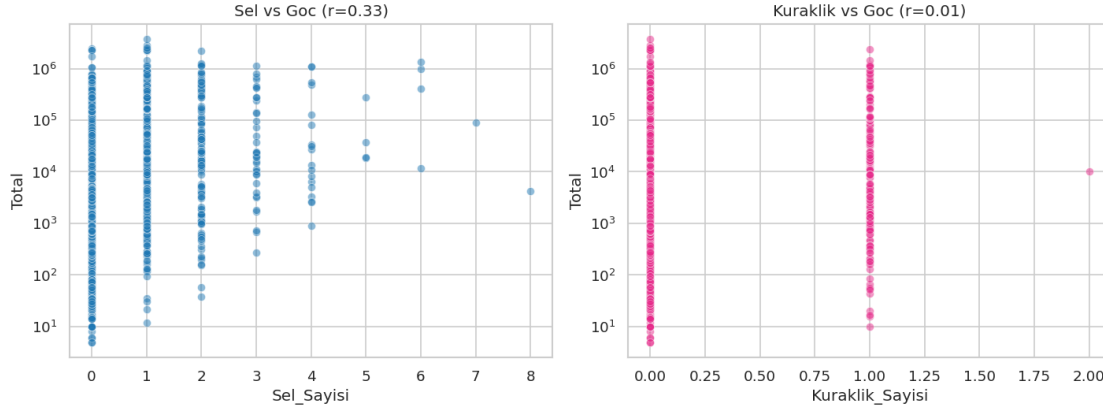


Şekil 3. H1 - Kuraklık sayısı ile göç arasındaki ilişki.

H7 - Sel, kuraklıktan daha güçlü ve ani bir etki yaratır

Sel_Sayisi ile göç arasındaki Spearman korelasyonu ($r = 0,326$), Kuraklik_Sayisi ile göç arasındaki korelasyondan ($r = 0,011$) belirgin şekilde daha yüksek çıkmıştır.

Sonuç: Hipotez destekli. Ani felaketler (sel) kronik/yavaş gelişen felaketlere (kuraklık) kıyasla sınır ötesi göç ile daha güçlü ilişkilidir.

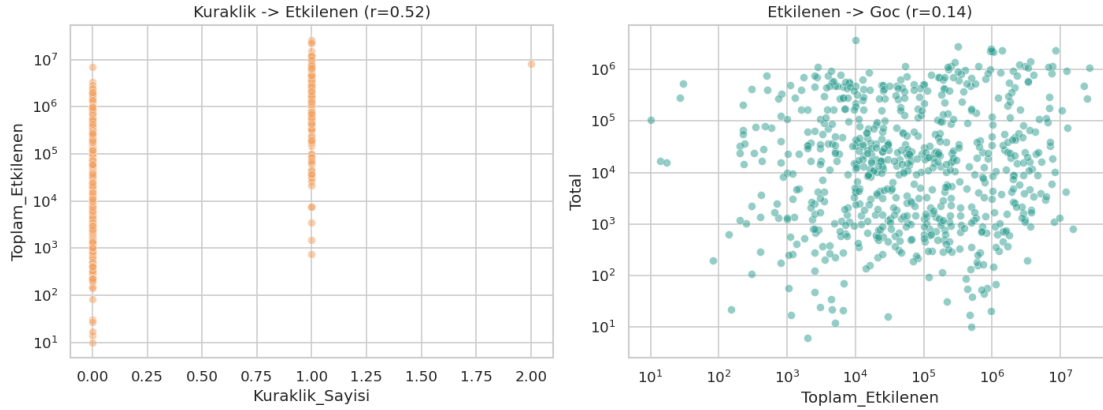


Şekil 4. H7 - Sel ve kuraklığın göçle ilişkisinin karşılaştırılması.

H5 - Kuraklığın etkisi Toplam Etkilenen üzerinden dolaylı olarak göçe yansır

Kuraklık - Toplam_Etkilenen korelasyonu: $r = 0,521$ (güçlü). Toplam_Etkilenen - Göç korelasyonu: $r = 0,139$ (zayıf). Kuraklık - Göç doğrudan korelasyonu çok zayıf.

Sonuç: Kısmi destek. Kuraklık, etkilenen kişi sayısını güçlü şekilde artırıyor, ancak bu etki sınır ötesi göçe zayıf yansıyor. Bu durum, göç literatüründeki "hapsolmuş nüfus" kavramıyla açıklanabilir: kuraklıktan etkilenen insanların bir kısmı, sınır ötesi göç edecek maddi/lojistik imkâna sahip olmadığı için etki büyük ölçüde ülke içinde kalıyor olabilir.

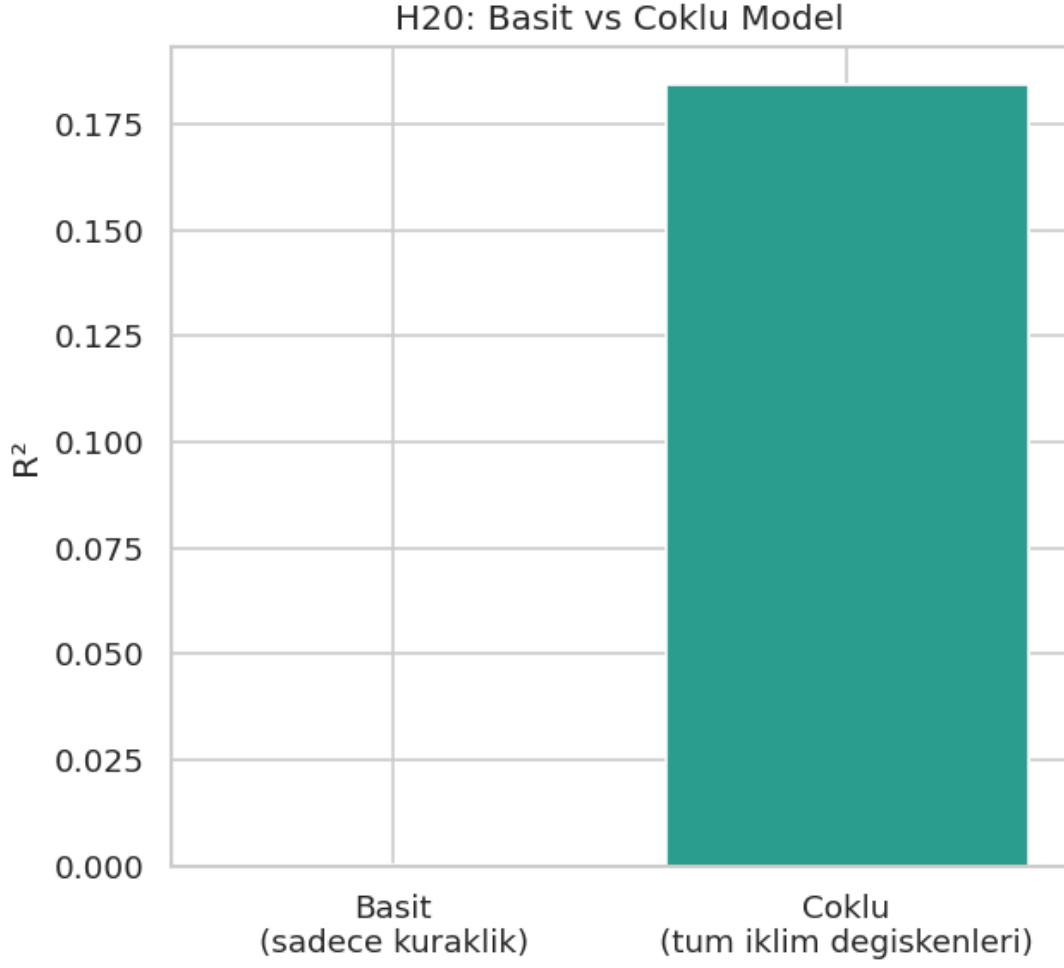


Şekil 5. H5 - Kuraklığın etkilenen nüfus üzerinden göçe dolaylı yansımaları.

H20 - Çok değişkenli iklim modeli, tek değişkenliden daha iyi tahmin eder

Sadece Kuraklik_Sayisi ile kurulan basit regresyon modeli $R^2 = 0,0003$ verirken, 6 iklim/afet değişkeninin (Avg_Temp, Avg_Precip, Kuraklik_Sayisi, Sel_Sayisi, Firtina_Sayisi, Asiri_Sicaklik_Sayisi) birlikte kullanıldığı çok değişkenli model $R^2 = 0,184$ vermiştir.

Sonuç: Hipotez destekli. Göç, tek bir iklim değişkeniyle değil, birden fazla iklim/afet değişkeninin birlikte etkisiyle daha iyi açıklanabiliyor - bu bulgu, modelleme aşamasında tek değişkenli yaklaşımlar yerine çoklu değişkenli / makine öğrenmesi modellerine geçişin gerekçesini oluşturmuştur.



Şekil 6. H20 - Tek değişkenli ve çok değişkenli modelin R² karşılaştırması.

Hipotez	Sonuç	Değerlendirme
H1: Kuraklık -> Göç	Pearson r = 0,013	Desteklenmedi
H7: Sel vs Kuraklık	Sel r = 0,326 > Kuraklık r = 0,011	Destekli
H5: Dolaylı etki (mediation)	Kuraklık->Etkilenen r=0,521; Etkilenen->Göç r=0,139	Kısmi destek
H20: Basit vs çoklu model	Basit R ² =0,0003; Çoklu R ² =0,184	Destekli

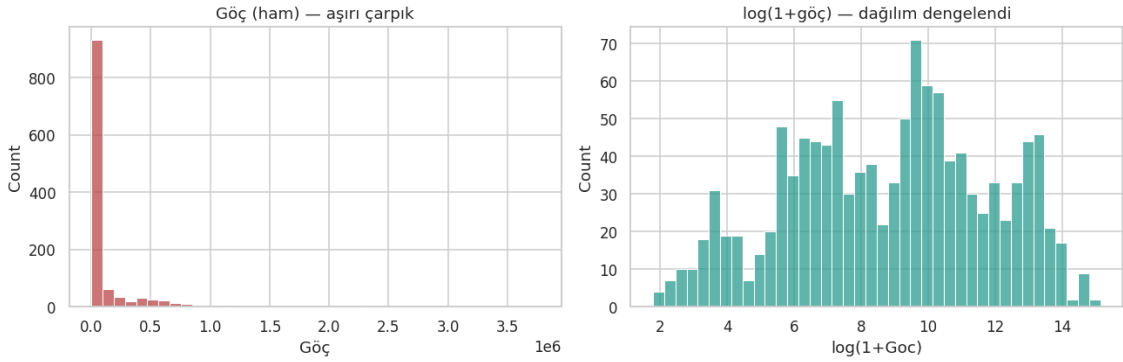
9. Modelleme Çalışması 1: CatBoost & Ridge

Bu çalışmada birlesik_tablo verisi üzerinde, hedef değişken $\log_{1p}(\text{Total})$ olacak şekilde iki farklı model ailesi karşılaştırılmıştır: ağaç tabanlı CatBoost ve doğrusal Ridge regresyon. Amaç, iklim/afet değişkenleri ile göç arasındaki ilişkinin doğrusal mı yoksa eşik/etkileşim içeren daha karmaşık bir yapıda mı olduğunu test etmektir.

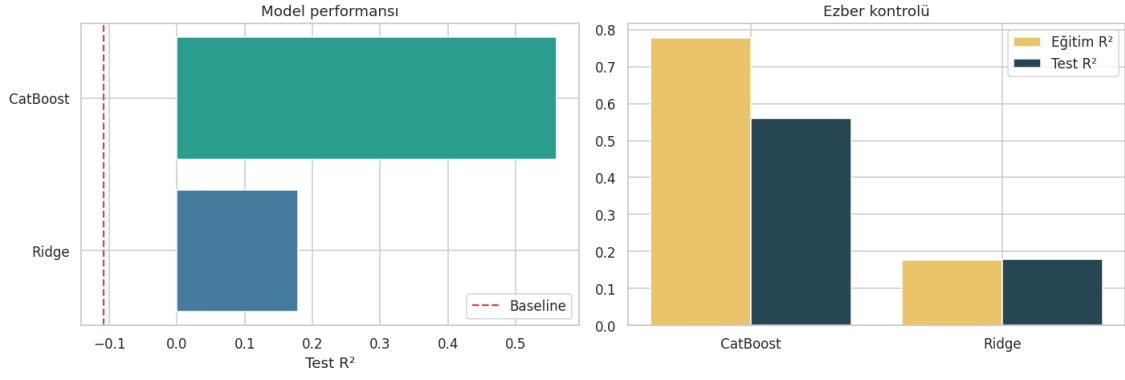
Model	Test R ²	Test RMSE	Test MAE	Eğitim R ²
CatBoost	0,560	1,912	1,500	0,778
Ridge Regression	0,179	2,612	2,012	0,177
Baseline (ortalama tahmin)	-0,108	-	-	-

5 katlı zaman serisi çapraz doğrulaması sonuçları:

Model	Ortalama R ² (CV)	En kötü fold
CatBoost	+0,545 ± 0,062	-
Ridge Regression	-0,142 ± 0,584	-1,303

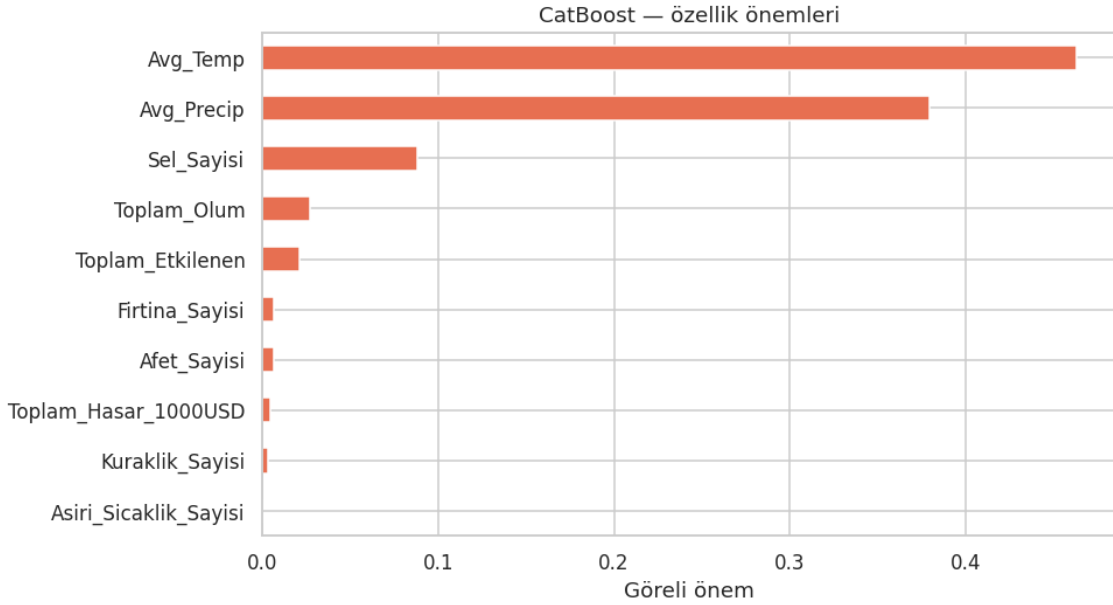


Şekil 7. Hedef değişkenin (Göç) ham ve $\log_{1p}$ dönüşümlü dağılımı (Çalışma 1).



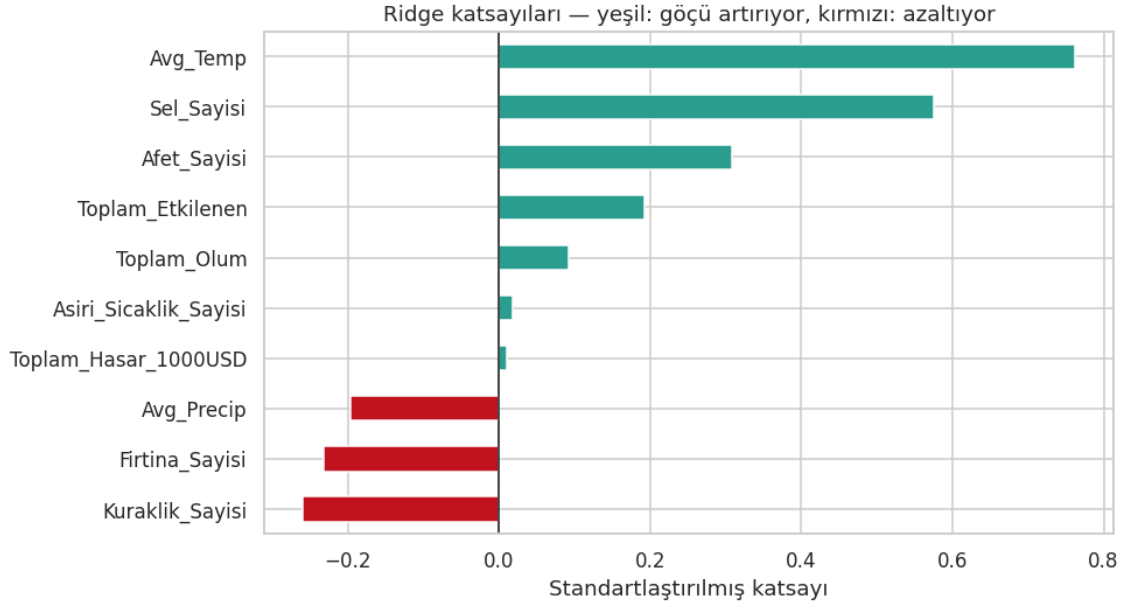
Şekil 8. CatBoost ve Ridge için test R^2 performansı ve eğitim/test R^2 karşılaştırması (ezber kontrolü).

CatBoost, hem tek train/test bölmesinde hem de 5 katlı CV'de Ridge'i açık ara geçmiştir. Bu sonuç, iklim/afet değişkenleri ile göç arasındaki ilişkinin doğrusal olmadığını, eşik değerler ve değişkenler arası etkileşimler içerdiğini göstermektedir (H20 bulgusuyla da tutarlıdır).

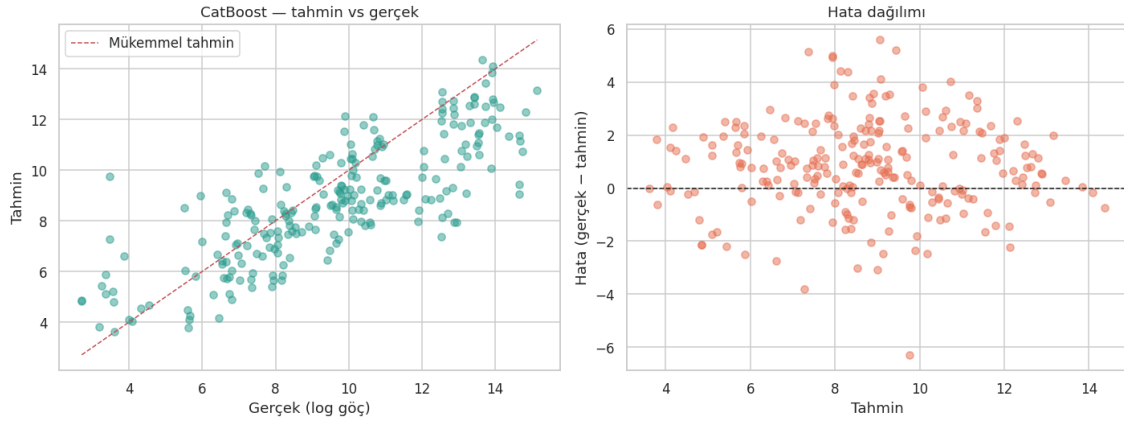


Şekil 9. CatBoost özellik önemleri.

Ridge modelinde kuraklık değişkeninin katsayısının negatif çıkması dikkat çekicidir. Bu durum, H5 bulgusuyla birlikte değerlendirildiğinde "hapsolmuş nüfus" kavramıyla açıklanabilir: aşırı kuraklık bazı durumlarda insanların sınır ötesi göç etme kapasitesini (maddi kaynak, ulaşım imkânı) de azaltabilmektedir.



Şekil 10. Ridge regresyon katsayıları (standartlaştırılmış).



Şekil 11. CatBoost tahmin-gerçek karşılaştırması ve hata dağılımı.

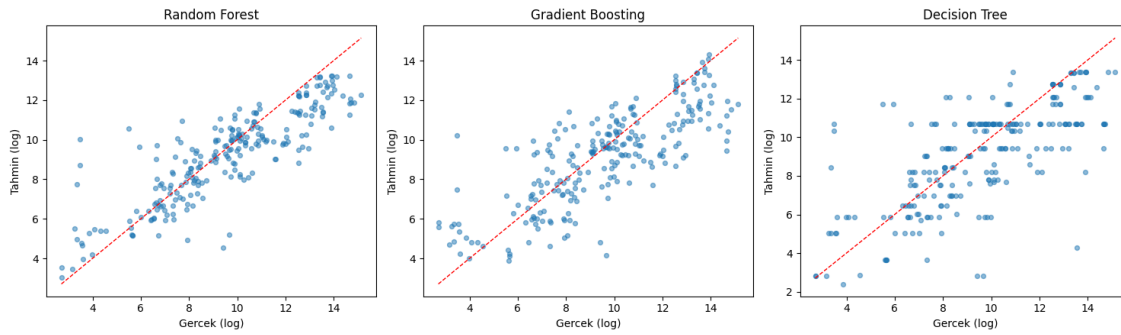
10. Modelleme Çalışması 2: Ağaç Tabanlı Model Karşılaştırması

İkinci modelleme çalışmasında, aynı birlesik_tablo verisi (1.190 satır, 48 ülke, 2001-2025) üzerinde 6 farklı regresyon modeli karşılaştırılmıştır: **Random Forest, Gradient Boosting, Decision Tree, KNN, Ridge ve Linear Regression**. Hedef değişken yine $\log_{1p}(\text{Göç})$ olarak alınmış, aşırı çarpık olan Toplam_Olum, Toplam_Etkilenen ve Toplam_Hasar değişkenleri de log dönüşümüne tabi tutulmuştur.

Eğitim/test ayrımı zaman bazlı yapılmıştır: 2001-2020 arası eğitim (950 satır, %79,8), 2021-2025 arası test (240 satır, %20,2). İlk denemede kullanılan rastgele ayrımın veri sızıntısına yol açtığı tespit edilip düzeltilmiştir - model artık yalnızca geçmiş yıllardan öğrenip henüz görmediği yılları tahmin etmektedir; bu, gerçek dünya senaryosuna daha uygun ve daha güvenilir bir yaklaşımdır.

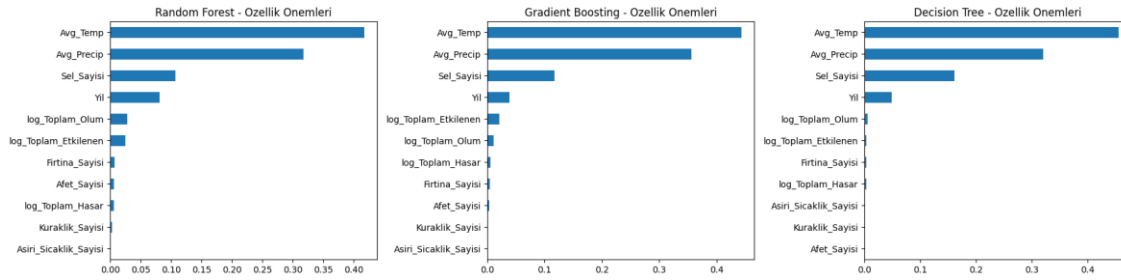
Model	R ²	RMSE (log)	RMSE (kişi)	Sıra
Random Forest	0,7151	1,5388	483.645	1
Gradient Boosting	0,5903	1,8454	504.045	2
Decision Tree	0,4340	2,1689	487.242	3
KNN Regressor	0,3406	2,3411	542.870	4
Ridge Regression	0,2634	2,4744	535.210	5
Linear Regression	0,2634	2,4743	535.164	6

Ağaç tabanlı 3 model (Random Forest, Gradient Boosting, Decision Tree) doğrusal modellerin (Ridge, Linear, R² = 0,26) önüne geçmiştir. Bunun nedeni, afet değişkenlerinin çoğu ülke-yılda sıfır olup bazı yıllarda birden yüksek değerlere sıçramasıdır; bu tür keskin eşikli yapıyı doğrusal modeller yakalayamamaktadır.



Şekil 12. Random Forest, Gradient Boosting ve Decision Tree için tahmin-gerçek karşılaştırması.

Özellik önem analizinde, Random Forest ve Gradient Boosting'in her ikisi de aynı iki değişkene en fazla ağırlığı vermiştir: Avg_Temp (ortalama sıcaklık) ve Avg_Precip (ortalama yağış), toplam önemin yaklaşık %70-75'ini oluşturmaktadır. Sel_Sayisi üçüncü sırada gelmiş, ancak payı çok daha küçük kalmıştır. Dikkat çekici bir bulgu, Afet_Sayisi ve Toplam_Etkilenen gibi "göç ile daha doğrudan ilişkili görünen" değişkenlerin modelde neredeyse hiç ağırlık almamasıdır - yani model esas olarak ülkenin genel iklim karakterinden (sıcak/kurak mı, yağışlı mı) yola çıkmakta, tekil afet olaylarından değil.



Şekil 13. Random Forest, Gradient Boosting ve Decision Tree için özellik önemleri.

En iyi model olan Random Forest ($R^2 = 0,72$), Gradient Boosting'i ($R^2 = 0,59$) hem genel başarı hem de hata payı açısından geçmiştir. Bunun olası nedeni, Random Forest'in bagging yapısı (birbirinden bağımsız ağaçların ortalamasını alma) sayesinde varyansı düşürmesidir; veri seti nispeten küçük olduğu için (950 eğitim satırı) bu durum önemli bir avantaj sağlamaktadır.

11. Sonuç

Bu çalışma, **Doğu-Güney Afrika ve Batı-Orta Afrika bölgelerinde** 2001-2025 yılları arasında **iklim/afet değişkenleri ile sınır ötesi göç arasındaki ilişkiyi** hem hipotez testleri hem de makine öğrenmesi modelleriyle incelemiştir. Temel bulgular şunlardır:

- Tek başına kuraklık, göç miktarını zayıf açıklamaktadır (H1 desteklenmedi); ancak sel gibi ani felaketler kuraklığa göre göç ile daha güçlü ilişkilidir (H7 destekli).
- Kuraklığın etkisi büyük ölçüde ülke içinde kalmakta, sınır ötesi göçe sınırlı yansımaktadır; bu durum "hapsolmuş nüfus" kavramıyla açıklanabilir (H5 kısmi destek).
- Göç, tek bir iklim değişkeniyle değil, birden fazla değişkenin birlikte etkisiyle daha iyi tahmin edilebilmektedir (H20 destekli).
- Her iki bağımsız modelleme çalışmasında da ağaç tabanlı modeller (CatBoost, Random Forest, Gradient Boosting) doğrusal modelleri (Ridge, Linear) belirgin şekilde geçmiştir; en iyi performans Random Forest ile $R^2 = 0,72$, ikinci yaklaşımda CatBoost ile $R^2 = 0,56$ elde edilmiştir.
- Özellik önem analizleri, modelin esas olarak genel iklim karakterinden (ortalama sıcaklık ve yağış) yola çıktığını, tekil afet sayılarının göreceli olarak daha az belirleyici olduğunu göstermiştir.

Genel olarak, iklim ve afet verileri sınır ötesi göçü tek başına güçlü bir şekilde açıklamasa da, çoklu değişken ve doğrusal olmayan modellerle anlamlı ve tutarlı bir tahmin gücü sunmaktadır. Bu da küresel ısınma - göç ilişkisinin basit ve doğrudan değil, dolaylı ve eşik temelli bir yapıda olduğunu düşündürmektedir.

12. Kısıtlar ve Gelecek Çalışmalar

- Kişi bazında hata payı (RMSE) hâlâ yüksektir (~480 bin kişi); bunun temel nedeni, savaş/kriz yaşayan bazı ülkelerin göç sayılarının milyonlara ulaşip dağılımı aşırı çarpık hale getirmesidir. Modelin bu uç değerleri tam olarak yakalayamaması beklenen bir durumdur.
- Veri, 2 bölgedeki 48 ülke ile sınırlıdır; sonuçların diğer bölgelere genellenebilirliği test edilmemiştir.
- Nüfus, kişi başı GSYİH gibi sosyoekonomik değişkenler modele dahil edilmemiştir; bu değişkenlerin eklenmesi tahmin gücünü artırabilir.
- Hiperparametre optimizasyonu (örn. GridSearch) yapılmamıştır; varsayılan parametrelerle çalışılmıştır.

- İç yerinden edilme (IDMC) verisiyle yapılan ayrı keşif çalışması, projenin sınır ötesi göç tanımı gereği ana bulgulara dahil edilmemiştir; ancak iklim kaynaklı iç göç dinamiklerini anlamak için ileride ayrı bir çalışma konusu olabilir.

Proje Ekibi

İsim	Görev	Açıklama
Beyza Fatıma Çekerekli	Takım Lideri	Proje kapsamının ve veri/analiz kararlarının belirlenmesi, tüm ekip çıktılarının tutarlılık kontrolü, literatür taramasına katkı, nihai rapor ve uygulamanın koordinasyonu.
Dilan Sazan	Modelleme (CatBoost & Ridge)	Sınır ötesi göç tahmini için ağaç tabanlı ve doğrusal model karşılaştırması; veri taraması, birleştirme ve literatür taramasına katkı sağladı.
Esmâ Nur Beğbağ	Modelleme (Random Forest / Gradient Boosting / Decision Tree / KNN / Linear)	Çoklu model karşılaştırması ve zaman bazlı doğrulama; veri taramasına katkı sağladı.
Feyza Nur Demirbaş	Veri Toplama, Veri Temizleme ve Birleştirme	Ham göç, iklim ve afet verilerinin derlenmesi, temizlenmesi ve nihai veri setinin oluşturulması.
Hilal Üçüncü	Hipotez Geliştirme	Literatüre dayalı hipotezlerin belirlenmesi ve test edilecek hipotez setinin tasarımı; literatür taramasına katkı sağladı.
Şevval Ok	Keşifsel Veri Analizi (EDA)	Veri setinin dağılım, korelasyon ve örüntü analizleriyle keşfedilmesi; veri taramasına katkı sağladı.

13. Kaynakça

- [1] Cai, R., Feng, S., Oppenheimer, M., & Pytlikova, M. (2016). The migration response to increasing temperatures. *NBER Working Paper*.
- [2] Cattaneo, C., Beine, M., ve ark. (2019). The impact of climate change on migration: a synthesis of recent empirical insights. *Climatic Change, Springer*.
- [3] Bermejo, N., ve ark. (2026). Linking Climate Change to Migration: Insights, Challenges and Future Directions. *IntechOpen*.
- [4] IMF Working Paper. (2024). Climate Variability and Worldwide Migration: Empirical Evidence and Projections (WP/2024/058).
- [5] Journal of Computer Science and Technology Studies. (2025). Big Data and Migration Forecasting: Predictive Insights into Displacement Patterns Triggered by Climate Change and Armed Conflict.
- [6] Marchiori, L., Maystadt, J.-F., & Schumacher, I. (2012). The impact of weather anomalies on migration in sub-Saharan Africa. *ScienceDirect*.
- [7] Sedova, B., & Kalkuhl, M., ve ark. (2024). Climate anomalies and international migration: A disaggregated analysis for West Africa. *ScienceDirect*.