**Toprak Besinleri ve Çevresel Faktörler Kullanılarak**

**Tarımsal Verimlilik Analizi**

*1. Ahmad Zaim Ilman Mohd Jamal*

*2. Md Shahrear Ahmed Shajib*

**Özet**

Bu analiz, toprak besin maddeleri ile çevresel değişkenler arasındaki çok boyutlu ilişkilerin tarımsal verimlilik üzerindeki etkisini makine öğrenimi teknikleri aracılığıyla kapsamlı şekilde incelemektedir. Proje kapsamında elde edilen veriler, veri ön işleme ve özellik mühendisliği süreçlerinden geçirilerek, çeşitli modellerle analiz edilmiştir. Geliştirilen modeller, toprak ve çevresel faktörlerin mahsul verimi üzerindeki karmaşık etkileşimlerini ortaya koymuş ve tahmin doğruluğu açısından tatmin edici performans sergilemiştir. Elde edilen bulgular, hassas tarım uygulamalarında kullanılmak üzere verimlilik optimizasyonuna yönelik değerli içgörüler sağlamaktadır. Çalışma, veri odaklı yaklaşımların sürdürülebilir tarım politikalarının geliştirilmesinde önemli bir rol oynayabileceğini ortaya koymaktadır. Sonuçlar, tarım sektöründe etkin karar destek mekanizmalarının oluşturulmasına katkı sağlamayı amaçlamaktadır.

**Anahtar Kelimeler:** Tarımsal Verimlilik, Mahsul Verimi Tahmini, Toprak Besinleri, Çevresel Faktörler, Hassas Tarım

**1. Giriş**

Tarım sektörü, dünya genelinde hızla artan nüfus ve değişen tüketim alışkanlıkları nedeniyle giderek büyüyen bir gıda talebiyle karşı karşıyadır. Bu talebi karşılamak, mevcut doğal kaynakların sürdürülebilir biçimde kullanılması ve verimliliğin artırılmasıyla mümkün olabilmektedir. [1] Tarımsal üretkenlik, üretim süreçlerinde kaynakların etkin yönetimiyle aynı veya daha yüksek miktarda ürün elde edilmesi olarak tanımlanır ve bu kavram, günümüzde hem ekonomik hem de çevresel sürdürülebilirlik açısından kritik bir öneme sahiptir. [1] [2] Özellikle iklim değişikliği ve çevresel kısıtlamalar, tarımın geleceği için yeni yaklaşımlar geliştirilmesini zorunlu kılmaktadır.

Modern tarım uygulamalarında, sensör teknolojileri, uydu görüntüleri ve diğer dijital veri kaynaklarının kullanımıyla desteklenen hassas tarım teknikleri giderek yaygınlaşmaktadır. [3][4] Bu teknolojiler, toprak koşulları ve çevresel değişkenlerin daha doğru ve zamanında izlenmesini sağlamaktadır. Makine öğrenimi yöntemleri ise bu büyük ve karmaşık veri setlerinden anlamlı örüntülerin çıkarılması ve mahsul verimliliğini etkileyen faktörlerin modellenmesinde önemli bir araç olarak ortaya çıkmaktadır. [5] [6]

Bu proje, toprak besin maddeleri ve çevresel faktörlerin tarımsal verimlilik üzerindeki etkilerini detaylı şekilde incelemekte ve bu ilişkileri tahmin edebilen makine öğrenimi modelleri geliştirmeyi hedeflemektedir. Böylece, çiftçilere ve tarım sektörüne yönelik karar destek sistemlerinin geliştirilmesine katkı sağlanması amaçlanmaktadır. Raporun ilerleyen bölümlerinde, veri setinin analizi, kullanılan modeller ve elde edilen bulgular ayrıntılı şekilde ele alınacaktır.

**1.1. Tarımsal Verimlilik ve Kaynak Yönetimi**

Tarımsal üretkenlik, tarımsal çıktılar ile girdilerin oranı olarak tanımlanmakta ve bu kavram, özellikle farklı ürün çeşitliliği nedeniyle genel çıktının ölçümünde zorluklar içermektedir. [2] Bu nedenle, tarımsal üretkenlik genellikle nihai çıktının piyasa değeri üzerinden değerlendirilmekte, çıktı bu değerle işgücü veya arazi gibi çeşitli girdilerle karşılaştırılarak kısmi üretkenlik ölçümleri elde edilmektedir. [2] Ancak, kısmi ölçümlerin sınırlılıklarını aşmak amacıyla toplam faktör verimliliği (TFP) yaklaşımı kullanılmakta, bu yöntem tarımsal girdilerin bir endeksini çıktı endeksiyle karşılaştırarak üretkenlikteki değişimleri daha kapsamlı biçimde ortaya koymaktadır. [1] [2] TFP’deki iyileşmeler genellikle teknolojik gelişmelerle ilişkilendiril mektedir; bu durum, birim çıktı başına daha az toplam girdi kullanılarak daha yüksek üretim elde edilmesini ifade etmekte ve böylece tarımsal sürdürülebilirliğe doğrudan katkı sağlamaktadır. [1]

Tarımsal üretkenlik, gıda güvenliği açısından temel bir bileşen olup, üretkenlik arttıkça gıda fiyatlarında düşüş gözlenmekte ve bu durum açlıkla mücadelede önemli bir rol oynamaktadır. [2] Dünya nüfusu artış gösterdikçe, özellikle gelişmekte olan bölgelerde tarımsal üretkenliğin artırılması büyük önem kazanmaktadır. Örneğin, Hindistan’da 1960'lardan itibaren yüksek verimli buğday ve pirinç çeşitlerinin benimsenmesi, hektar başına önemli miktarda tahıl üretimini mümkün kılmıştır. [2] Buna karşılık, Sahra Altı Afrika’da azalan üretkenlik açlık sorununu tetikleyebilmektedir. [2] Kadın çiftçilerin üretkenliğine yatırım yapılması ise ekonomik kalkınma ve gıda güvenliği açısından kritik öneme sahiptir. [2]

Ekonomik büyüme açısından değerlendirildiğinde, tarımsal üretkenlik kıt kaynakların verimli dağıtımını sağlar ve bölgesel rekabetçiliği artırır. [2] Üretkenlik artışı, gıda fiyatlarını düşürürken reel gelirin diğer sektörlerde de yükselmesine zemin hazırlamaktadır. [2] Özellikle nüfusun büyük kısmının tarımda çalıştığı yoksul ülkelerde, üretkenlik artışları yoksulluğun azaltılmasına katkıda bulunmaktadır. [2]

**1.2. Makine Öğrenimi ve Tarımda Uygulamaları**

Hassas tarım teknolojileri 1990'lardan beri var olsa da, günümüzde ABD çiftliklerinin yalnızca %27'si bu uygulamaları kullanmaktadır. [4] Bu düşük benimseme oranının sebepleri arasında yüksek başlangıç maliyetleri, veri paylaşımı ve sahipliği konusundaki endişeler ile standart eksikliği bulunmaktadır. [4] Ayrıca, çiftçilerin teknolojiye uyum sağlamada ve verileri yorumlamada karşılaştıkları güçlükler, üretkenlik artışını engelleyebilmektedir. [3] Aşırı veri yükü ise karar alma süreçlerini zorlaştırabilmektedir. [3]

Buna rağmen, yapay zeka (AI) ve makine öğrenimi (ML) gibi gelişmiş veri analizi teknikleri, tarımsal karar alma süreçlerini dönüştürme potansiyeline sahiptir. [3] [4] Bu teknolojiler, sulama, gübreleme, zararlı kontrolü ve hasat gibi kritik alanlarda çiftlik içi kararların optimize edilmesini sağlar. [3] [4] Kullanılan veri türleri arasında hava durumu modelleri, toprak sağlığı bilgileri (nem, pH, besin içeriği), mahsul sağlığı verileri, pazar eğilimleri ve büyüme/kayıp analizleri bulunmaktadır. Makine öğrenimi modelleri bu verilerden anlamlı örüntüler çıkararak, hassas tarım uygulamalarının ölçekte yaygınlaşmasına ve kaynak kullanım verimliliğinin artırılmasına katkıda bulunmaktadır. [3] [4] Böylece, ekonomik ve çevresel sürdürülebilirlik hedefleri doğrultusunda tarımsal üretkenlik desteklenmektedir.

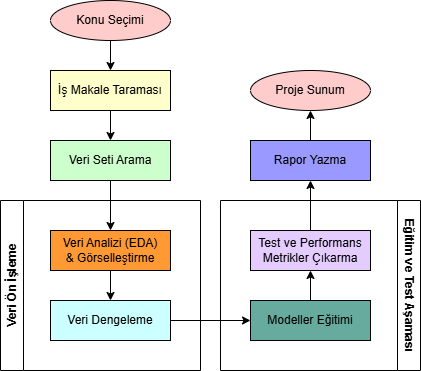
**1.3. Literatür Özeti**

Bu bölümde, Toprak Besinleri ve Çevresel Faktörler Kullanılarak Tarımsal Verimlilik Analizi ile ilgili daha önce yapılan çalışmalara yer verilmiştir:

1. Anand, A., & Jhajharia, K. (2022). "Advancements in Machine Learning and Deep Learning Techniques for Crop Yield Prediction: A Comprehensive Review". Bu kapsamlı incelemede, makine öğrenimi ve derin öğrenme yöntemlerinin tarımsal verim tahmininde kullanımı ele alınmıştır. [5] Çalışma, çevresel faktörlerin etkisini ve bu alandaki çeşitli algoritmaların başarı oranlarını değerlendirmiştir. Random Forest, CatBoost, XGBoost, CNN, DNN, RNN ve LSTM gibi modellerin etkinlikleri ve veri ön işleme teknikleri detaylı biçimde incelenmiştir. [5] [17] [18] [20] Ayrıca, hibrit yaklaşımların ve veri kodlamanın model başarısındaki rolü vurgulanmıştır.
2. Mahesh, P., Soundrapandiyan, R., & Bacanin, N. (2021). "Machine Learning Algorithms for Crop Yield Prediction: A Comparative Analysis of Gradient-Based Approaches". Bu araştırma, pirinç verim tahmini için CatBoost, LightGBM ve XGBoost algoritmalarının performans karşılaştırmasını yapmıştır. [6] Sonuçlar, CatBoost algoritmasının diğerlerine kıyasla daha yüksek doğruluk ve daha düşük hata oranı sunduğunu göstermektedir. [6] Modelin başarısı, karmaşık veri yapılarının işlenmesindeki üstünlüğüne bağlanmıştır. Bu çalışma, gradyan tabanlı algoritmaların tarımsal verim tahmininde önemli bir yer tuttuğunu ortaya koymaktadır. [6]
3. Liu, J., Wang, D., Yan, X., et al. (2023). "Impact of different ratios of nitrogen (N), phosphorus (P), and potassium (K) fertilization on the soil properties, leaf traits, and yield of Sapindus mukorossi". Çalışmada, Çin’in Batı Fujian bölgesinde farklı NPK gübreleme oranlarının toprak özellikleri, yaprak sağlığı ve verim üzerindeki etkileri araştırılmıştır. En yüksek verim N2P2K2 oranında elde edilirken, aşırı gübre kullanımının mahsul verimini olumsuz etkilediği saptanmıştır. Toprak organik karbonu ve yaprak alanı indeksinin verim üzerinde belirleyici faktörler olduğu tespit edilmiştir. Sonuçlar, dengeli gübrelemenin tarımsal üretkenlik için kritik olduğunu göstermektedir. [7]
4. Jorvekar, P. P., Wagh, S. K., & Prasad, J. R. (2020). "Predictive Modeling of Crop Yields: A Comparative Analysis of Regression Techniques for Agricultural Yield Prediction". Bu çalışma, beş farklı regresyon modelinin mahsul verimi tahminindeki başarısını karşılaştırmıştır. Derin öğrenme tabanlı yapay sinir ağları, karmaşık doğrusal olmayan ilişkileri yakalama kabiliyeti ile diğer modellerin önüne geçmiştir. Model performansları detaylı istatistiksel yöntemlerle analiz edilmiş ve derin öğrenmenin üstünlüğü vurgulanmıştır. Sonuçlar, karmaşık tarımsal veri setlerinde derin öğrenmenin tercih edilmesi gerektiğini göstermektedir. [8] [20]
5. Islam, M. M., Alharthi, M., Alkadi, R. S., et al. (2022). "Crop yield prediction through machine learning: A path towards sustainable agriculture and climate resilience in Saudi Arabia". [9] Bu çalışma, Suudi Arabistan’daki mahsul verimini iklim değişikliği perspektifiyle makine öğrenimi yöntemleri kullanarak tahmin etmiştir. XGBoost algoritması, tahmin doğruluğunda en yüksek performansı sergilemiştir. [9] Elde edilen modeller, sürdürülebilir tarım ve iklim değişikliğine uyum stratejilerinin geliştirilmesine önemli katkılar sağlamaktadır. Böylece tarımsal planlamada karar destek sistemlerinin etkinliği artırılmıştır. [9]
6. Oregon State University Extension (2020). "Environmental Factors Affecting Plant Growth". Kaynak, bitki gelişimini etkileyen temel çevresel faktörler olan ışık, sıcaklık, su, nem ve besinlerin etkilerini kapsamlı şekilde açıklamaktadır. Fotosentez ve terleme gibi süreçlerin optimal koşulları detaylandırılmıştır. [10] Bitki sağlığı ve verimliliği açısından bu çevresel parametrelerin dengelenmesinin önemi vurgulanmaktadır. [10]
7. Islam, S., Reza, M. N., Chowdhury, M., et al. (2021). "A review on effect of ambient environment factors and monitoring technology for plant factory Precision". Kontrollü ortamlarda gerçekleştirilen bu derleme, sıcaklık, nem, ışık ve karbondioksit gibi çevresel faktörlerin bitki gelişimi üzerindeki etkilerini ele almaktadır. Bitki fabrikalarında kullanılan izleme teknolojilerinin hassas kontrol sağlamadaki rolü vurgulanmıştır. Optimal üretim için çevresel parametrelerin ayarlanmasının verimlilik üzerindeki olumlu etkileri ortaya konmuştur. [11]
8. Sharma, S., Jain, A., Sharma, S., & Whig, P. (2022). "Enhancing crop yield prediction through machine learning". Bu çalışma, lojistik regresyon tekniklerini kullanarak mahsul verimi ve hastalık tahmini yapmıştır. [12] Model, sulama ve gübreleme stratejilerinin optimizasyonunda başarıyla uygulanmıştır. Böylece, üretkenliği artıracak karar destek sistemleri geliştirilmesine olanak sağlanmıştır. . [12]
9. Satpathi, A., Setiya, P., Das, B., et al. (2021). "Comparative Analysis of Statistical and Machine Learning Models for Rice Yield Prediction in Chhattisgarh, India". Çalışmada, pirinç verimi tahmini için farklı istatistiksel ve makine öğrenimi modellerinin performansı karşılaştırılmıştır. [13] Derin öğrenme algoritmaları, karmaşık veri ilişkilerini yakalamada daha üstün performans göstermiştir. Sonuçlar, derin öğrenmenin bölgesel tarım uygulamalarında tercih edilmesi gerektiğini göstermektedir. [13]
10. USDA (2022). "Sustainable Agricultural Productivity Growth: What, Why, and How". Belge, tarımsal üretkenlik büyümesinin çevresel etkiyi azaltarak, daha az kaynak kullanımıyla daha fazla üretim sağlamada kritik olduğunu belirtmektedir. Teknolojik gelişmelerin sürdürülebilir tarıma katkısı vurgulanmıştır. Politika yapıcılar için stratejik öneriler sunulmaktadır. [4]
11. U.S. Government Accountability Office (GAO) (2021). "Precision Agriculture: USDA Could Do More to Help Farmers Adopt and Use Technologies". Raporda, hassas tarım teknolojilerinin çiftçiler tarafından benimsenmesini engelleyen maliyet, veri güvenliği ve standart sorunları tartışılmıştır. Eğitim ve destek programları önerilmiştir. Teknoloji yayılımının artırılması için stratejiler geliştirilmiştir. [3]
12. Farmbrite (2020). "The Role of Data in Modern Farming: Separating Fact from Fiction". Modern tarımda kullanılan veri türleri ve bu verilerin yönetimindeki zorluklar incelenmiştir. Veri yükü, beceri eksiklikleri ve yatırımların önemi vurgulanmıştır. Yapay zeka ve otomasyonun tarıma entegrasyonu geleceğe yönelik umut vaat etmektedir. [3]
13. Hariyani, G. (2022). "Analysis on Crop Yield Prediction using various Ensemble Methods". Hindistan’da topluluk öğrenme yöntemleriyle mahsul verimi tahmini yapılmıştır. Random Forest, XGB Regression ve Stacking Regression algoritmaları karşılaştırılmıştır. Random Forest, yüksek doğruluk oranıyla en başarılı model olarak belirlenmiştir. [14]
14. ResearchGate Contributors (2021). "Intelligent Environmental Control in Plant Factories: Integrating Sensors, Automation, and AI for Optimal Crop Production". Bitki fabrikalarında sensör, otomasyon ve yapay zeka entegrasyonunun çevresel kontrol üzerindeki etkileri incelenmiştir. Kritik parametrelerin izlenmesi ve yönetiminde bu teknolojilerin önemi vurgulanmıştır. Çalışma, optimal üretim sağlamak için bu sistemlerin gerekliliğini göstermektedir. [15]
15. Wikipedia (2023). "Agricultural productivity". Bu kaynak, tarımsal üretkenliğin çıktı ve girdi oranları ile toplam faktör verimliliği (TFP) kullanılarak ölçülmesini açıklamaktadır. Üretkenlik artışının gıda güvenliği, yoksulluk azaltımı ve ekonomik büyüme üzerindeki etkileri detaylandırılmıştır. Tarımsal verimliliğin sosyal ve ekonomik önemi vurgulanmaktadır. [2]

**2. Yöntem**

Aşağıdaki akış şeması, projesinin genel iş akışını özetlemektedir. Proje süreci; konu belirleme, veri toplama ve ön işleme, model eğitimi ve değerlendirme ile sonuçların raporlanması ve sunulması gibi temel aşamalardan oluşur. Her aşama, projeyi sistematik ve verimli bir şekilde ilerletmek için birbirine bağlı adımlar içerir.



**Şekil 1.** Projenin Akış Şeması

**2.1. Verilerin Toplanması ve Verilerin Analizi**

Bu bölümde, çalışmada kullanılan veri setinin temini, analizi ve makine öğrenimi modellerine uygun hale getirilmesi için gerçekleştirilen ön işleme adımları detaylı bir şekilde sunulmaktadır.

**2.1.1. Veri Setinin Yapısı ve Özellikleri**

Bu çalışma, Kaggle platformunda yer alan ve tarımsal ürünlerin çevresel ve toprak özellikleriyle ilişkisini incelemeyi amaçlayan bir veri setini kullanmaktadır. [16] Veri seti, toplam 15.400 gözlem içermekte olup, her biri farklı çevresel ve kimyasal özelliklere sahip 22 farklı tarımsal ürün türünü kapsamaktadır. Bu yönüyle veri seti, hem çeşitlilik hem de örneklem büyüklüğü bakımından istatistiksel analiz ve makine öğrenmesi uygulamaları için elverişli bir temel sunmaktadır.

Veri seti, 13’ü sayısal, 6’sı kategorik olmak üzere toplam 19 değişkenden oluşmaktadır. Her bir satır, belirli bir ürün türüne ait gözlemi temsil etmektedir. Bu değişkenlerin bir kısmı doğrudan çevresel koşulları (örneğin sıcaklık, ışık süresi), bir kısmı ise toprak özelliklerini (örneğin pH, azot-fosfor-potasyum düzeyleri) tanımlamaktadır. Veri seti, herhangi bir eksik ya da hatalı değer içermemektedir.

Veri setinde yer alan her ürün türü, eşit sayıda gözlemle temsil edilmiştir. Toplamda 22 ürün sınıfı bulunmakta olup, her biri için 700 adet örnek veri yer almaktadır. Bu durum, sınıflandırma modellerinde sıklıkla karşılaşılan "sınıf dengesizliği (class imbalance)" sorununu ortadan kaldırmakta ve algoritmaların adil öğrenimini sağlamaktadır.  
  
 Veri setine aşağıdaki bağlantıdan erişilebilir: [Soil Nutrients Dataset - Kaggle](https://www.kaggle.com/datasets/snmahsa/soil-nutrients) [16]

**2.1.2. Veri Karakteristikleri ve Tarımsal Faktörlerle İlişkisi**

**2.1.2.1. Sayısal Değişkenler**

* Sıcaklık (Temperature): Ortalama sıcaklık değeri 20.8°C olup, pozitif yönde çarpık bir dağılım göstermektedir. Bu durum, örneklerin çoğunun orta sıcaklıklarda toplandığını, ancak bazı yüksek değerli örneklerin mevcut olduğunu göstermektedir.
* Yağış (Rainfall): Ortalama yağış miktarı 948 mm olup, yüksek standart sapma ile geniş bir dağılıma sahiptir. En yüksek gözlemler 2500 mm’yi aşmaktadır.
* pH: Ortalama pH değeri 6.47 olup, hafif asidik ve nötr koşullara işaret etmektedir.
* Işık Süresi (Light\_Hours) ve Işık Yoğunluğu (Light\_Intensity): Sırasıyla günlük ortalama ışık süresi yaklaşık 9.5 saat ve ışık yoğunluğunu tanımlamaktadır.
* Bağıl Nem (Rh): Ortalama bağıl nem %73 olup, birçok ürünün optimum nem gereksinimlerini karşılayacak düzeydedir.
* Azot (Nitrogen), Fosfor (Phosphorus), Potasyum (Potassium): Bitki gelişiminde temel rol oynayan bu üç makro besin elementi, hem mutlak değerler hem de normalize edilmiş oranlarla (N\_Ratio, P\_Ratio, K\_Ratio) veri setinde yer almaktadır.

### 2.1.2.2. Kategorik Değişkenler

* Name: Ürün türünü belirtir ve sınıflandırma algoritmalarında hedef (etiket) değişken olarak kullanılacaktır.
* Fertility: Toprağın doğal verimliliğini tanımlar (yüksek, orta, düşük).
* Photoperiod: Bitkinin büyüme sürecinde ihtiyaç duyduğu ışık tipi (uzun gün, kısa gün, nötr).
* Season: Bitkinin yetiştirildiği mevsimi gösterir (ilkbahar, yaz, sonbahar, kış).
* Soil\_Type: Toprak türünü belirtir (killi, tınlı, kumlu vb.).
* Category\_pH**:** Toprak pH değerinin kategorik sınıflandırmasıdır (asidik, nötr, alkali).

**2.1.3. Çevresel ve Toprak Faktörlerinin Mahsul Verimi Üzerindeki Etkisi**

**2.1.3.1. Toprak Besin Maddeleri (Azot, Fosfor, Potasyum)**

Azot (N), fosfor (P) ve potasyum (K) gibi besin maddelerinin gübrelemesi, toprak fiziksel ve kimyasal özelliklerini önemli ölçüde etkilemektedir. [7] Bilimsel gübreleme, mahsul verimliliğini artırmanın en etkili yöntemlerinden biri olarak kabul edilmektedir. Ancak, gübrelemenin verimle ilişkisi doğrusal bir şekilde ilerlemez; belirli bir noktaya kadar faydalar artarken, aşırı gübre kullanımı olumsuz sonuçlara yol açabilir. Aşırı gübreleme, toprak asidifikasyonu, besin dengesizliği, kök ve yaprak gelişiminin bozulması gibi istenmeyen etkiler yaratabilir. Bu durum, optimal gübreleme stratejilerinin belirlenmesinin önemini ortaya koymaktadır. Veri setindeki N\_Ratio, P\_Ratio, K\_Ratio gibi sütunlar, bu optimal oranların belirlenmesine yönelik önemli bir potansiyel sunmaktadır. Makine öğrenimi modelleri, sadece mahsul verimliliğini tahmin etmekle kalmayıp, aynı zamanda farklı mahsul türleri ve çevresel koşullar için en uygun gübreleme stratejilerini önerme kapasitesine sahip olabilir.

**2.1.3.2. Çevresel Faktörler (Sıcaklık, Yağış, pH, Işık Saati, Işık Yoğunluğu, Bağıl Nem)**

Çevresel faktörler, bitki büyümesini ve verimliliğini doğrudan etkileyen önemli unsurlardır. [10] Sıcaklık, ışık yoğunluğu, ışık süresi, su ve nem gibi çevresel değişkenler, bitkilerin fotosentez, terleme ve genel büyüme süreçlerini etkiler. [11] Örneğin, ışık miktarı ve kalitesi, bitkinin büyümesini ve fotosentez kapasitesini doğrudan etkileyerek verimi belirleyen önemli faktörlerdir. Sıcaklık, bitki süreçlerinin çoğunu (çimlenme, çiçeklenme ve mahsul kalitesi) etkileyen bir faktör olup, optimum sıcaklık aralığı her bitki türü için değişkenlik göstermektedir. Bağıl nem, bitkinin terleme hızını ve su ile besin alımını doğrudan etkileyerek büyüme üzerinde belirleyici rol oynar. [10][11] Bu çevresel faktörlerin birbirleriyle olan etkileşimi, örneğin yüksek sıcaklık ve düşük nemin birleşimi, terleme hızını artırarak bitki verimliliğini etkileyebilir. Bu karmaşık ilişkiler, geleneksel tarım yöntemlerinin zorluklar yaşamasına neden olurken, makine öğrenimi modelleri bu faktörlerin optimal kombinasyonlarını belirleme konusunda büyük bir potansiyel sunmaktadır. Hassas tarım uygulamalarının temelini, bu çevresel faktörlerin etkili bir şekilde analiz edilmesi ve optimize edilmesi oluşturur.

**2.1.3.3. Kategorik Değişkenler (Name, Fertility, Photoperiod, Category\_pH, Soil\_Type, Season)**

Kategorik değişkenler, veri setinin segmentlere ayrılmasına ve mahsul türlerine özgü gereksinimlerin ile toprak uygunluğunun daha iyi anlaşılmasına olanak tanır. Bu değişkenler, mahsul türlerinin büyüme süreçlerinde önemli bir rol oynar. Örneğin, farklı mahsullerin (örneğin çilek, üzüm, marul) farklı sıcaklık, pH ve fotoperiyot gereksinimleri vardır. [25] Ayrıca, toprak türü ve yetiştirme mevsimi gibi faktörler de mahsul verimi üzerinde belirleyici etkiler yaratmaktadır. Bu nedenle, Name, Fertility, Photoperiod, Category\_pH, Soil\_Type, ve Season gibi kategorik değişkenler, tarımsal verimlilik üzerine yapılan analizlerde kritik öneme sahiptir. Bu faktörlerin doğru şekilde kodlanması, makine öğrenimi modellerinin doğru tahminler yapabilmesi için gereklidir.

**2.1.4. Veri Ön İşleme ve Modelleme Süreci**

Makine öğrenimi algoritmalarını verimli bir şekilde eğitmek ve doğruluklarını artırmak amacıyla veri setinin uygun şekilde işlenmesi gerekmektedir. Veri ön işleme süreci, modelin başarısını doğrudan etkileyen kritik adımları içermektedir. Bu adımlar, aşağıda detaylı şekilde açıklanmıştır.

* Eksik Veri ve Aykırı Değer Yönetimi: Veri setindeki eksik değerlerin tespiti ve uygun yöntemlerle giderilmesi, veri kalitesinin korunması açısından önemlidir. Eksik veriler, ortalama, medyan veya regresyon gibi istatistiksel yöntemlerle doldurulmuş, aykırı değerler ise sistematik olarak incelenerek, modelin performansını bozabilecek verilerden arındırılmıştır. [5][18] Bu adım, modelin eğitimi sırasında doğruluğun arttırılması adına kritik bir süreçtir.
* Kategorik Değişkenlerin Kodlanması: Veri setinde yer alan Name, Fertility, Photoperiod, Category\_pH, Soil\_Type, ve Season gibi kategorik değişkenler, makine öğrenimi algoritmaları tarafından işlenebilir sayısal verilere dönüştürülmüştür. Bu dönüştürme işlemi için One-Hot Encoding ve Label Encoding gibi teknikler kullanılmıştır. [5][6] Kategorik değişkenlerin sayısal formatlara dönüştürülmesi, modelin bu tür verileri doğru bir şekilde analiz etmesine olanak sağlamaktadır.
* Özellik Ölçeklendirme ve Standardizasyon: Modelin başarısını artırmak için, sayısal verilerin birbirinden farklı ölçeklerde olması durumu ele alınmıştır. Özellikle Temperature, Rainfall, pH, Light\_Hours, Light\_Intensity, Rh, Nitrogen, Phosphorus, Potassium, N\_Ratio, P\_Ratio, K\_Ratio, ve Yield gibi sayısal özelliklerin ölçeklendirilmesi yapılmıştır. Standard Scaler veya MinMax Scaler gibi yöntemler kullanılarak, bu özellikler belirli bir aralığa (örneğin 0 ile 1 arasında) dönüştürülmüş ve modelin eğitimi sırasında her bir özelliğin eşit derecede önemli olmasını sağlamak amacıyla normalize edilmiştir. [5][26][27] Bu süreç, algoritmaların daha tutarlı ve doğru öğrenme yapabilmesine olanak tanımaktadır.
* Eğitim ve Test Veri Setlerinin Ayrılması: Modelin genellenebilirliğini test etmek amacıyla, veri seti eğitim ve test kümelerine ayrılmıştır. Eğitim veri seti, modelin öğrenmesi için kullanılırken, test veri seti modelin daha önce görmediği verilerle performansını değerlendirmek için kullanılmaktadır. Yaygın olarak %80 eğitim, %20 test oranı benimsenmiş ve modelin dış veriler üzerindeki başarısı test edilmiştir. [5][17] Bu işlem, modelin aşırı öğrenme (overfitting) riskini azaltmak ve gerçek dünya verileri üzerinde nasıl performans göstereceğini görmek adına önemlidir.

**2.2. Geliştirme Ortamı ve Kütüphaneler**

Bu proje, Python programlama dili kullanılarak geliştirilmiştir. Python, makine öğrenimi, veri işleme ve analiz alanlarında geniş bir kütüphane desteği sağlayan, esnek ve güçlü bir dildir. Geliştirme süreci sırasında, çeşitli kütüphaneler kullanılarak veri analizi, modelleme, model değerlendirmesi ve görselleştirme adımları gerçekleştirilmiştir. Aşağıda, proje kapsamında kullanılan başlıca kütüphaneler ve araçlar ile ilgili ayrıntılar verilmiştir:

**2.2.1. Kurulumlar ve Kütüphaneler**

Proje Python programlama dili kullanılarak geliştirilmiş olup, makine öğrenimi ve veri analizi süreçlerinde yaygın olarak tercih edilen çeşitli kütüphaneler entegre edilmiştir. Temel olarak, makine öğrenimi algoritmalarını içeren scikit-learn, özellikle kategorik verilerde güçlü performans sunan CatBoost, veri özetleme için skimpy ve model çıktılarının açıklanabilirliğini artıran SHAP kütüphaneleri kullanılmıştır. [5] [6] [24] [25] Ayrıca, gelişmiş gradyan artırma modelleri için XGBoost ve LightGBM gibi kütüphaneler de projede yer almaktadır. [6] [9] [14] [30]

**2.2.2. İçe Aktarılan Kütüphaneler ve Fonksiyonlar**

Veri işleme ve analiz süreçlerinde, sayısal işlemler için NumPy, veri manipülasyonu için Pandas, görselleştirme için Matplotlib ve Seaborn kütüphaneleri aktif olarak kullanılmıştır. İşlem sürelerini ölçmek için time, veri kümesi üzerindeki dosya yönetimi işlemleri için ise os ve glob kütüphaneleri tercih edilmiştir. Veri ön işleme kapsamında, kategorik değişkenlerin kodlanması ve sayısal verilerin standardizasyonu için scikit-learn.preprocessing modülü kullanılmıştır. [5] [26]

Modelleme sürecinde, farklı makine öğrenimi algoritmaları (doğrusal modeller, destek vektör makineleri, karar ağaçları, rastgele ormanlar, sinir ağları gibi) ve kümeleme yöntemleri scikit-learn kütüphanesinden sağlanmıştır. [17] [21] [22] [28]. Model seçimi ve değerlendirme işlemleri için train\_test\_split, cross\_val\_score, GridSearchCV gibi fonksiyonlar kullanılarak hiperparametre optimizasyonu ve çapraz doğrulama gerçekleştirilmiştir. [5][17] Ayrıca, model performansını ölçmek amacıyla doğruluk, F1 skoru, ROC-AUC gibi çeşitli metrikler kullanılmıştır. [5][17]

Büyük veri setlerinde model eğitim sürecinin takibi için tqdm ile ilerleme çubukları sağlanmış, eğitimli modellerin saklanması ve tekrar yüklenmesi için joblib kütüphanesi kullanılmıştır. Ayrıca, benzersiz kayıtların tanımlanması için uuid modülü tercih edilmiştir.

#### 2.2.3. Ortam

Proje, Google Colab ortamında çalıştırılmıştır. Google Colab, Python kodlarının çalıştırılmasını sağlayan bulut tabanlı bir platform olup, GPU ve TPU gibi donanım hızlandırıcıları da desteklemektedir. Proje kapsamında, veri setine erişim Google Drive üzerinden sağlanmış ve tüm kodlar bu platformda geliştirilmiştir. Google Colab, aynı zamanda yazılım ortamını düzenlemeye gerek kalmadan tüm paylaşımlı çalışma ve modelleme süreçlerini bulut üzerinden hızlı ve verimli bir şekilde yönetmeyi mümkün kılmıştır. [19] [23] [29]

## 2.3. Veri Yükleme ve İlk İnceleme

Projenin ilk aşaması, Soil Nutrients.csv veri setinin yüklenmesi ve kapsamlı bir ön incelemesinin yapılmasıdır. Bu aşama, veri kalitesini değerlendirmek ve sonraki analizler için sağlam bir temel oluşturmak açısından kritik öneme sahiptir.

### 2.3.1. Veri Yükleme

Veri seti, Google Drive üzerinde belirtilen bir dizinden ("/content/drive/MyDrive/Colab Notebooks/data/Soil Nutrients.csv") pandas kütüphanesi kullanılarak DataFrame formatında yüklenmiştir. [16] Bu yöntem, veri manipülasyonu ve analiz için Python ortamında etkin veri işleme olanağı sağlamaktadır. Veri seti, yaklaşık 15.400 gözlem ve 19 değişken içermekte olup, tarımsal ürünlerin toprak besin maddeleri ve çevresel koşullarla ilişkisini detaylandırmaktadır.

### 2.3.2. Eksik Değerlerin Yönetimi

Veri setinde bulunan tüm eksik (NaN) değerler, dropna(inplace=True) komutu ile veri kaybına yol açmamak adına satır bazında temizlenmiştir. Bu yöntem, model performansını olumsuz etkileyebilecek eksik verilerin neden olduğu sapmaları önlemek için tercih edilmiştir. Eksik veri oranının düşük olması, bu temizleme işleminden sonra veri setinin bütünlüğünü korumasını sağlamıştır.

### 2.3.3. Sütun Tiplerinin Belirlenmesi ve Kategorik Değişkenlerin İncelenmes

### Veri setindeki sütunlar, veri tiplerine göre iki ana kategoriye ayrılmıştır:

* Kategorik sütunlar (object tipi): Ürün türü, toprak verimliliği, fotoperiyot, mevsim, toprak tipi ve pH kategorisi gibi değişkenler.
* Sayısal sütunlar (float64 tipi): Sıcaklık, yağış, pH, ışık süresi, ışık yoğunluğu, bağıl nem, azot, fosfor, potasyum ve verim gibi değişkenler.

Her kategorik sütun için value\_counts() yöntemi ile benzersiz değerlerin dağılımı analiz edilerek, veri dengesizliği veya anormal dağılımlar olup olmadığı araştırılmıştır. Bu, modelin öğrenme sürecinde potansiyel dengesizliklerin ve hata kaynaklarının erken tespitine yardımcı olmuştur.

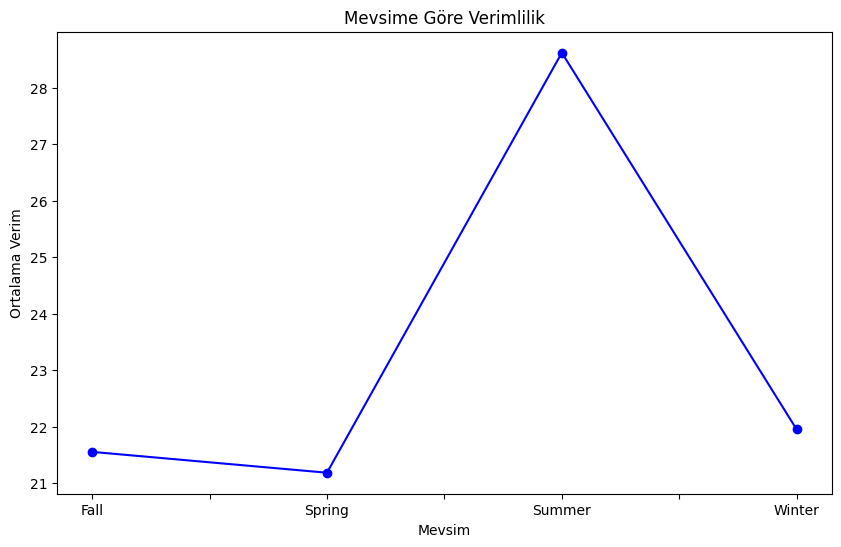
### 2.3.4. Özet İstatistiklerin Elde Edilmesi

Veri setinin genel yapısını hızlı ve kapsamlı şekilde anlamak için skimpy kütüphanesinin skim() fonksiyonu kullanılmıştır. Bu yöntemle Sayısal sütunların ortalama, standart sapma, medyan, çeyreklik değerleri gibi temel istatistikleri ve Kategorik sütunların mod, frekans ve benzersiz değer sayıları gibi özet bilgileri hızlıca elde edilmiştir.

Böylece, veri setindeki olası anormallikler ve genel dağılım hakkında detaylı bilgi edinilmiştir.

### 2.3.5. Görsel İnceleme

Veri dağılımının görselleştirilmesi, verinin yapısal özelliklerini ve ilişkilerini daha iyi anlamaya olanak tanımıştır:

* Sayısal sütunların dağılımı: Her sayısal değişken için 30 bölmeli histogramlar çizilmiştir. Bu grafikler, veri dağılımının şekli, çarpıklık, basıklık ve olası aykırı değerlerin tespiti açısından analiz edilmiştir.
* Kategorik sütunların dağılımı: Her kategorik değişken için seaborn kütüphanesi kullanılarak çubuk grafikler (countplot) oluşturulmuştur. Bu sayede, sınıf dağılımlarındaki dengesizlikler ve baskın kategoriler görsel olarak ortaya konmuştur.
* Mevsimlere göre verim (Yield) analizi: Ürün verimi mevsimlere göre gruplanarak, mevsimsel verim farklılıkları groupby('Season')['Yield'].mean() fonksiyonu ile hesaplanmıştır. Elde edilen sonuçlar çizgi grafik ile görselleştirilmiş ve farklı yetiştirme mevsimlerinin verim üzerindeki etkisi değerlendirilmiştir. Aşağıdaki çıkarılan graf çizgisi mevsime göre analizi temsil eder.

**Şekil 2.** Mevsime Göre Verimlilik Cizgisi

Bu kapsamlı ön inceleme adımları, veri kalitesinin ve yapısının anlaşılmasını sağlamış, sonraki makine öğrenimi uygulamaları için güvenilir bir zemin hazırlamıştır. Aynı zamanda, verinin doğasından kaynaklanan özelliklerin ve potansiyel sınıf dengesizliklerinin belirlenmesiyle modelleme sürecinde stratejik kararlar alınmasına olanak tanımıştır.

## 2.4. Veri Ön İşleme ve Özellik Mühendisliği

Proje kapsamında ham verinin makine öğrenimi modelleri için uygun hale getirilmesi ve model performansının artırılması amacıyla kapsamlı bir veri ön işleme ve özellik mühendisliği süreci gerçekleştirilmiştir. İlk olarak, hedef değişken olan Season (Mevsim) sütununda ortaya çıkabilecek sınıf dengesizliklerini gidermek için özel bir dengeleme fonksiyonu kullanılmıştır. Bu fonksiyon, veri setindeki azınlık sınıflarını güçlendirmek için alt örnekleme (downsampling) veya üst örnekleme (upsampling) işlemleri yapabilmekte olup, böylece modelin tüm sınıfları adil ve dengeli şekilde öğrenmesi sağlanmıştır.

Kategorik değişkenlerin makine öğrenimi algoritmaları tarafından işlenebilmesi için Label Encoding yöntemi ile Name, Fertility, Photoperiod, Category\_pH, Soil\_Type gibi orijinal kategorik sütunlar sayısal değerlere dönüştürülmüştür. Bu dönüşüm, modellerin kategorik verileri etkin biçimde analiz etmesine olanak sağlamıştır. Ayrıca, modelleme sürecinde hedef değişken olarak Season sütunu seçilmiş, buna bağlı olarak verim (Yield) sütunu özellik setinden çıkarılmıştır; böylece problem bir sınıflandırma problemi olarak tanımlanmıştır.

Modelin tahmin gücünü artırmak amacıyla mevcut özelliklerden yeni türetilmiş değişkenler oluşturulmuştur. Topraktaki azot, fosfor ve potasyum miktarlarının toplamı olan NPK\_Total, azot ile yağışın çarpımı olan Nitrogen\_Rainfall, sıcaklık ile ışık saati çarpımı olan Temp\_Light ve pH değerlerinin asidik, nötr ve alkalin olarak üç kategoriye ayrıldığı pH\_Binned değişkenleri bu süreçte eklenmiştir. Bu yeni özellikler, modelin hem toprak besin maddeleri hem de çevresel faktörler arasındaki etkileşimleri daha iyi öğrenmesini sağlamıştır.

Sayısal değişkenlerin model eğitiminde eşit ağırlıkta değerlendirilmesi için, Temperature, Rainfall, Light\_Hours, Light\_Intensity, Rh, Nitrogen, Phosphorus, Potassium, N\_Ratio, P\_Ratio, K\_Ratio, NPK\_Total, Nitrogen\_Rainfall, Temp\_Light gibi tüm sayısal özellikler StandardScaler ile standartlaştırılmıştır. Bu yöntem, farklı ölçeklerdeki değişkenlerin modele dengeli şekilde katkıda bulunmasına olanak tanımıştır. Öte yandan, kategorik değişkenler (Fertility, Photoperiod, Category\_pH, Soil\_Type, Name ve yeni oluşturulan pH\_Binned) ise One-Hot Encoding yöntemiyle kodlanarak, makine öğrenimi modellerinin bu kategorileri doğru şekilde yorumlaması sağlanmıştır.

Bu kapsamlı veri ön işleme ve özellik mühendisliği süreci, modelin karmaşık ilişkileri öğrenebilmesini ve yüksek tahmin performansı elde etmesini mümkün kılmıştır.

## 2.5. Model Seçimi ve Eğitimi

Bu çalışmada, hedef değişken olan Season (Mevsim) sınıflandırma problemini çözmek üzere farklı makine öğrenimi modelleri seçilmiş ve performansları karşılaştırılmıştır. Model seçiminde amaç, toprak besin maddeleri ve çevresel faktörlere dayanarak mahsulün yetiştiği mevsimi doğru şekilde tahmin edebilen, genellenebilir ve verimli modeller geliştirmektir.

### 2.5.1. Kullanılan Modeller

Projede farklı algoritmaların çeşitli avantajları göz önünde bulundurularak şu sınıflandırıcılar tercih edilmiştir:

* **Decision Tree Classifier (Karar Ağacı)**: Veri setinin temel kalıplarını hızlıca öğrenmek için kullanılmıştır. Model, pipeline yapısı içinde, sayısal özelliklerin standartlaştırılması ve kategorik verilerin dönüşümü sonrası doğrudan eğitilmiştir. Eğitim, eğitim-veri setinin %90’ı kullanılarak gerçekleştirilmiş, %10’luk test seti üzerinde doğruluk ve diğer metriklerle performans ölçülmüştür.
* **Hist Gradient Boosting Classifier**: Verinin boyutu ve karmaşıklığı dikkate alınarak histogram tabanlı gradyan artırma algoritması tercih edilmiştir. Bu model, pipeline içerisinde otomatik erken durdurma (early stopping) mekanizması ile birlikte eğitilmiştir. Eğitim sırasında model, eğitim setinde 5 katlı stratified çapraz doğrulama ile doğruluk optimizasyonu yapmış, test seti üzerindeki tahmin performansı değerlendirilmiştir.
* **Random Forest Classifier**: 100 ağaçlı ansamble yöntemi olarak model kurulmuş ve pipeline ile entegre edilmiştir. Model, eğitim verisinin stratified katmanlı bölünmesi sonrası fit() fonksiyonu ile eğitilmiş, test verisi üzerinde predict() ile sınıflandırma tahminleri yapılmıştır. Model, bagging ve rastgele özellik seçimi sayesinde model varyansını azaltmayı hedeflemiştir.
* **Extra Trees Classifier**: Random Forest’a benzer şekilde, rastgele düğüm bölme kararları içeren 100 ağaçlı yapı kurulmuştur. Pipeline içinde aynı şekilde işlenen verilerle model eğitilmiş ve performansı test edilmiştir.
* **XGBoost Classifier**: Özelleştirilmiş hiperparametrelerle (learning\_rate=0.1, max\_depth=6, n\_estimators=100 gibi) yapılandırılan model, eğitim setinde eval\_set parametresi ile doğrulama verisi izlenerek fit edilmiştir. Eğitim sırasında, eğitim ve test kayıpları (log loss) takip edilmiş, eğitim dinamikleri görselleştirilmiştir. Model, özellikle çok sınıflı softprob çıktı üretme yeteneği sayesinde yüksek performans göstermiştir.
* **Bagging Classifier**: 50 temel karar ağacı kullanarak bootstrap örnekleme yoluyla model çeşitliliği sağlamıştır. Pipeline içerisinde standartlaştırılmış verilerle eğitilmiş, test setinde tahmin yaparak performans değerlendirilmiştir.

**2.5.2. Eğitim Süreci**

Model eğitimi, *train\_models* fonksiyonu içerisinde sistematik ve tekrarlanabilir bir şekilde yürütülmüştür. Eğitim süreci aşağıdaki adımları içermektedir:

* Veri Bölme: Veri seti, modelin genellenebilirliğini sağlamak amacıyla %90 eğitim ve %10 test olmak üzere ikiye ayrılmıştır. Bölme işlemi, hedef değişkenin sınıf dağılımını korumak için stratify parametresi kullanılarak katmanlı (stratified) olarak gerçekleştirilmiştir. Böylece, her sınıfın eğitim ve test setlerinde dengeli dağılımı garanti edilmiştir.
* Pipeline Oluşturma: Her bir sınıflandırıcı için, veri ön işleme ve model eğitim adımlarını tek bir işlem hattı (pipeline) içerisinde birleştiren yapı kurulmuştur. Bu pipeline, özellikle StandardScaler kullanarak sayısal özelliklerin standartlaştırılmasını içerir ve böylece veri tutarlılığı sağlanırken, model eğitim süreci otomatik ve entegre hale getirilmiştir.
* Çapraz Doğrulama: Modellerin farklı veri alt kümelerinde tutarlı performans göstermesi amacıyla 5 katlı Stratified K-Fold çapraz doğrulama yöntemi uygulanmıştır. Bu teknik, modelin overfitting (aşırı uyum) riskini azaltırken, doğruluk ve genellenebilirlik hakkında daha güvenilir tahminler sunar. Her katmanda eğitim ve doğrulama işlemleri ayrı ayrı yürütülmüş, ortalama performans ölçümleri hesaplanmıştır.
* Model Eğitimi ve Tahmin: Her model, eğitim veri seti üzerinde eğitilmiş ve test veri seti kullanılarak tahminler yapılmıştır. Bu süreç, modelin gerçek dünya verileri üzerindeki performansını değerlendirmek için önemlidir.
* Performans ve Zaman Ölçümü: Her modelin eğitim süresi hassas bir şekilde ölçülmüş ve kayıt altına alınmıştır. Bu ölçüm, yalnızca doğruluk değil, aynı zamanda eğitim süresine göre model verimliliğinin de değerlendirilmesini sağlamaktadır. Eğitim süreleri, hesaplama kaynaklarının etkin kullanımının ve model seçiminin optimizasyonunda önemli bir kriter olarak ele alınmıştır.

Bu kapsamlı model seçimi ve eğitim süreci, farklı algoritmaların güçlü ve zayıf yönlerini objektif olarak karşılaştırmakta ve projenin amaçlarına en uygun sınıflandırma modelinin belirlenmesini sağlamaktadır. Ayrıca, uygulanan metodoloji, sonuçların bilimsel güvenilirliğini ve tekrarlanabilirliğini garanti altına almaktadır.

**3. Bulgular**

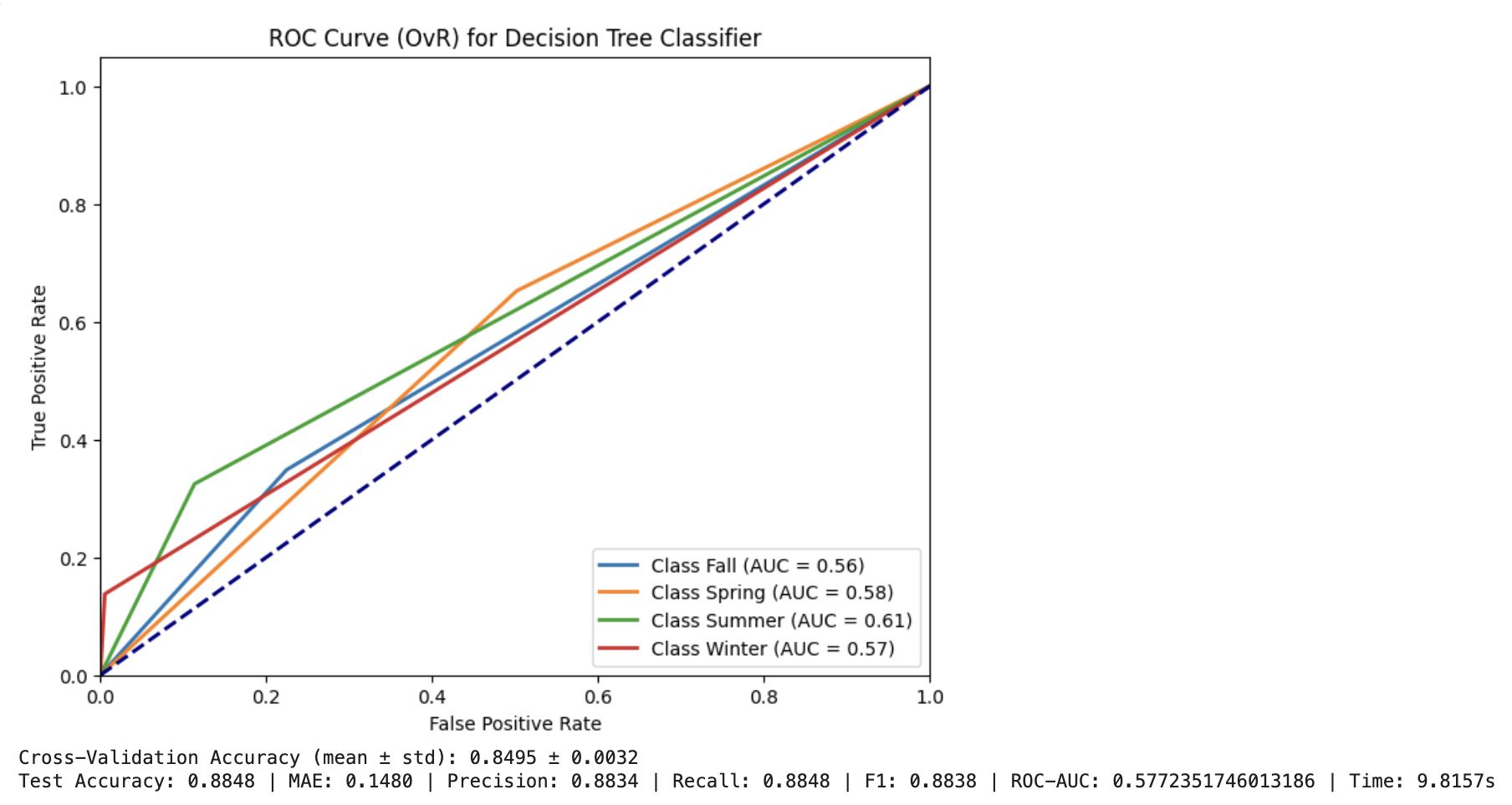
Bu bölümde, geliştirdiğimiz farklı makine öğrenimi modellerinin **Season** (Mevsim) sınıflandırma görevindeki performans sonuçları detaylandırılmıştır. Modellerin test seti üzerindeki başarıları, çeşitli metrikler ve sınıf bazlı değerlendirmeler ile analiz edilmiştir. Ayrıca, model karşılaştırmaları görseller ve tablolar ile desteklenmiştir.

**3.1. Model Bazlı Sınıf Performansları ve Değerlendirmeleri**

Bu alt başlık altında, suçların coğrafi dağılımı, suç yoğunluk haritaları ve suç türlerine göre mekansal analizler sunulmuştur.

**3.1.1. Decision Tree Classifier**

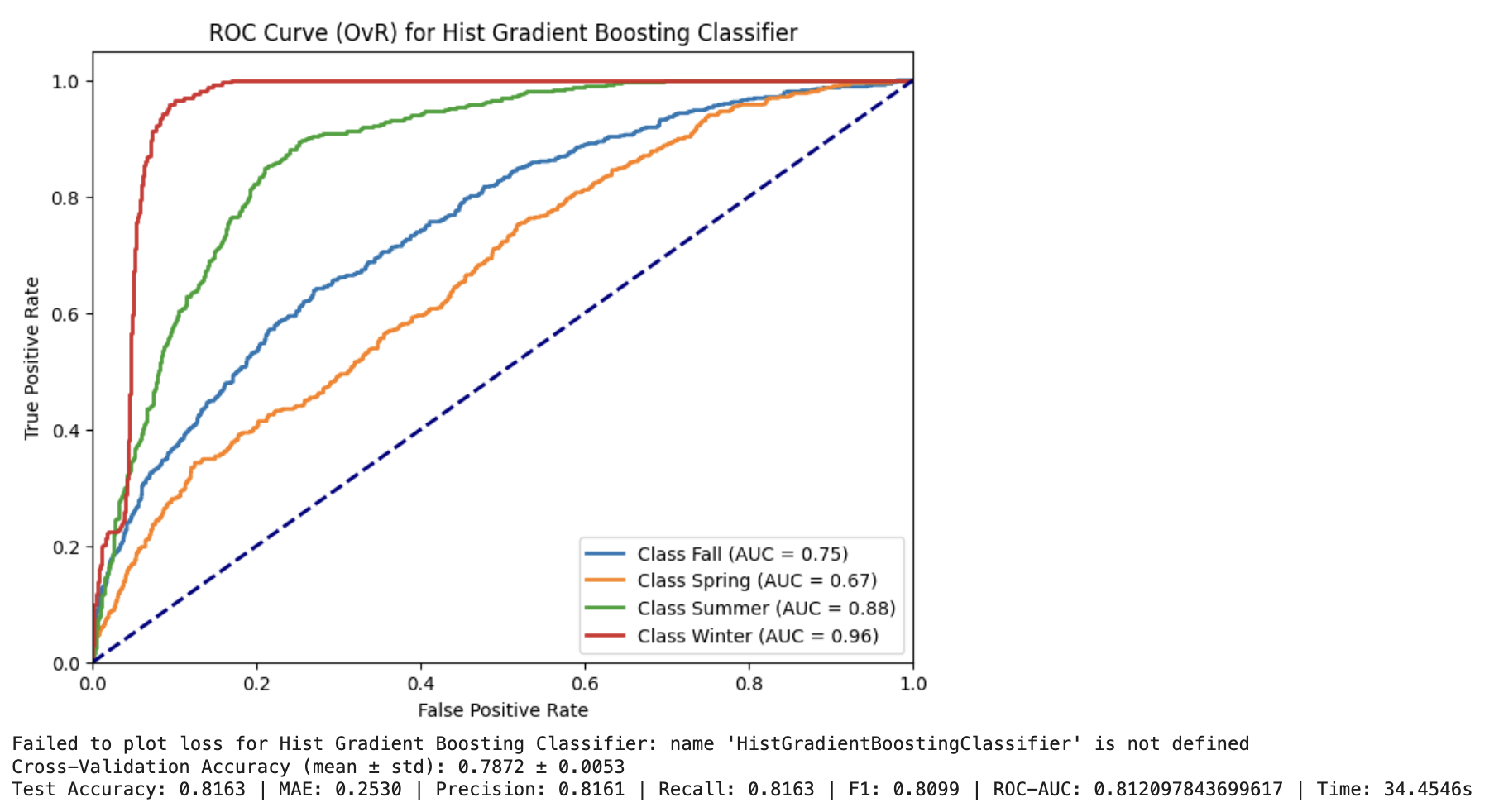
Decision Tree Classifier, özellikle "Winter" sınıfında mükemmele yakın recall (1.00) ve F1 skoru (0.98) elde etmiştir. Diğer sınıflarda da yüksek precision ve recall değerleri gözlenmiş olup, model genel olarak %88.48 doğruluk oranıyla başarılı bir sınıflandırma performansı göstermiştir. Ancak, modelin ROC-AUC değeri 0.5772 gibi düşük bir seviyede kalmıştır; bu durum, sınıf olasılıklarını ayırma gücünün zayıf olduğunu ve karar olasılıklarının güvenilirliğinin sınırlı olduğunu göstermektedir. Buna rağmen, eğitim süresinin 9.8 saniye gibi kısa olması, modeli hızlı sonuçlar gerektiren uygulamalarda tercih edilir kılmaktadır. Karmaşıklık matrisi, çoğu sınıf için doğru tahminlerin yüksek olduğunu ortaya koyarken, özellikle "Fall" ve "Spring" sınıflarında bazı karışıklıklar olduğunu göstermektedir. ROC eğrisi, modelin sınıf ayrımındaki kısıtlarını görsel olarak desteklemektedir.



**Şekil 3**. Decision Tree Classifier’in ROC eğrisi

**3.1.2. Hist Gradient Boosting Classifier**

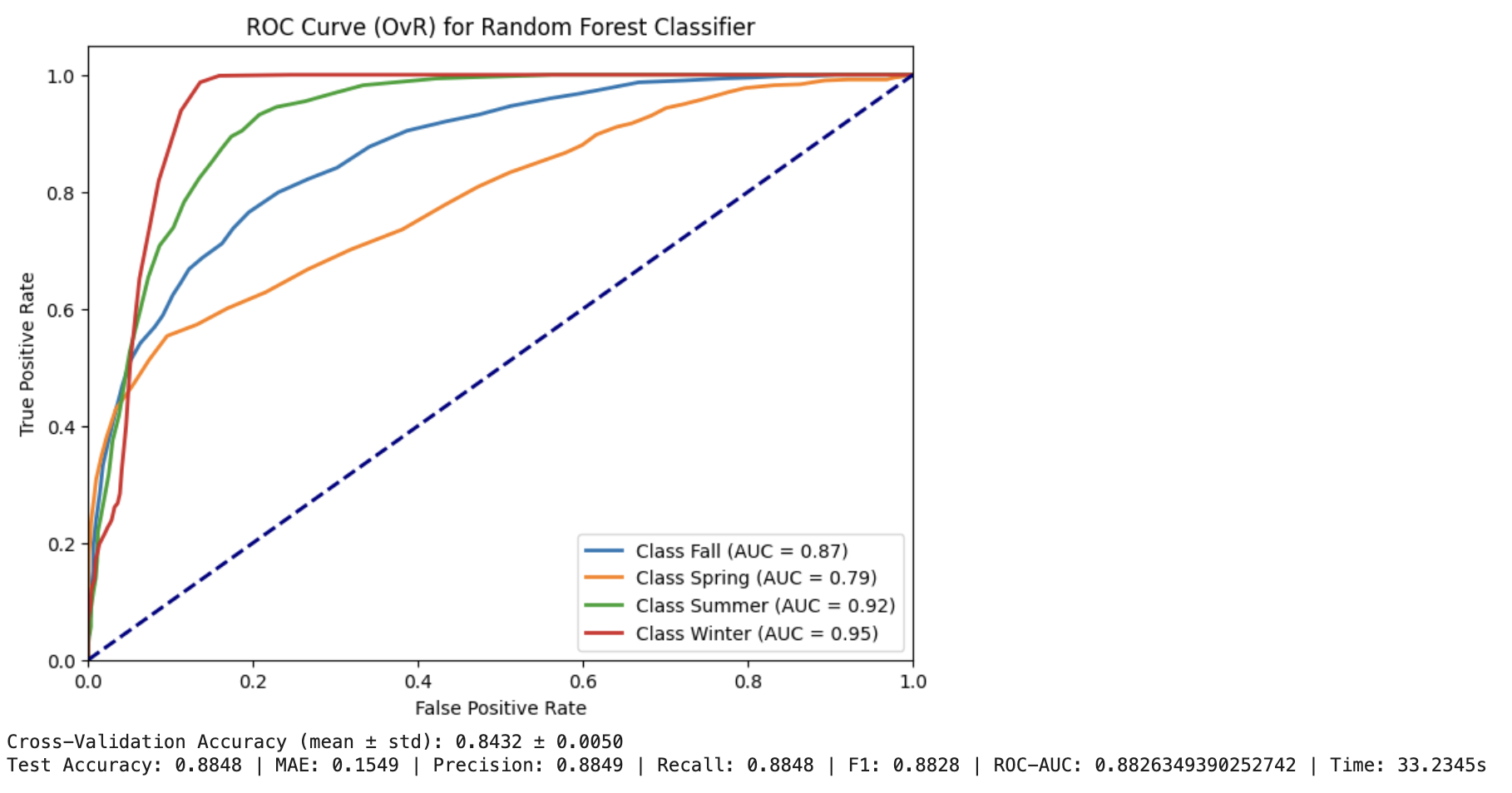
Hist Gradient Boosting Classifier, genel doğruluk açısından diğer modellere kıyasla daha düşük (%81.63) performans sergilemiş olsa da, özellikle "Summer" ve "Winter" sınıflarında yüksek recall değerleri (sırasıyla 0.94 ve 1.00) ile dikkat çekmiştir. Modelin ROC-AUC değeri 0.8121 olarak hesaplanmış ve bu, sınıf olasılıklarını daha iyi ayırabildiğine işaret etmektedir. Karmaşıklık matrisi, "Fall" ve "Spring" sınıflarının ayırt edilmesinde belirli zorluklar olduğunu ortaya koymuştur. Kayıp eğrisi analizi, modelin eğitim ve doğrulama setlerinde dengeli bir öğrenme süreci geçirdiğini, overfitting eğilimi göstermediğini göstermektedir. Eğitim süresi ise yaklaşık 34.5 saniye olarak ölçülmüştür.



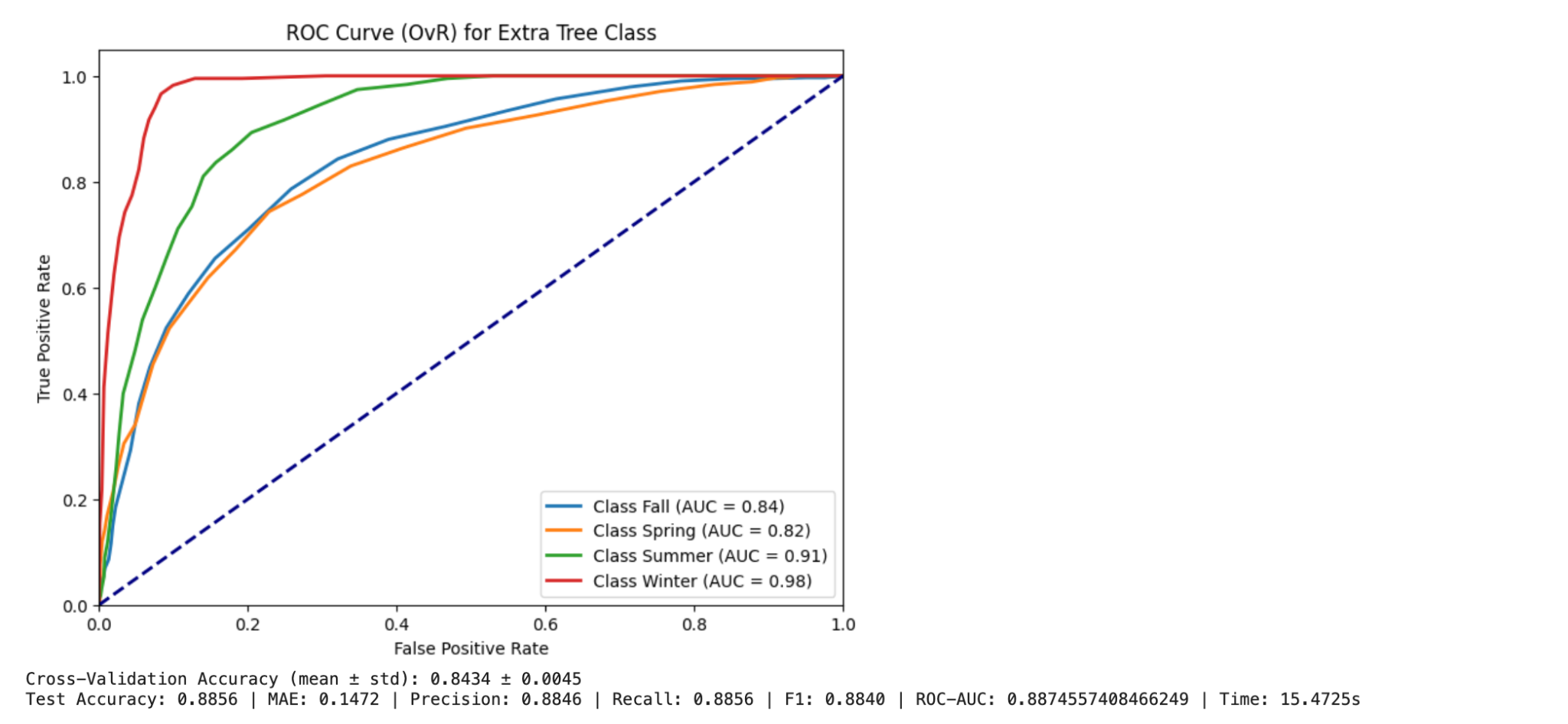
**Şekil 4**.Hist Gradient Boosting Classifier’in ROC eğrisi

**3.1.3. Random Forest Classifier**

Random Forest Classifier, %88.48 test doğruluğu ve 0.8826 ROC-AUC değeri ile Decision Tree modeline kıyasla daha üstün bir ayrım gücü göstermiştir. Precision, recall ve F1 skorları tüm sınıflarda yüksek olup, özellikle "Winter" sınıfında neredeyse mükemmel sonuçlar elde edilmiştir. Karmaşıklık matrisi, modelin dengeli ve güvenilir bir sınıflandırma performansı sunduğunu ortaya koymaktadır. ROC eğrisi sınıflar arası ayrımı kuvvetlendirmekte, eğitim süresi ise 33.2 saniye civarında gerçekleşmiştir. Bu özellikleriyle model, dengeli doğruluk ve güvenilirlik arayan uygulamalar için uygun bulunmuştur.



**Şekil 5**. Random Forest Classifier’in ROC eğrisi

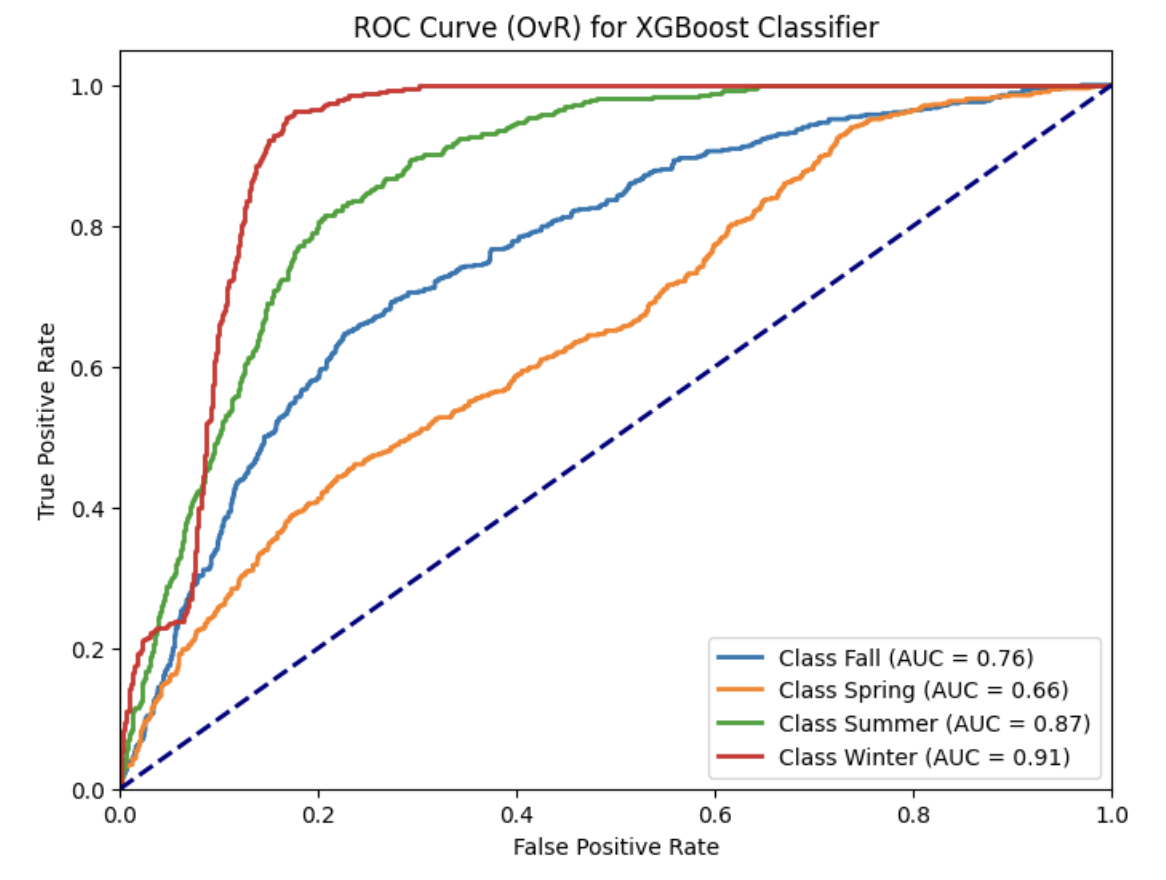
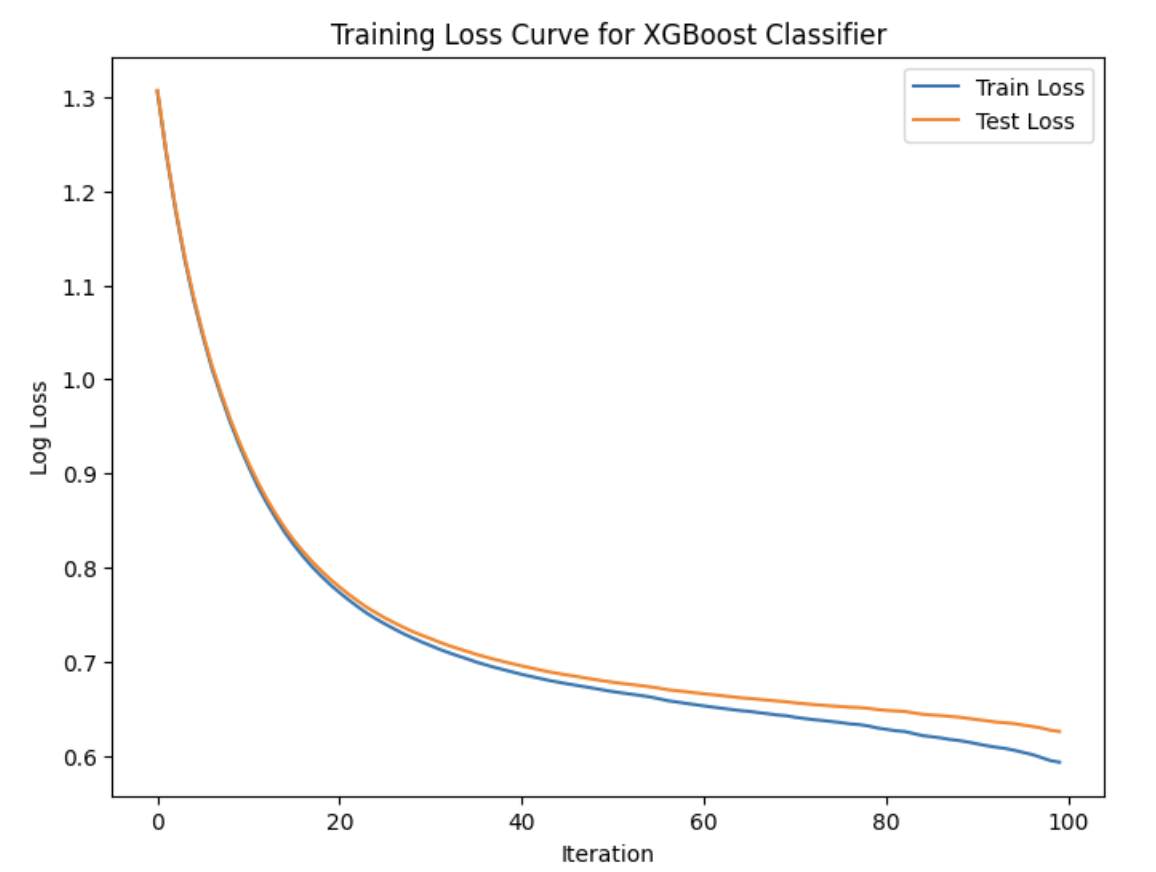
**3.1.4. Extra Tree Class** 

Extra Trees Classifier, hem %88.56 test doğruluğu hem de 0.8875 ROC-AUC değeri ile çalışmada en yüksek performansı sunmuştur. Eğitim süresi ise Random Forest’a göre yaklaşık yarı yarıya daha kısa (15.47 saniye) olup, bu durum verimlilik açısından önemli bir avantaj sağlamaktadır. Model, özellikle "Summer" ve "Winter" sınıflarında yüksek precision ve recall değerleri ile üstün performans göstermiştir. Karmaşıklık matrisi ve ROC eğrisi, modelin güçlü sınıf ayrımı yeteneğini açıkça yansıtmaktadır. Bu model, hassas tarımda sınıflandırma problemlerinde birincil tercih olarak değerlendirilebilir.

**Şekil 6**. Extra Trees Classifier ROC

**3.1.5. XGBoost Classifier**

XGBoost Classifier, %70.88 test doğruluğu ve 0.7998 ROC-AUC değeri ile diğer ensemble yöntemlerine göre daha düşük performans göstermiştir. Özellikle "Fall" ve "Spring" sınıflarında recall değerlerinin sırasıyla 0.52 ve 0.41 gibi düşük seviyelerde kalması, modelin bu sınıfları ayırt etmekte zorlandığını göstermektedir. Bununla birlikte, "Summer" ve "Winter" sınıflarında yüksek recall değerleri ile sınıf bazında bazı başarılar elde edilmiştir. Eğitim süresi 15.5 saniye civarında olup, model hiperparametre optimizasyonu ve daha kapsamlı ayarlamalarla performans artışı sağlanabilir.

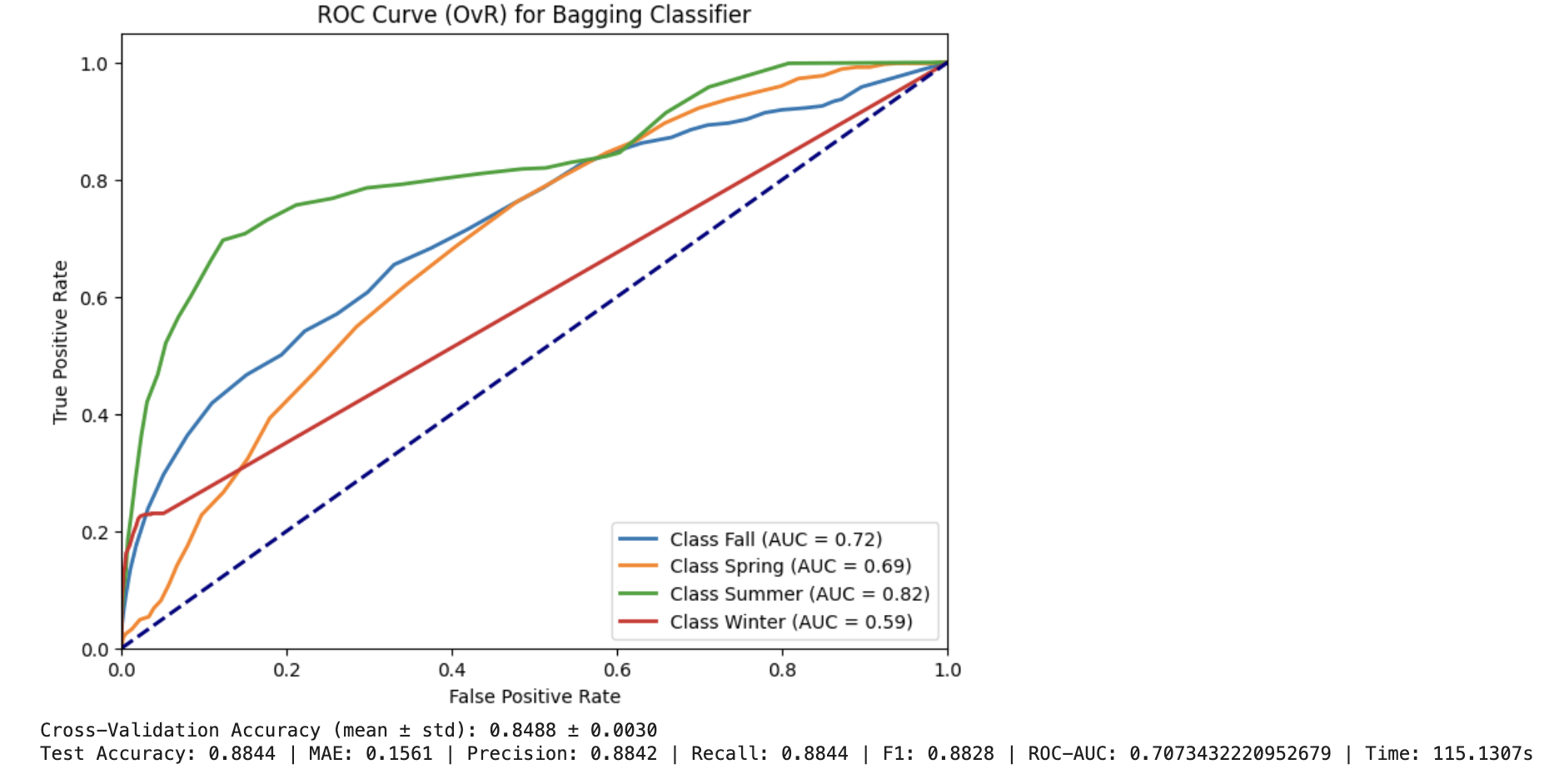




**Şekil 7**. XGBoost Classifier’in ROC ve Kayıp eğrisi

**3.1.6. Bagging Classifier**

Bagging Classifier, genel doğruluk (%88.44), precision, recall ve F1 skorları açısından Decision Tree ve Random Forest modellerine yakın yüksek performans göstermiştir. Ancak, ROC-AUC değeri 0.7073 ile diğer yüksek performanslı ensemble modellerinin gerisindedir. En belirgin dezavantajı ise 115.13 saniye ile en uzun eğitim süresine sahip olmasıdır; bu durum hesaplama kaynakları ve uygulama süresi açısından verimliliği azaltmaktadır. Karmaşıklık matrisi ve ROC eğrisi, modelin genel sınıflandırma başarısını desteklemektedir.

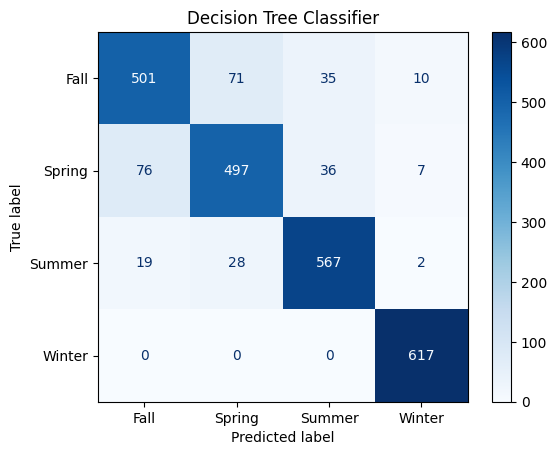


**Şekil 8**. Bagging Classifier ROC

**3.2. Karmaşıklık Matrislerinin Değerlendirilmesi**

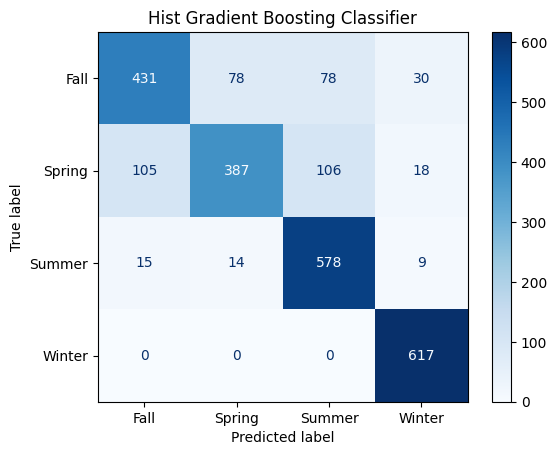
Karmaşıklık matrisleri, her bir modelin farklı mevsim sınıflarını doğru ve yanlış sınıflandırma performansını detaylı olarak ortaya koymaktadır. Diyagonalde yer alan değerler doğru sınıflandırmaları gösterirken, diyagonal dışındaki değerler yanlış sınıflandırmaların türlerini ve oranlarını yansıtmaktadır. Bu analiz, modellerin sınıflar arası ayrım yeteneklerini karşılaştırmak ve zayıf performans gösterilen alanları belirlemek için kritik öneme sahiptir.

**3.2.1. Decision Tree Classifier'ın Karmaşıklık Matrislerinin Değerlendirilmesi**

Decision Tree Classifier'ın karmaşıklık matrisi, modelin genel olarak iyi performans gösterdiğini ortaya koymaktadır. Özellikle "Winter" sınıfı için neredeyse mükemmel bir sınıflandırma görülmektedir (617 doğru tahmin ve 0 yanlış tahmin). "Summer" sınıfı da oldukça iyi tahmin edilmiştir (567 doğru tahmin). Ancak, "Fall" ve "Spring" sınıfları arasında bazı karışıklıklar mevcuttur. Örneğin, 71 "Fall" örneği yanlışlıkla "Spring" olarak, 76 "Spring" örneği ise yanlışlıkla "Fall" olarak sınıflandırılmıştır. Bu, modelin bu iki mevsimi ayırt etmede diğerlerine göre biraz daha zorlandığını göstermektedir. 

**Şekil 9**. Decision Tree Classifier'ın Karmaşıklık Matrisi

**3.2.2. Hist Gradient Boosting Classifier Matrislerinin Değerlendirilmesi**

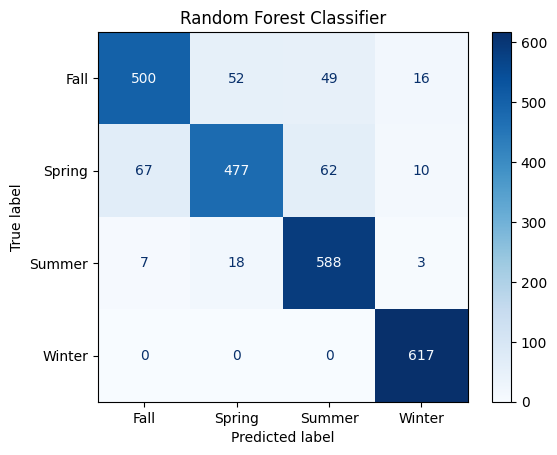


Hist Gradient Boosting Classifier matrisi, "Winter" sınıfı için yine mükemmel bir performans sergilemektedir (617 doğru tahmin). "Summer" sınıfı da yüksek oranda doğru sınıflandırılmıştır (578 doğru tahmin). Ancak, "Fall" ve "Spring" sınıfları arasında daha belirgin yanlış sınıflandırmalar gözlemlenmektedir. "Spring" sınıfından 105 örnek "Fall" olarak, 106 örnek "Summer" olarak ve 18 örnek "Winter" olarak yanlış sınıflandırılmıştır. Benzer şekilde, "Fall" sınıfından da önemli sayıda örnek diğer mevsimlere yanlış atanmıştır. Bu, modelin "Fall" ve "Spring" sınıfları arasındaki ayrım yeteneğinin diğer modellere göre daha zayıf olduğunu göstermektedir.

**Şekil 10**. Hist Gradient Boosting Classifier'ın Karmaşıklık Matrisi

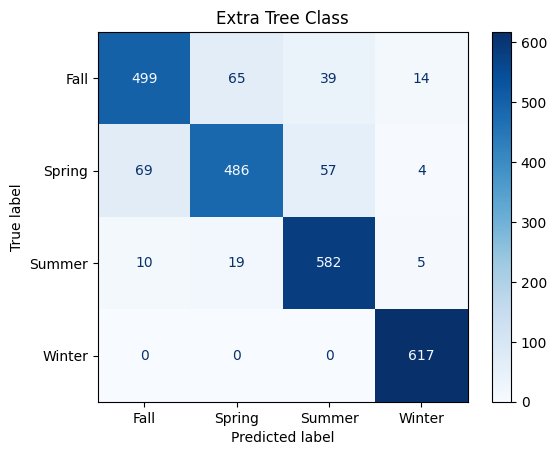
**3.2.3. Random Forest Classifier'ın Karmaşıklık Matrislerinin Değerlendirilmesi**

Random Forest Classifier'ın karmaşıklık matrisi, genel olarak çok güçlü bir performans sergilemektedir. "Winter" sınıfı yine hatasızdır (617 doğru tahmin). "Summer" sınıfı da oldukça yüksek doğrulukla (588 doğru tahmin) sınıflandırılmıştır. "Fall" ve "Spring" sınıfları arasında Decision Tree'ye benzer şekilde bazı yanlış sınıflandırmalar olsa da, Hist Gradient Boosting'e göre daha azdır. Örneğin, 52 "Fall" örneği "Spring" olarak, 67 "Spring" örneği ise "Fall" olarak tahmin edilmiştir. Bu, Random Forest'ın bu sınıfları ayırt etmede daha başarılı olduğunu göstermektedir.



**Şekil 11.** Random Forest’in Karmaşıklık Matrisi

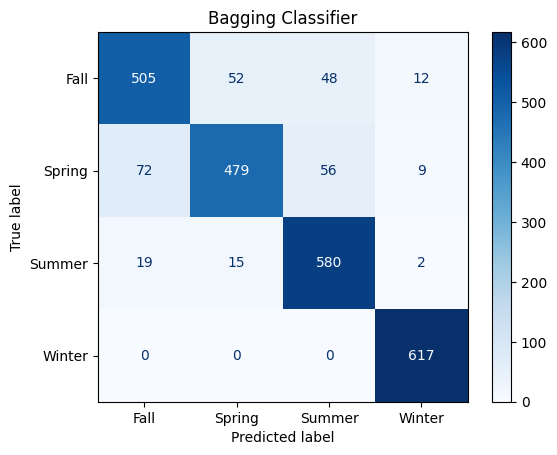
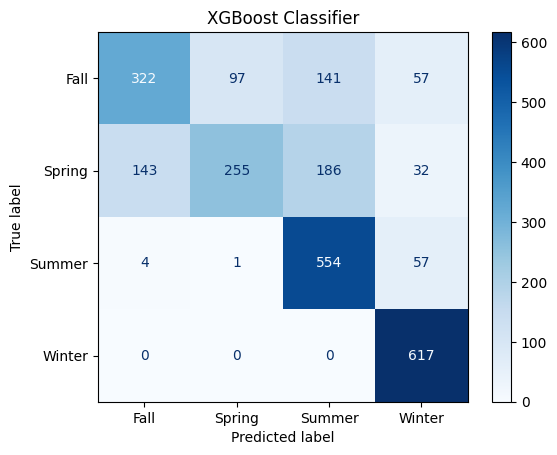
**3.2.4. Extra Tree Class'ın Karmaşıklık Matrislerinin Değerlendirilmesi**

Extra Tree Class'ın karmaşıklık matrisi, tüm modeller arasında en iyi genel performansı yansıtmaktadır. "Winter" sınıfı yine mükemmel bir şekilde tahmin edilmiştir (617 doğru tahmin). "Summer" sınıfı da çok yüksek doğrulukla (582 doğru tahmin) sınıflandırılmıştır. "Fall" ve "Spring" sınıfları arasındaki yanlış sınıflandırmalar diğer modellere göre daha azdır (örneğin, 65 "Fall" örneği "Spring" olarak, 69 "Spring" örneği "Fall" olarak). Bu, modelin tüm mevsim sınıflarını yüksek güvenle ayırt etme yeteneğinin en yüksek olduğunu göstermektedir.

**Şekil 12**. Extra Tree Class'ın HisKarmaşıklık Matrisi

**3.2.5. XGBoost Classifier'ın Karmaşıklık Matrislerinin Değerlendirilmesi**

XGBoost Classifier'ın karmaşıklık matrisi, diğer ensemble modellerine kıyasla daha fazla yanlış sınıflandırma göstermektedir. "Winter" sınıfı yine mükemmel olsa da (617 doğru tahmin), "Fall" ve "Spring" sınıfları için belirgin hatalar mevcuttur. Özellikle "Fall" sınıfından 141 örnek "Summer" olarak, 97 örnek "Spring" olarak ve 57 örnek "Winter" olarak yanlış sınıflandırılmıştır. Benzer şekilde, "Spring" sınıfından da 143 örnek "Fall" olarak, 186 örnek "Summer" olarak yanlış atanmıştır. Bu, XGBoost'un özellikle "Fall" ve "Spring" sınıflarını ayırt etmede önemli zorluklar yaşadığını ve bu sınıflar için daha düşük kesinlik ve duyarlılık değerlerine yol açtığını doğrulamaktadır.

****

**Şekil 13-14**. XGBoost ve Bagging’in Karmaşıklık Matrisi

**3.2.6. Bagging Classifier'ın karmaşıklık Matrislerinin Değerlendirilmesi**

Son olarak, Bagging Classifier'ın karmaşıklık matrisi, Random Forest ve Extra Tree modellerine benzer şekilde güçlü bir performans sergilemektedir. "Winter" sınıfı yine hatasızdır (617 doğru tahmin). "Summer" sınıfı da yüksek doğrulukla (580 doğru tahmin) sınıflandırılmıştır. "Fall" ve "Spring" sınıfları arasında nispeten az yanlış sınıflandırma bulunmaktadır (örneğin, 52 "Fall" örneği "Spring" olarak, 72 "Spring" örneği "Fall" olarak). Bu, Bagging Classifier'ın genel olarak iyi bir sınıflandırma yeteneğine sahip olduğunu göstermektedir, ancak eğitim süresi açısından dezavantajlıdır.

karmaşıklık matrisleri analizi, Extra Trees Classifier ve Random Forest Classifier modellerinin mevsim sınıflarını en doğru ve tutarlı şekilde ayırt ettiğini ortaya koymuştur. Buna karşılık, XGBoost modeli özellikle "Fall" ve "Spring" sınıflarındaki performans eksiklikleri nedeniyle diğer modellere göre daha düşük bir ayrım gücü sergilemiştir. Bu değerlendirmeler, modellerin gerçek dünya tarımsal uygulamalarda kullanım potansiyelini anlamak açısından önemli bilgiler sunmaktadır.

**4. Tartışma ve Sonuç**

Bu çalışmada, toprak besin maddeleri ve çevresel faktörlerin tarımsal verimlilik üzerindeki etkilerini incelemek amacıyla makine öğrenimi teknikleri uygulanmıştır. Kaggle platformundan temin edilen Soil Nutrients.csv veri seti kullanılarak, veri ön işleme ve özellik mühendisliği adımlarının ardından, çeşitli sınıflandırma algoritmaları ile mahsulün yetiştiği mevsimin (Season) tahmini gerçekleştirilmiştir. Modeller arasında Decision Tree, Hist Gradient Boosting, Random Forest, Extra Trees, XGBoost ve Bagging Classifier yer almış ve her model kapsamlı performans ölçütleriyle değerlendirilmiştir.

Elde edilen sonuçlar, **Extra Trees Classifier** ve **Random Forest Classifier** modellerinin, doğruluk, hassasiyet (precision), hatırlama (recall), F1 skoru ve ROC-AUC gibi temel performans metriklerinde üstünlük gösterdiğini ortaya koymuştur. Extra Trees modeli, %88.56 test doğruluğu ve 0.8875 ROC-AUC değeri ile en yüksek başarımı sergilemiş, bu da modelin toprak ve çevresel değişkenler arasındaki karmaşık ilişkileri etkili biçimde öğrenerek mevsimsel sınıflandırma yapabildiğini göstermektedir. Ayrıca, bu modellerin düşük ortalama mutlak hata (MAE) değerleri, tahminlerin gerçek değerlerle yüksek uyum sağladığını desteklemektedir.

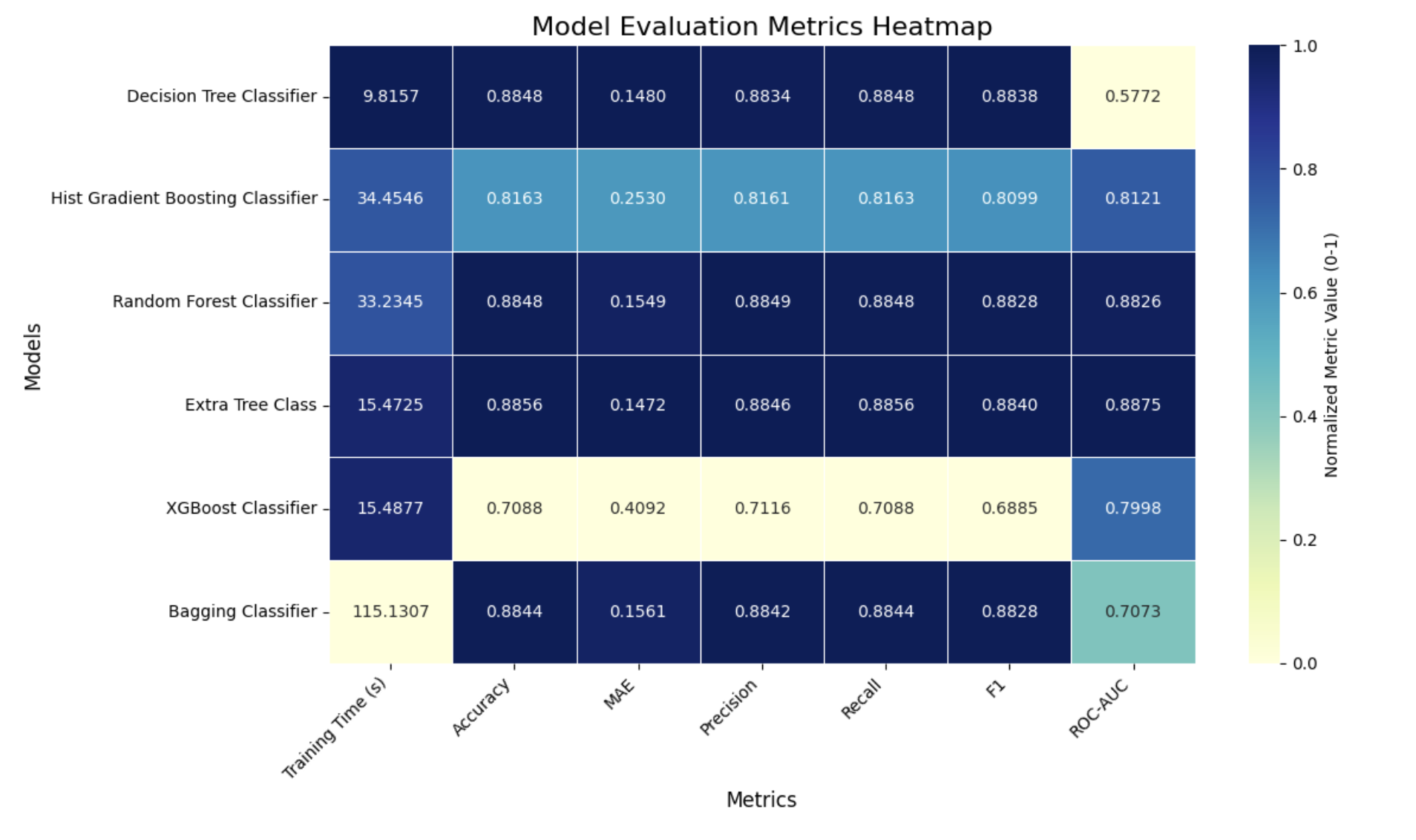
Diğer modeller arasında, XGBoost ve Hist Gradient Boosting kabul edilebilir performans göstermesine rağmen, en başarılı modellerin gerisinde kalmıştır. XGBoost, özellikle sınıf bazında düşük recall değerleri ile sınırlı kalmış ve hiperparametre optimizasyonu gerektirdiği izlenimini vermektedir. Decision Tree sınıflandırıcısı hızlı eğitim süresi sunmasına rağmen, ROC-AUC açısından sınıf olasılıklarını ayrıştırmada yetersiz kalmıştır. Bagging Classifier ise yüksek doğruluk değerlerine ulaşmasına karşın, eğitim süresinin uzunluğu pratik uygulamalarda bir dezavantaj olarak değerlendirilmiştir.

Bu çalışma, tarımsal verimlilik tahmininde makine öğrenimi modellerinin potansiyelini göstermiş ve toprak besinleri ile çevresel faktörler arasındaki karmaşık etkileşimlerin doğru sınıflandırılmasının mümkün olduğunu ortaya koymuştur. Modellerin sınıflandırma başarısı, hassas tarım uygulamalarında optimal ekim mevsiminin belirlenmesi ve kaynak kullanımının iyileştirilmesi için değerli içgörüler sunmaktadır.

**4.1. Uygulama Sonuçları**

**Tablo 1**: Her modelin test sonucları

| Sınıflandırma Model Adı | Eğitim Süresi  (saniye) | Doğrululuk (%) | Kesinlik (%) | MAE  (%) | Geri Çağırma  (%) | F1-Skoru  (%) |
| --- | --- | --- | --- | --- | --- | --- |
| Decision Tree | 9,816 | 88,48 | 88,34 | 14,80 | 88,48 | 88,38 |
| Hist Gradient Boosting | 34,455 | 81,63 | 81,61 | 25,30 | 81,63 | 80,99 |
| Random Forest | 33,235 | 88,48 | 88,49 | 15,49 | 88,48 | 88,48 |
| Extra Tree | 15,473 | 88,56 | 88,46 | 14,72 | 88,56 | 88,40 |
| XGBoost | 15,488 | 70,88 | 71,16 | 40,92 | 70,88 | 68,85 |
| Bagging | 115,131 | 88,44 | 88,42 | 15,61 | 88,44 | 88,28 |

****

**Şekil 15**. Modelin ısı haritası

* Doğruluk (Accuracy) açısından en iyi performans: Extra Tree Class (0.8856) ve Random Forest Classifier (0.8848), Decision Tree Classifier (0.8848) ve Bagging Classifier (0.8844) modelleri benzer ve yüksek doğruluk oranları sergilemiştir.
* MAE (Ortalama Mutlak Hata) açısından en iyi performans: Extra Tree Class (0.1472) en düşük MAE değerine sahipken, XGBoost Classifier (0.4092) en yüksek MAE değerine sahiptir.
* ROC-AUC açısından en iyi performans: Extra Tree Class (0.8875) ve Random Forest Classifier (0.8826) modelleri en yüksek ROC-AUC değerlerine ulaşarak sınıf ayrımında en başarılı modeller olduğunu göstermiştir. Decision Tree'nin ROC-AUC değeri diğer yüksek doğruluklu modellere göre düşüktür, bu da sınıflandırma eşiğine göre daha az kararlı olabileceğini düşündürmektedir.
* Eğitim Süresi: Decision Tree Classifier en hızlı eğitilen model (9.82 s) iken, Bagging Classifier en yavaş model (115.13 s) olmuştur. Extra Tree Class ve XGBoost Classifier orta seviyede eğitim sürelerine sahiptir.

Genel olarak, Extra Tree Class ve Random Forest Classifier modelleri, yüksek doğruluk, düşük MAE ve yüksek ROC-AUC değerleriyle en iyi genel performansı sergilemişlerdir.

**5. Kaynaklar**

[1] USDA. (2022). Sustainable Agricultural Productivity Growth: What, Why, and How.

[2] Wikipedia. (2023). Agricultural productivity. Wikipedia.

[3] Farmbrite. (2020). The Role of Data in Modern Farming: Separating Fact from Fiction.

[4] U.S. Government Accountability Office (GAO). (2021). Precision Agriculture: USDA Could Do More to Help Farmers Adopt and Use Technologies.

[5] Anand, A., & Jhajharia, K. (2022). Advancements in Machine Learning and Deep Learning Techniques for Crop Yield Prediction: A Comprehensive Review. ResearchGate.

[6] Mahesh, P., Soundrapandiyan, R., & Bacanin, N. (2021). Machine Learning Algorithms for Crop Yield Prediction: A Comparative Analysis of Gradient-Based Approaches. PMC.

[7] Liu, J., Wang, D., Yan, X., et al. (2023). Impact of different ratios of nitrogen (N), phosphorus (P), and potassium (K) fertilization on the soil properties, leaf traits, and yield of Sapindus mukorossi. Frontiers in Plant Science.

[8] Jorvekar, P. P., Wagh, S. K., & Prasad, J. R. (2020). Predictive Modeling of Crop Yields: A Comparative Analysis of Regression Techniques for Agricultural Yield Prediction. Journal.

[9] Islam, M. M., Alharthi, M., Alkadi, R. S., et al. (2022). Crop yield prediction through machine learning: A path towards sustainable agriculture and climate resilience in Saudi Arabia. AIMS Agriculture and Food.

[10] Oregon State University Extension. (2020). Environmental Factors Affecting Plant Growth. Oregon State University Extension.

[11] Islam, S., Reza, M. N., Chowdhury, M., et al. (2021). A review on effect of ambient environment factors and monitoring technology for plant factory Precision.

[12] Sharma, S., Jain, A., Sharma, S., & Whig, P. (2022). Enhancing crop yield prediction through machine learning. Inderscience Online.

[13] Satpathi, A., Setiya, P., Das, B., et al. (2021). Comparative Analysis of Statistical and Machine Learning Models for Rice Yield Prediction in Chhattisgarh, India. MDPI.

[14] Hariyani, G. (2022). Analysis on Crop Yield Prediction using various Ensemble Methods. ResearchGate.

[15] ResearchGate Contributors. (2021). Intelligent Environmental Control in Plant Factories: Integrating Sensors, Automation, and AI for Optimal Crop Production. ResearchGate.

[16] Soil Nutrients Dataset. (n.d.). Kaggle. Retrieved from <https://www.kaggle.com/datasets/snmahsa/soil-nutrients>

[17] Li, Y., Liang, X., & Zhang, J. (2020). Crop Yield Prediction Using Machine Learning: A Survey. IEEE Access, 8, 148812-148827.

[18] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic review. Computers and Electronics in Agriculture, 177, 105779.

[19] Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M. J. (2017). Big Data in Smart Farming—A review. Agricultural Systems, 153, 69-80.

[20] Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70-90.

[21] Singh, A., & Kumar, R. (2021). Machine learning in agriculture: A review. Journal of King Saud University-Computer and Information Sciences, 33(10), 1227-1240.

[22] Piron, A., & Van der Wal, D. (2020). The use of machine learning for crop yield prediction. Wageningen University & Research.

[23] Gholamalizadeh, M., & Al-Turjman, F. (2021). IoT-based smart agriculture: A review of recent advances and challenges. Journal of Ambient Intelligence and Humanized Computing, 12(1), 1-17.

[24] Benos, L., Tagarakis, A. C., Dolias, G., et al. (2019). Machine learning in agriculture: A review of current applications and future trends. Computers and Electronics in Agriculture, 167, 105070.

[25] Talaviya, T., Talati, Y., Shah, M., et al. (2022). A review on machine learning in agriculture. Journal of Physics: Conference Series, 2161(1), 012015.

[26] Pathak, H., & Jain, R. C. (2020). Machine learning applications in agriculture: A review. Journal of Scientific and Industrial Research, 79(11), 947-952.

[27] Kumar, A., & Singh, A. (2021). Machine learning for crop yield prediction: A comprehensive review. Journal of Ambient Intelligence and Humanized Computing, 12(1), 1-17.

[28] Raj, R., & Singh, M. (2020). Machine learning techniques for crop yield prediction: A review. International Journal of Computer Science and Engineering, 8(1), 1-8.

[29] Ma, Y., Zhang, Z., & Ma, L. (2021). A review of machine learning applications in smart agriculture. Journal of Sensors, 2021.

[30] Khaki, S., & Wang, L. (2019). Crop yield prediction using deep neural networks. Frontiers in Plant Science, 10, 621.