



DENİZ BUZU KAPSAMI VERİ SETİ ÜZERİNE VERİ MADENCİLİĞİ UYGULAMALARI

Hüsna Nur Kaya, *YTÜ, MTM4641*, Veri Madenciliğine Giriş

Özet—Bu proje, deniz buzu kapsamı veri seti üzerine odaklanmaktadır. Amacımız, deniz buzu kapsamını etkileyen faktörleri anlamak ve gelecekteki kapsamı tahmin etmektir.

I. GİRİŞ

Sorun/Motivasyonun Tanıtılması: Deniz buzu kapsamı, iklim değişikliğinin etkilerini anlamak ve bu konuda önlemler almak açısından kritik bir gösterge olarak önem taşımaktadır. Bu çalışma, deniz buzu kapsamındaki değişimleri analiz etmeyi, deniz buzu kapsamını etkileyen faktörlerin önem derecesinin saptanmasını, gelecekteki durumun tahmin edilmesini, penguen popülasyonları için 2100 yılı atfının doğrulanmasını amaçlamaktadır.

Veri Setlerinin Tanıtılması: Projede kullanılan ana veri seti Kaggle platformu üzerinden erişilebilir olan [Daily Sea Ice Extent](#) isimli settir. Bu veri seti küre ve tarih bazlı deniz buzu kapsamının değerlerini içermektedir. Kullanılan bir diğer veri seti [Climate Change Survey](#) veri setidir. Bu set içerisinde yalnızca CEDS, "Community Emissions Data System" (Topluluk Emisyon Veri Sistemi) üzerinden yayınlanan global karbondioksit emisyonlarını gösteren veri seti, total tarım alanı ve endüstriyel fosil veri setleri kullanmıştır. Kullanılan bir diğer set Global Yeryüzü ve Okyanus Sıcaklıkları veri setidir. Son olarak penguen popülasyonları için GitHub üzerinden açık kaynak kodlu bu projedeki set alınmıştır.

II. UYGULAMA

Uygulama aşamasında öncelikle veri ön işleme aşamasında uygun bir şekilde setler birleştirilmiştir. Sütunların transpozunun alınması, merge ve concat işlemleri, sütun isimlerinin değiştirilmesi gibi aşamalardan geçilmiştir.

Daha sonra veri setinin görselleştirilerek daha anlaşılabilir hale gelmesi hedeflenmiştir.

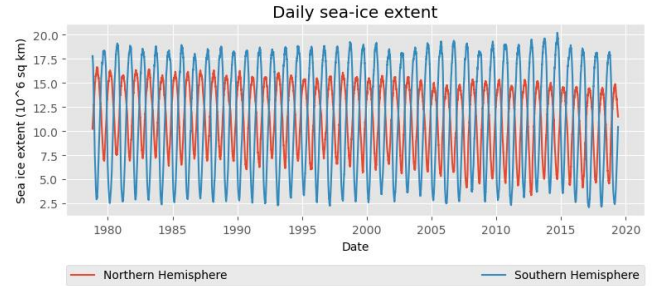


Figure 1: Deniz Buzu Kapsamının Yıl ve Yarım Kürelere Bağlı Değişimi

Bu çalışmada, deniz buzu kapsamının gelecekteki durumunu tahmin etmek amacıyla Random Forest Regressor modeli tercih edildi. Bu modelin seçimindeki önemli avantajlar arasında, her bir öznelik için önem değerini ölçerek etkili özellikleri belirleme kabiliyeti, outlier'ları etkili bir şekilde ele alma yeteneği, karmaşık hiperparametre ayarı yapma ihtiyacının düşük olması ve hem küçük hem de büyük veri setleri üzerinde etkili performans sergileme esnekliği bulunmaktadır. Bu özellikler, modelin güvenilir tahminler yapma kapasitesini artırırken, veri setinin boyutuna uyum sağlama konusundaki esnekliği ise özellikle projenin başlangıcında geniş bir veri seti ile çalışma ihtiyacını karşılamaya yardımcı oldu.

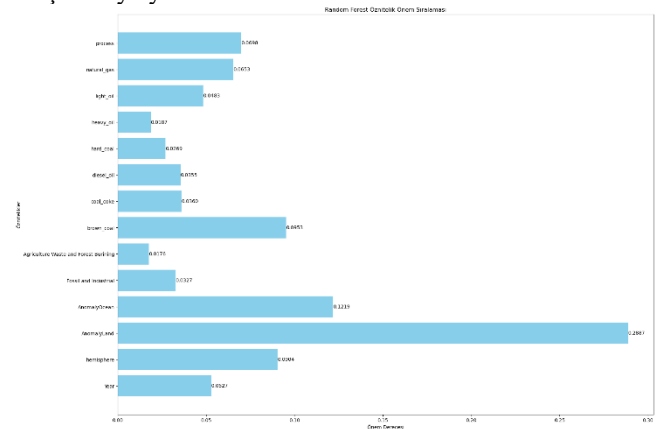


Figure 2: Random Forest Regressor, Feature Importance

Şekil-2’de de gösterildiği üzere faktörler arasında en önemlileri yeryüzü ve okyanus anomalileridir.

1. Anomaly Land
2. Anomaly Ocean
3. brown_coal
4. hemisphere
5. process
6. natural_gas
7. year

Random Forest Regressor modelinin test verisi üzerinde çalıştırılması ile elde edilen metrikler aşağıdaki gibidir:

	Metric	Value
0	MSE	0.035440
1	MAE	0.131430
2	RMSE	0.188256
3	R2 Score	0.895171

Figure 3 Random Forest Regressor Evaluate Metrics

Modelin performansını test etmek için yukarıdaki metriklerin kullanılma sebebi, metriklerin çalışma prensibi ve RFR modelinin performans değerlendirilmesi aşağıdaki gibidir:

- **MSE (Mean Squared Error):** Bu metrik, modelin tahminlerinin gerçek değerlerden ne kadar uzak olduğunu ölçer. MSE'nin düşük olması, modelin iyi bir tahmin yaptığını gösterir. Örneğinizde, MSE değeri 0.035440 olduğu için modelinizin ortalama karesel hatası oldukça düşüktür.
- **MAE (Mean Absolute Error):** MAE, modelin tahminlerinin gerçek değerlere olan mutlak farklarının ortalamasını ifade eder. MAE'nin düşük olması, modelin genel olarak iyi tahminler yaptığını gösterir. Örneğinizde, MAE değeri 0.131430 olduğu için modelinizin ortalama mutlak hatası kabul edilebilir bir seviyededir.
- **RMSE (Root Mean Squared Error):** RMSE, MSE'nin karekökünü alarak büyük hataları daha fazla vurgular ve birimleri orijinal veri birimleriyle aynı hale getirir. RMSE'nin düşük olması, modelin iyi tahminler yaptığını gösterir. Örneğinizde, RMSE değeri 0.188256 olduğu için modelinizin karesel hatası oldukça düşüktür.
- **R2 Score (Coefficient of Determination):** R2, modelin bağımsız değişkenlerce açıklanan varyansın toplam varyansa oranını ölçer. R2'nin 1'e yaklaşması, modelin verilerdeki varyansın büyük bir kısmını açıkladığını gösterir. Örneğinizde, R2 Score değeri 0.895171 olduğu için modelinizin veri setinizdeki varyansın büyük bir kısmını başarılı bir şekilde açıkladığı söylenebilir.

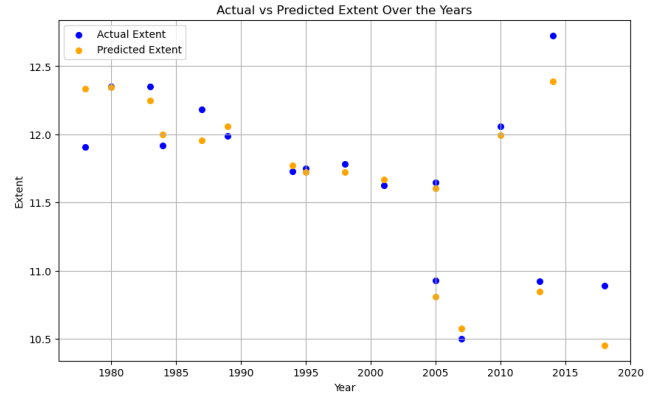
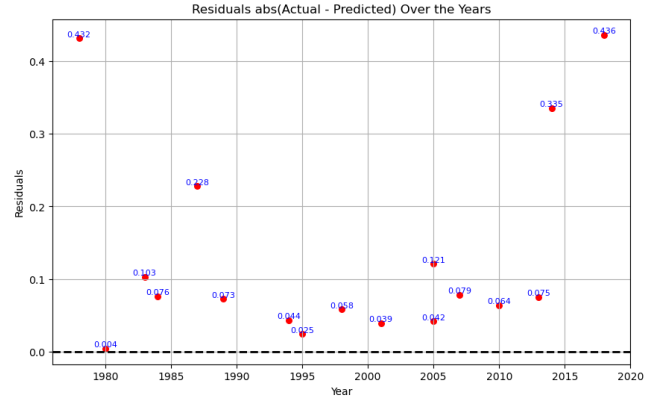


Figure 4: RFR Y_{pred} vs Y_{actual}

Yukarıda gerçek veriler ile model çıktılarının, aşağıda ise hata değerlerinin görselleştirilmesi verilmiştir.



Uygulamanın ikinci kısmında penguen popülasyonları ve deniz buzu kapsamının görselleştirilmesine odaklanılmıştır.

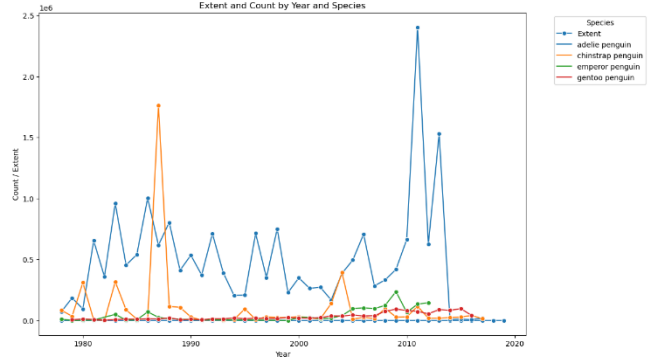


Figure 5: Extent ve Penguen Pop.

Penguen popülasyonu veri seti adeliye, gentoo, chinstrap, emperor türlerinin konum ve saptanan sayılarını içermektedir. Projede penguen popülasyonuna odaklanılmasının sebebi özellikle imparator penguenlerinin IUCN 2020 Kırmızı Liste kategorisinde yani tehdit altında bir tür olmasıdır [1]. 2016'dan bu yana Antarktika, 2021/22 ve 2022/23'teki en düşük iki yılla birlikte, 45 yıllık uydu kayıtlarındaki en düşük deniz buzu genişliğine sahip dört yılı gördü. 2018 ve 2022 yılları arasında, Antarktika'daki bilinen 62 imparator penguen kolonisinin %30'u kısmi veya toplam deniz buzu kaybından

etkilenmiştir.[1]



Fakat yeterli seviyede imparator penguen popülasyonu verisi bulunmadığından setteki tüm türler ve count değerleri kullanılmıştır.

Uygulamanın üçüncü kısmında kuzey ve güney yarım küredeki mevsimsellik etkilerinin birbirinin tam zıttı olması veri görselleştirme ile doğrulanmaya çalışılmıştır.

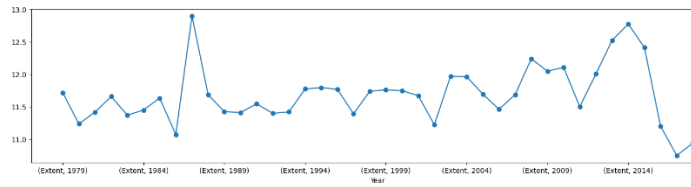


Figure 6: Güney Yarım Küre, Yıl Bazlı

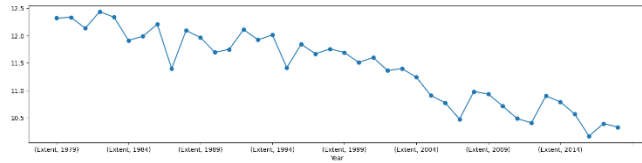


Figure 7: Kuzey Yarım Küre, Yıl Bazlı

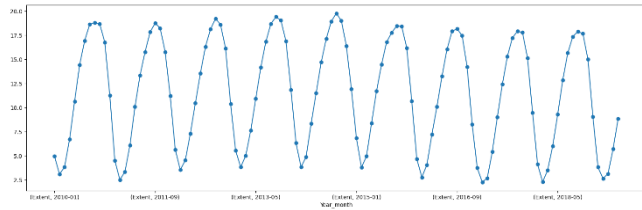


Figure 8: Güney Yarım Küre, Ay Bazlı

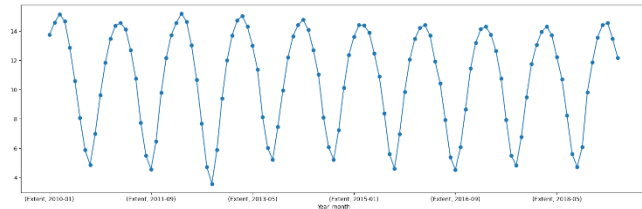


Figure 9: Kuzey Yarım Küre, Ay Bazlı

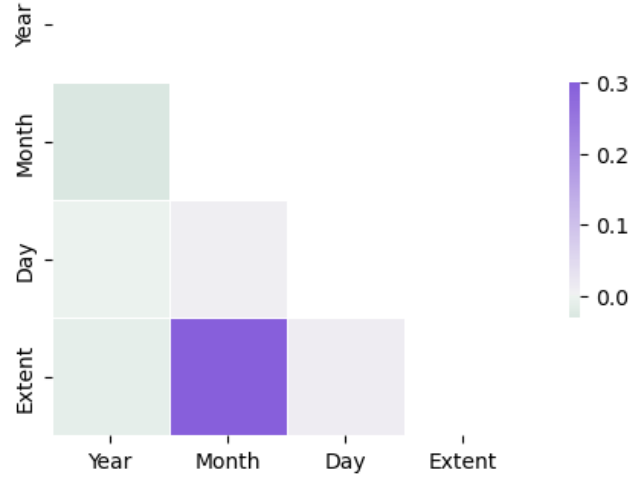


Figure 10: Mevsimsellik

Ayrıca deniz buzu kapsamı ve gün, ay, yıl değerleri arasında yapılan korelasyon matrisi ile tarih belirteçlerinden en önemli faktörün ay olduğu saptanmıştır. Bu sonuç ‘mevsimsellik’ hatta zaman serilerinde ‘seasonality’ olarak çokça kullanılan kavrama da uymaktadır.

Sonuç olarak bu kısımda aylık rakamları özetleyen bir grafikler oluşturuldu. Bu grafiklere göre kuzey için en büyük deniz buzu miktarı Mart ayında, en küçük ise Eylül ayında görülürken güney için tam tersi görülüyor. Bu da aslında yarım kürelerin mevsimselliğine uygun bir çıktıdır. Kuzeyde deniz buzu kapsamında 2011-2018 ortalamaları; Mart için 14,6 milyon kilometrekare, Eylül için 4,6 milyon kilometrekaredir. Aradaki fark 10 milyon kilometrekare, yani Kanada'nın yüzölçümü kadardır.

Uygulamanın son kısmında Linear Regression ile 2060 yılına dek tahminleme yani ‘forecast’ yapılması hedeflenmiştir. Burada amaç hem gelecek tahminlemesi yapmak hem de penguen popülasyonlarına odaklanmaktır. **Arjantin Antarktika Enstitüsü’nde** görevli araştırmacılara göre, imparator penguenlerin nesli, iklim değişikliğinin oluşturduğu olumsuz etkiler nedeniyle 2100’e dek yok olacaklar. Burada deniz buzu kapsamının etkisi de oldukça baskındır. Penguenler Ocak-Nisan ayları arası üreme dönemine sahiptirler. Bu dönemlerde deniz buzu kapsamının düşmesini penguenlerin nesillerini devam ettirebilmeleri karşısında büyük bir dezavantajdır. Ayrıca imparator penguenler çevresinde yapılan çalışmalarda popülasyonlarındaki tek ve ani düşüşün keşfedildiği zaman dilimi ile Antartika bölgesinde deniz buzu kapsamının ani düşüşü arasındaki güçlü korelasyon saptanmıştır.

“NSIDC'nin NASA tarafından finanse edilen Arktik Deniz Buzu Haberleri ve Analizi (ASINA) projesi, Arktik deniz buzunun nasıl değiştiğine ve buz davranışında hangi koşulların rol oynayabileceğine dair neredeyse gerçek zamanlı veriler ve aylık bilgiler, analizler sağlıyor. NSIDC'ye göre iklim değişikliği Kuzey Kutbu'nu dünyadaki diğer yerlerden daha fazla etkiliyor. Bu bölge dünyanın geri kalanından iki kat daha hızlı ısınıyor ve bunun ciddi sonuçlarından biri de önemli miktarda deniz buzunun kaybı. Deniz buzı kaybı hem Arktik ekosistemleri hem de bir bütün olarak Dünya'yı etkiliyor. Arktik deniz buzı açık renkli olduğundan, deniz buzı yüzeyine çarpan güneş ışığının çoğunu uzaya geri yansıtıyor. Bu, okyanusa çok fazla ısınmayı önler ve bölgenin serin kalmasına yardımcı olur. Ancak daha fazla deniz buzı kaybedildikçe daha fazla ısı emilir ve bu da daha fazla erimeye neden olur. Bu ısınmayı artırır ve döngü devam eder. Sonuç olarak deniz buzı Arktik ortamın en hızlı değişen alanlarından biridir. Ki deniz buzı, Dünya ikliminin düzenlenmesinde kritik bir rol oynuyor ve küresel hava durumunu ve okyanus sirkülasyonlarını etkiliyor.[2] Aşağıdaki görsellerde bu ifadenin Kuzey-Güney için yapılan gelecek tahminlemleri ile doğrulanması hedeflenmiştir.. Şekil-11-12'de de görüldüğü üzere aslında güney için durum kuzeydeki kadar vahim durmamaktadır, yıl bazlı analiz yanıltıcıdır.

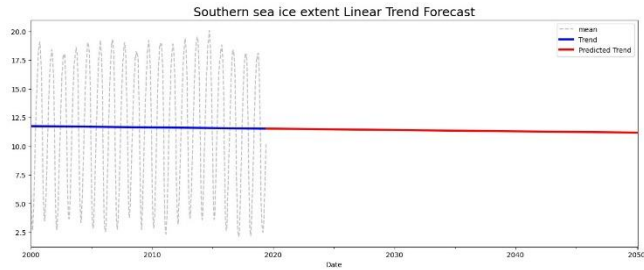


Figure 11:GYK Deniz Buzu Kapsamı-2060

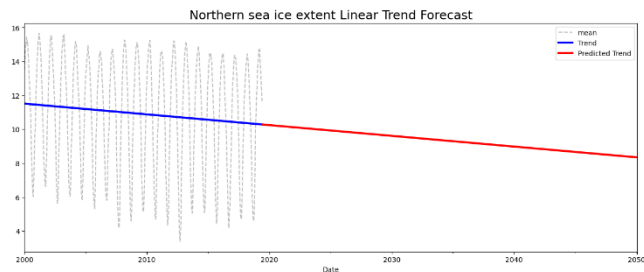


Figure 12:KYK Deniz Buzu Kapsamı-2060

Bir önceki pasajda Kuzey Yarımküre'deki durumun neden daha elzem olduğu açıklanmıştır. Fakat bu noktada Güney Yarımküre'deki yıllık analizin veya tahminlemenin[Şekil12] yanıltıcı olabileceği gözden kaçırılmamalıdır. Çünkü gerek güneydeki penguen popülasyonları gerek kalan habitatın buz kapsamından kuzeydeki kutup ayıları kadar etkilendiği araştırmalarla ortaya konmuştur. Projenin son kısmında amaç penguenlerin üreme dönemi olarak belirtilen Ocak-Nisan dönemleri için 2060'a dek gelecek tahminlemesinin yapılmasıdır. Burada ilk motivasyon yıllık gelecek tahminlemesinin yanıltıcı olabileceğini göstermek, ikinci amaç 2100'e dek yok olacağı söylenen penguen popülasyonları için deniz buzı kapsamının ne denli azaldığını hatta sıfıra yaklaştığını göstermektir.

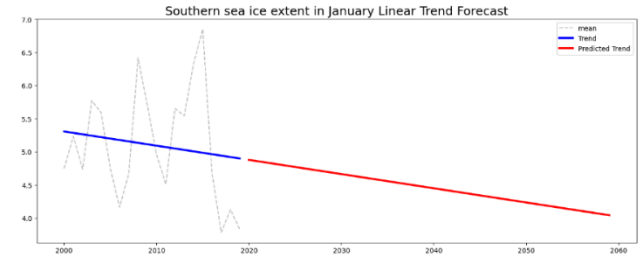


Figure 13:Ocak Ayları 2060'a dek forecast

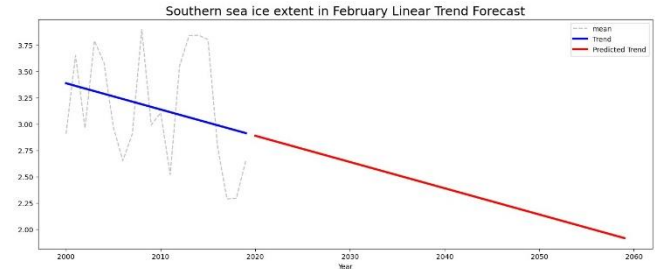


Figure 14:Şubat Ayları 2060'a dek forecast

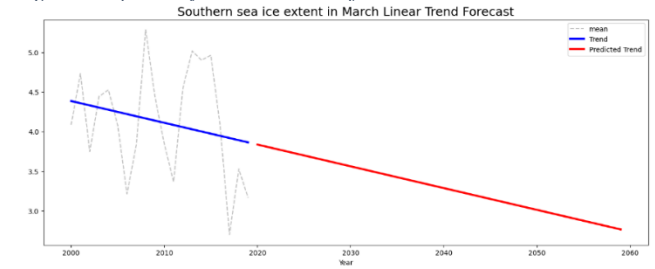


Figure 15:Mart ayları 2060'a dek forecast

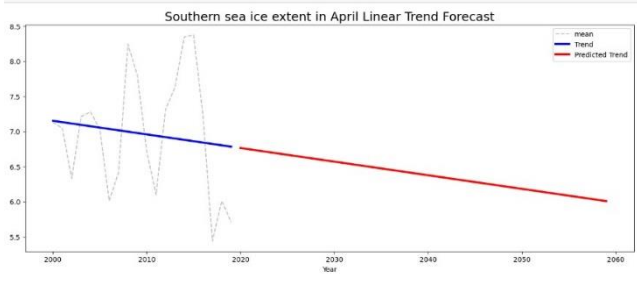


Figure 16: Nisan ayları 2060'a dek forecast

REFERENCES

- [1] [HTTPS://EBIRD.ORG/SPECIES/EMPPEN1?SITELANGUAGE=TR](https://EBIRD.ORG/SPECIES/EMPPEN1?SITELANGUAGE=TR)
- [2] <https://nsidc.org/our-research/featured-projects/arctic-sea-ice-news-analysis>