**Orman Tipi Sınıflandırması: UCI Forest Type Mapping Veri Seti Üzerine Makine Öğrenmesi Uygulaması**

*Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye*

*Ayşe AKDOĞAN 23181616011* [*ayse.akdogan@gazi.edu.tr*](mailto:ayse.akdogan@gazi.edu.tr)

*İrem CİBELEK 23181616024* [*irem.cibelek@gazi.edu.tr*](mailto:irem.cibelek@gazi.edu.tr)

*Şevvalsu AKTAŞ 23181616051* [*sevvalsu.aktas@gazi.edu.tr*](mailto:sevvalsu.aktas@gazi.edu.tr)

**ÖZET**

Orman ekosistemlerinin sürdürülebilir yönetimi; çevresel planlama süreçlerinde doğru, güncel ve güvenilir bilgiye duyulan ihtiyacın artmasıyla birlikte daha da kritik hale gelmektedir. Bu sebeple, orman tipi sınıflandırması; arazi kullanım planlamasından yangın riski yönetimine, biyoçeşitliliğin korunmasından ekosistem hizmetlerinin değerlendirilmesine kadar pek çok alanda önemli ve kritik bir rol üstlenmektedir.

Eski ve gelenekselleşmiş yöntemlerin yerini almaya başlayan makine öğrenmesi tabanlı otomatik sınıflandırma teknikleri, bu süreci daha hızlı, verimli ve geniş ölçekli analizlere uygun hale getirmektedir. Bu çalışmada, Japonya'nın ormanlık bir bölgesine ait çok zamanlı **ASTER uydu görüntülerinden** türetilmiş ve **UCI Machine Learning Repository** tarafından sağlanan **Forest Type Mapping veri seti** kullanılarak farklı makine öğrenmesi algoritmaları ile orman tipi sınıflandırması gerçekleştirilmiştir. Veri seti; spektral ve topoğrafik özellikleri yansıtan toplam **27 öznitelik** ve **dört farklı sınıf** etiketinden oluşmaktadır: **Sugi, Hinoki, yaprak döken karışık orman** ve **orman dışı alan**. Bu çalışmada beş farklı makine öğrenmesi algoritması uygulanmıştır. Bunlar; **K-En Yakın Komşu (k-NN)**, **Karar Ağaçları**, **Destek Vektör Makineleri (SVM)**, **Rastgele Orman (Random Forest)** ve **Lojistik Regresyondur**. Algoritmaların sınıflandırma performansları karşılaştırıldığında, **%85 doğruluk oranı** ile **k-NN ve Destek Vektör Makineleri** en başarılı sonuçları vermiştir. Bu durum orman tipi sınıflandırması gibi karmaşık çevresel problemler için oldukça etkili olduğunu ortaya koymaktadır. En düşük başarı ise Karar Ağaçları ve Rastgele Orman algoritma yöntemlerinde elde edilmiştir.

Elde edilen sonuçlarlarla makine öğrenmesi tabanlı modellerin sadece akademik çalışmalarda değil, aynı zamanda uygulamalı ormancılık faaliyetlerinde de önemli katkılar sağlayabileceği anlaşılmaktadır. Özellikle orman planlaması, ağaç türü sebepli yangın riski değerlendirmesi ve doğa turizmi ile kamp alanlarının mekansal planlaması gibi alanlarda, bu tür sınıflandırmalar karar verme süreçlerini destekleyici niteliktedir.

**Anahtar Kelimeler:** Makine Öğrenmesi, Orman Tipi Sınıflandırması, Veri Madenciliği, UCI Forest Type Mapping

1. **Giriş**

Orman ekosistemleri; dünyadaki biyolojik çeşitliliğin korunmasında, karbon döngüsünün düzenlenmesinde ve iklim değişikliğiyle mücadelede hayati bir rol üstlenmektedir. Bu sistemler yalnızca doğal yaşamı desteklemekle kalmayıp aynı zamanda insanlar için önemli ekosistem hizmetleri de sunar: karbon depolama, su kaynaklarının korunması, erozyonun önlenmesi ve mikro iklim düzenlenmesi bunlardan bazılarıdır. Bu nedenle, ormanların türlerine göre doğru bir şekilde sınıflandırılması ve haritalanması; sürdürülebilir arazi yönetimi, doğal kaynakların korunması, yangın riski analizi ve biyoçeşitlilik değerlendirmesi gibi pek çok çevresel planlama süreci açısından kritik öneme sahiptir.

Geleneksel yöntemlerle yapılan orman sınıflandırmaları genellikle zaman alıcı, maliyetli ve geniş alanlar için uygulanması güç yöntemlerdir. Günümüzde uzaktan algılama teknolojilerinin gelişmesi ve yüksek çözünürlüklü uydu görüntülerinin elde edilebilmesiyle birlikte, bu alanda makine öğrenmesi tabanlı otomatik sınıflandırma sistemleri öne çıkmaktadır. Bu sistemler, özellikle spektral veriler üzerinden orman örtüsünün detaylı analizine olanak tanıyarak daha hızlı ve etkili çözümler sunar.

Bu çalışma, UCI Machine Learning Repository'de yer alan Forest Type Mapping veri setini temel alarak farklı orman tiplerinin sınıflandırılmasını hedeflemektedir. Veri seti, Japonya'daki bir ormanlık alandan üç farklı tarihte elde edilen ASTER uydu görüntülerine dayalı olarak hazırlanmıştır ve dört ana sınıfı içermektedir: Sugi (Cryptomeria japonica), Hinoki (Chamaecyparis obtusa), karışık yaprak döken orman ve orman dışı alanlar. Veri setinde, her bir gözlem için farklı tarihlere ait dokuz spektral bantın ölçümleri ile mekansal enterpolasyonla elde edilen tahmin-gözlem farkları yer almaktadır.

Çalışmada, veri seti üzerinde çeşitli makine öğrenmesi algoritmaları (k-NN, Karar Ağaçları, SVM, Rastgele Orman, Lojistik Regresyon) uygulanmış; bu modellerin sınıflandırma doğrulukları karşılaştırılmış ve orman tipi haritalamada otomatik sistemlerin etkinliği değerlendirilmiştir. Bu sayede, çevresel analizlerde makine öğrenmesinin potansiyeli ortaya konularak, gelecekteki sürdürülebilir ormancılık uygulamaları için güçlü bir temel oluşturulması amaçlanmaktadır.

**2. Materyal ve Yöntem**

**2.1. Veri Seti**

Bu çalışmada kullanılan veri seti, **UCI Machine Learning Repository** üzerinden sağlanan ve Japonya’daki ormanlık bir alandan elde edilen **çok zamanlı uzaktan algılama verilerine** dayalı **Forest Type Mapping** veri setidir. Veri seti, Japonya’nın Tottori bölgesinde bulunan ormanlık alanlara ait 2010–2011 yıllarında üç farklı tarihte (26 Eylül 2010, 19 Mart 2011, 8 Mayıs 2011) alınan **ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer)** uydu görüntülerinden oluşturulmuştur.

Veri seti, her biri **27 sayısal öznitelikten** oluşan toplam **326 örnek (piksel)** içermektedir. Öznitelikler; yeşil, kırmızı ve yakın kızılötesi (NIR) spektral bantlara ait yansıma değerleri ile bu değerler arasındaki farkları temsil etmektedir. Her bir örnek, **dört orman türünden birine** ait olarak etiketlenmiştir.

Orman sınıfları şunlardır:

1. **s**: Sugi ormanı (Cryptomeria japonica)
2. **h**: Hinoki ormanı (Chamaecyparis obtuse)
3. **d**: Yaprak döken karışık orman
4. **o**: Diğer (orman dışı alanlar)

Veri setinde eksik değer bulunmamaktadır. Sınıf dengesi gözetilerek veriler hem eğitim hem test aşamasında eşit dağılımla kullanılmak üzere bölünmüştür.

Bu veri seti, uzaktan algılama temelli orman sınıflandırması üzerine makine öğrenmesi algoritmalarının performansını karşılaştırmak için oldukça uygundur. Çalışmada kullanılan sınıflandırma algoritmaları dört sınıfı en yüksek doğrulukla tahmin etmeyi hedeflemektedir.

**2.2. Kullanılan Makine Öğrenmesi Algoritmaları**

Makine öğrenmesi, yapay zeka alanının bir alt dalı olup; bilgisayarların doğrudan programlanmadan, matematiksel işlemler yoluyla verilerden öğrenmesini sağlayan bir süreçtir. Bu yöntem sayesinde, bilgisayarlar işledikleri verilerden elde ettikleri deneyimlerle karmaşık problemlere yönelik tahminlerde bulunabilirler. Makine öğrenmesi; doğal dil işleme, çevrimiçi arama, öneri sistemleri, akıllı kart uygulamaları ve veri güvenliği gibi pek çok farklı alanda etkin biçimde kullanılmaktadır.

### Makine öğrenmesi, yapay zekanın temel alt alanlarından biridir ve verilerden öğrenerek tahmin ve sınıflandırma yapabilen modeller geliştirmeyi amaçlamaktadır. Günümüzde bu yöntem, çevresel verilerin analizi, arazi örtüsünün sınıflandırılması, orman haritalarının oluşturulması, görüntü işleme ve uzaktan algılama gibi birçok farklı alanda yaygın biçimde kullanılmaktadır. Bu çalışma kapsamında, Japonya’ya ait çok zamanlı spektral verilerden oluşan Forest Type Mapping veri seti üzerinde, farklı orman türlerinin (s, h, d, o) sınıflandırılması için çeşitli makine öğrenmesi algoritmaları kullanılmıştır. **Bu çalışmada, aşağıda açıklanan makine öğrenme yöntemleri kullanılmıştır.**

* Rastgele Orman (Random Forest): Denetimli öğrenmeye dayalı, birden fazla karar ağacının rastgele alt kümeler üzerinde eğitilerek oluşturduğu topluluk modelidir. Her bir ağaç bağımsız çalışır ve sonuçlar oylanarak nihai tahmin yapılır.
* Karar Ağacı (Decision Tree): Her düğümde bir özellik seçerek, veri kümesini sınıflara göre ayırır. Ağacın dalları karar kurallarını temsil ederken, yaprak düğümler sınıf etiketlerine karşılık gelir.
* K-En Yakın Komşu (k-NN): Her bir veriyi öklidyen mesafe gibi ölçütlere göre komşularına bakarak sınıflandırır.
* Lojistik Regresyon (Logistic Regression): Lojistik regresyon, bir gözlemin belirli bir sınıfa ait olma olasılığını tahmin eder. Bu olasılık, bağımsız değişkenlerin (özelliklerin) doğrusal birleşimi ile hesaplanır.
* SVM (Support Vector Machine / Destek Vektör Makineleri): Denetimli bir makine öğrenmesi algoritmasıdır ve **sınıflandırma** ile **regresyon** problemlerinde kullanılır. En yaygın kullanım alanı sınıflandırmadır. SVM'nin temel amacı, farklı sınıfları birbirinden maksimum marjla ayıran en iyi doğrusal (veya doğrusal olmayan) sınırı (hiperdüzlem) bulmaktır.

Makine öğrenmesiyle sınıflandırma iki temel aşamada gerçekleştirilmiştir:

1. Öncelikle veri seti uygun biçimde ön işleme tabi tutulmuştur (ölçekleme ve sınıf etiketi kodlaması gibi).
2. Ardından Python ortamında geliştirilen modeller ile dört sınıfa (s, h, d, o) ait orman türleri tahmin edilmiştir.

Modellerin başarımı doğruluk (accuracy) metriği ile ölçülmüştür. Elde edilen sonuçlar bulgular kısmında verilmiş, hangi algoritmanın veri seti üzerinde daha iyi performans gösterdiği karşılaştırmalı olarak değerlendirilmiştir.

**2.3. Literatür Taraması**

Watanabe ve arkadaşları 2020 yılında, Japonya’daki orman türlerini uzaktan algılama teknikleriyle sınıflandırmak amacıyla çok zamanlı ASTER uydu görüntülerini kullanarak bir çalışma gerçekleştirmiştir. Çalışmada "Sugi", "Hinoki" ve "karışık yaprak döken" gibi türlerin ayrımı için spektral ve mekansal bilgilerden yararlanılmıştır ve kullanılan rastgele orman, destek vektör makineleri yöntemleriyle %87’ye varan sınıflandırma başarı oranı elde etmişlerdir. Bu çalışma, yüksek doğruluk oranlarıyla orman yönetimi ve karbon stoklarının izlenmesine katkı sağlamaktadır. [1]

Matsuoka ve Kubo 2021 yılında yaptıkları çalışmada, Sentinel-2 ve Landsat 8 uydu görüntülerini birleştirerek Japonya'nın orman örtüsünü yüksek çözünürlükte sınıflandırmayı hedeflemiştir. Makine öğrenmesi teknikleri kullanılarak yapılan bu çalışmada en yüksek sınıflandırma doğruluğu %89 ile Landsat 8 verilerinde elde edilmiş, Sentinel-2 verilerinin ise %77 doğruluk sağladığı rapor edilmiştir. Bu sonuç, çok kaynaklı veri kullanımının sınıflandırma başarısını artırdığını göstermektedir. [2]

Tanaka ve arkadaşları 2018 yılında yaptıkları çalışmada, Japonya’nın ulusal düzeyde 30 m çözünürlüklü bir arazi örtüsü haritasını oluşturmak amacıyla Rastgele Orman algoritmasıyla çalışmışlardır. Çalışmada; orman türleri yanında tarım alanları, su yüzeyleri ve yapılı çevre gibi sınıflar da analiz edilmiştir. Yapay zeka destekli bu sınıflandırma, ulusal düzeyde sürdürülebilir arazi kullanım politikalarının geliştirilmesine zemin hazırlamaktadır. [3]

Son olarak, Fujimoto ve arkadaşları 2023 yılında yaptıkları çalışmada Japonya’da orman karbon stoklarının belirlenmesi amacıyla LiDAR ve Sentinel-2 verilerini birleştirerek yüksek doğrulukta biyokütle tahminleri gerçekleştirmiştir. Bu çok kaynaklı yaklaşım, ormanların mekansal yapısını ayrıntılı olarak ortaya koymuş ve iklim değişikliği ile mücadele kapsamında veri tabanlı karar destek sistemlerine katkı sağlamıştır. [4]

**3.Bulgular**

Gerçekleştirilen çalışmada beş farklı makine öğrenmesi yöntemi kullanılmıştır. Bu yöntemlerden %85 doğruluk oranıyla k-NN ve Destek Vektör Makineleri en başarılı yöntemler olmuştur. Sonrasında %84 ile Lojistik Regresyon takip etmiştir. En düşük başarı ise Karar Ağaçları ve Rastgele Orman yöntemlerinde elde edilmiştir.

Aşağıda uygulanan algoritmaların doğruluk oranları tablo olarak sunulmuştur:

Tablo 1. Elde edilen sonuçlar

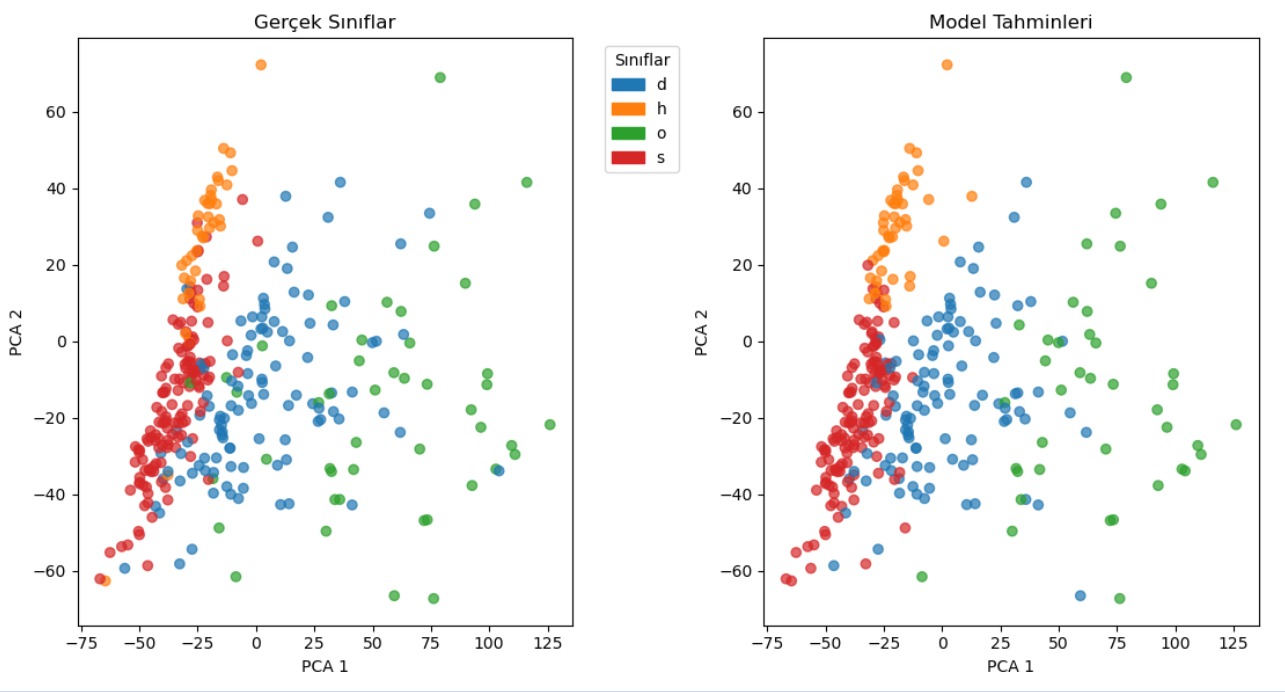
|  |  |
| --- | --- |
| Kullanılan Makine Öğrenmesi Yöntemi | Doğruluk (%) |
| K-En Yakın Komşu (k-NN) | 85 |
| Destek Vektör Makineleri (SVM) | 85 |
| Lojistik Regresyon | 84 |
| Rastgele Orman | 81 |
| Karar Ağaçları | 81 |

Tablo 2. Korelasyon matrisi

metin, kalıp, desen, düzen, mor, ekran görüntüsü içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Tablo 3. k-NN test ve tahmin verileri karşılaştırılması



Tablo 4. Kullanılan makine öğrenmesi yöntemlerinin doğruluklarının karşılaştırılması

metin, ekran görüntüsü, dikdörtgen, ekran, görüntüleme içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**4.Sonuç ve Öneriler**

Bu çalışmada, Japonya’da yer alan bir ormanlık bölgeye ait çok zamanlı uydu görüntülerinden yararlanılarak orman türlerinin sınıflandırılması ve haritalanması gerçekleştirilmiştir. Bu doğrultuda, ‘Sugi’, ‘Hinoki’, ‘karışık yaprak döken’ ve ‘diğer (orman dışı alanlar)’ olmak üzere dört farklı orman türünün ayırt edilebilmesi amacıyla çeşitli makine öğrenmesi modelleri kullanılmış ve bu modellerin performansları karşılaştırmalı olarak değerlendirilmiştir. Sınıflandırma sürecinde temel veri kaynağı olarak, spektral bant bilgileri, topografik veriler ve farklı zaman dilimlerinden elde edilen değişim bilgileriyle oluşturulan öznitelikler kullanılmıştır. Analizlerde kullanılan veri seti, 2010 ile 2011 yılları arasında elde edilen çok zamanlı ASTER uydu görüntülerine dayanmaktadır. Toplam 326 örneklemden oluşan veri seti, %70 oranında eğitim, %30 oranında ise test aşamasında kullanılarak modellenmiştir. Elde edilen sonuçlara göre, orman tipi sınıflandırmasında k-NN ve Destek Vektör Makineleri algoritmaları en yüksek doğruluğu sağlamış ve sınıflandırma performansı açısından ön plana çıkmıştır. En düşük doğruluk oranı ise Karar Ağaçları ve Rastgele Orman yöntemleri ile elde edilmiştir. Bu bulgular, çok zamanlı uzaktan algılama verilerinin orman türü haritalamada etkili bir şekilde kullanılabileceğini göstermektedir.

Bu çalışmayla, orman ekosistemlerinin haritalanmasına yönelik sürdürülebilir bir yaklaşım sunulmuş ve orman türlerinin uzaktan algılama ile izlenebilirliği gösterilmiştir. Ayrıca bu çalışma, Japonya’daki çok zamanlı spektral verilerin orman tür ayrımı için uygulanabilirliğini ortaya koyması açısından önemli bir örnek teşkil etmektedir. Gelecekte yapılacak çalışmalarda, daha yüksek çözünürlüklü veriler ve farklı sensör kombinasyonları kullanılarak orman yönetimi ve karbon stok tahminlerine katkı sağlayacak daha hassas sınıflandırmalar yapılabileceği sonucuna varılmıştır.

**KAYNAKÇA**

1. Watanabe, M., Saito, H., & Takahashi, H. (2020). Forest type classification using multi-temporal ASTER imagery and machine learning techniques in Japan. International Journal of Remote Sensing, 41(5), 1853–1871.
2. Matsuoka, M., & Kubo, S. (2021). High-resolution vegetation mapping in Japan by combining Sentinel-2 and Landsat 8 datasets through machine learning and cross-validation. Land, 6(3), 50.
3. Tanaka, Y., Suzuki, H., & Iizuka, K. (2018). A national-scale 30 m land cover map of Japan using Random Forests and remote sensing data. Remote Sensing, 10(2), 322.
4. Fujimoto, Y., Nishida, K., & Sato, M. (2023). High-resolution forest biomass estimation using multi-source remote sensing data in Japan: Integrating LiDAR, Sentinel-2, and ALOS PALSAR-2. Remote Sensing of Environment, 312, 114322.
5. University of California, Irvine. (2015). Forest type mapping data set. UCI Machine Learning Repository.

**GitHub Hesaplarımız:**

İrem Cibelek - https://github.com/iremcibelek/forest-type-mapping

Ayşe Akdoğan - https://github.com/ayseakdogann

Şevvalsu Aktaş - https://github.com/sevvalsuaktas