

Uluslararası Sürdürülebilir Mühendislik ve Teknoloji Dergisi
International Journal of Sustainable Engineering and Technology
ISSN: 2618-6055 / Cilt:9, Sayı:1, (2025), 70 – 79
Araştırma Makalesi – Research Article



TARIMSAL VERİMLİLİĞİ ARTIRMAK İÇİN AGRO-KLİMATİK KOŞULLARA BAĞLI OLARAK MAKİNE ÖĞRENMESİ TEMELLİ MAHSUL TAVSİYE SİSTEMİ

*Onur SEVLİ¹ 

¹Burdur Mehmet Akif Ersoy Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği
Burdur

(Geliş/Received: 19.05.2025, Kabul/Accepted: 17.06.2025, Yayınlanma/Published: 30.06.2025)

ÖZ

Tarım, dünya genelinde ekonomik ve toplumsal açıdan stratejik bir sektör olup, özellikle gıda güvenliği ve kırsal kalkınma açısından kritik bir role sahiptir. Bu sektörün sürdürülebilirliği, iklim değişikliği ve toprak özelliklerindeki farklılıklar gibi çok sayıda çevresel değişkenden etkilenmektedir. Çiftçilerin karşılaştığı en temel sorunlardan biri, mevcut agro-klimatik koşullara en uygun mahsulü belirleyebilmek ve bu doğrultuda üretim yapabilmektir. Bu çalışmanın temel amacı, makine öğrenmesi algoritmalarını kullanarak toprak ve çevre koşullarına göre en uygun mahsulü tahmin edebilen bir model geliştirmektir. Modelleme sürecinde, toprağın azot (N), fosfor (P), potasyum (K) içeriği ile sıcaklık, nem, pH değeri ve yağış miktarı gibi çevresel parametreler girdi olarak kullanılmıştır. Çalışmada Rastgele Orman, Lojistik Regresyon, Destek Vektör Makineleri, K-En Yakın Komşu ve Gradyan Artırma olmak üzere beş farklı makine öğrenmesi algoritması uygulanmıştır. Her bir modelin hiperparametrelerini optimize etmek için ızgara arama yaklaşımı ve modellerin doğruluğunu değerlendirmek için 5-kat çapraz doğrulama kullanılmıştır. Sonuçlara göre, %99.32'lik doğruluk oranı ile en başarılı model Rastgele Orman olmuştur. Diğer algoritmalar da yüksek doğruluk sağlamış olsa da Rastgele Orman'ın performansı belirgin şekilde öne çıkmıştır. Elde edilen sonuçlar, literatürde yer alan güncel çalışmalarla karşılaştırıldığında yöntemlerin etkinliğini ve uygulanabilirliğini ortaya koymaktadır. Bu çalışma, çiftçilerin veri destekli kararlar almasına olanak sağlayarak, tarımsal verimliliği ve kaynak kullanım etkinliğini artırabilecek akıllı bir mahsul tavsiye sistemi geliştirilmesine katkı sunmaktadır.

Anahtar kelimeler: Mahsul tavsiye sistemi, Makine öğrenmesi, Sınıflandırma

A MACHINE LEARNING-BASED CROP RECOMMENDATION SYSTEM BASED ON AGRO-CLIMATIC CONDITIONS TO ENHANCE AGRICULTURAL PRODUCTIVITY

ABSTRACT

Agriculture is a strategic sector worldwide from both economic and societal perspectives, playing a critical role especially in food security and rural development. The sustainability of this sector is influenced by numerous environmental variables such as climate change and variations in soil properties. One of the main challenges faced by farmers is identifying the most suitable crop for the prevailing agro-climatic conditions and producing accordingly. The primary objective of this study is to develop a model capable of predicting the most appropriate crop based on soil and environmental conditions using machine learning algorithms. In the modeling process, environmental parameters such as nitrogen (N), phosphorus (P), and potassium (K) content of the soil, along with temperature, humidity, pH level, and rainfall amount, were used as inputs. Five different machine learning algorithms were applied in this study, namely Random Forest, Logistic Regression, Support Vector Machines, K-Nearest Neighbor and Gradient Boosting. The grid search approach was used to optimize each model's

hyperparameters, and 5-fold cross-validation was used to assess the models' accuracy. According to the results, the most successful model was Random Forest with an accuracy of 99.32%. Although the other algorithms also achieved high accuracy, the performance of the Random Forest model stood out significantly. When compared with recent studies in the literature, the findings demonstrate the effectiveness and applicability of the proposed methods. This study contributes to the development of an intelligent crop recommendation system that can enhance agricultural productivity and resource use efficiency by enabling farmers to make data-driven decisions.

Keywords: Crop recommendation system, Machine learning, Classification

1. Giriş (Introduction)

Tarım sektörü, dünya nüfusunun beslenme ihtiyacını karşılamanın yanı sıra birçok ülke için önemli bir gelir kaynağıdır. Küresel nüfusun sürekli artışı, gıda güvenliğini sağlamak amacıyla tarımsal üretimde düzenli bir artışı zorunlu kılmaktadır [1]. Geleneksel tarım yöntemleri, iklim değişikliğinin ve artan çevresel baskıların etkisiyle verimlilikte düşüş yaşanma risklerine açıktır. İklim değişikliği, toprak verimliliğindeki azalmalar ve kaynakların sınırlı olması gibi faktörler tarımsal üretimi olumsuz etkilemektedir. Bu zorlukların üstesinden gelmek ve sürdürülebilir tarım uygulamalarını teşvik etmek için teknolojik yeniliklere ihtiyaç duyulmaktadır [2]. Bu bağlamda, sürdürülebilir ve verimli tarım uygulamaları, artan talebi karşılarken çevreye olumsuz etkiyi en aza indirmek için hayati öneme sahiptir. Hassas tarım, tarımsal girdileri ve uygulamalarını optimize etmek için veri odaklı bir yaklaşım sunarak bu ihtiyaca cevap vermektedir. Bu uygulamalar veri odaklı kararlar alarak kaynak kullanımını optimize etmeyi ve verimliliği artırmayı hedefler. Bu bağlamda, makine öğrenmesi teknikleri, tarımsal verilerden anlamlı bilgiler çıkararak çiftçilere rehberlik etme potansiyeline sahiptir [3], [4].

Çiftçilerin en kritik kararlarından biri, arazilerinin ve iklim koşullarının özelliklerine en uygun mahsulü seçmektir. Yanlış mahsul seçimi, düşük verim, kaynak israfı ve ekonomik kayıplara neden olur. Geleneksel yöntemlerle mahsul seçimi genellikle deneyime veya sınırlı bilgiye dayanır. Ancak, toprak özellikleri (mineraller, pH), iklimsel faktörler (sıcaklık, nem, yağış) ve mahsulün özel gereksinimleri gibi birçok karmaşık faktörün bir arada değerlendirilmesi gerekmektedir. Makine öğrenmesi algoritmaları, bu tür çok değişkenli verileri analiz ederek belirli koşullar için en uygun mahsulü tahmin edebilen modeller geliştirmeyi sağlar [5].

Teknolojinin yaygınlaşması ve yapay zekâ tekniklerinin gelişmesi ile birlikte son dönemde tarım alanında makine öğrenmesi algoritmaları kullanılarak yapılan çalışmaların sayısı giderek artmaktadır. Toprak besin maddeleri (Azot, Fosfor, Potasyum), sıcaklık, nem, pH ve yağış gibi agro-klimatik parametreler, mahsulün büyümesini ve verimini önemli ölçüde etkileyen temel faktörlerdir. Mohapatra ve Kale [6] agro-klimatik ve toprakla ilgili bir dizi parametre kullanarak çeşitli makine öğrenmesi teknikleri ile tahminlemeler gerçekleştirmiştir. Yağış, nem, toprak tipi, pH ve sıcaklık parametreleri üzerinde K-En Yakın Komşu (KNN), Yapay Sinir Ağları (YSA), Rastgele Orman (RO) ve Destek Vektör Makineleri (DVM) algoritmaları kullanılarak gerçekleştirilen sınıflandırmalarda KNN algoritması ile %97.95 doğruluğa ulaşmışlardır.

Madhuri ve Indiramma [7]; minimum, maksimum ve ortalama sıcaklık, güneşlenme süresi, toprak dokusu, toprak pH'ı, çakıl durumu, erozyon, eğim, derinlik ve toprak drenajı parametreleri kullanarak dört mahsul türünü (mısır, darı, pirinç, şeker kamışı) tahminlemeye çalışmışlardır. YSA kullanarak gerçekleştirdikleri analizlerde %96 doğruluk elde etmişlerdir. Bu çalışmalarında, çeşitli toprak ve iklim faktörlerini mahsul tavsiyesi için entegre etmede YSA'nın etkinliğini vurgulamaktadırlar.

Acharya vd. [8], üre, fosfor, potasyum, pH, sıcaklık gibi verilere dayalı olarak çiftçilere en uygun mahsulü öneren bir sistem sunulmuştur. 10 farklı mahsul içeren 10 bin örnekli veri seti üzerinde Naive Bayes (NB), DVM, Karar Ağacı (KA), ve RO algoritmaları ile sınıflandırmalar gerçekleştirmişlerdir. Her bir modeli bağımsız olarak değerlendirdikten sonra tümünün birleşimi ile ortaya koydukları hibrit modeli test etmişler ve %98.98 doğruluk elde etmişlerdir.

Deepika ve Andrew [9]; azot, fosfor, potasyum oranları, sıcaklık, nem, pH ve yağış miktarı parametrelerini kullanarak ortamda yetişmesi muhtemel 22 farklı mahsul türünü tahmin etmeye çalışmışlardır. Çok katmanlı algılayıcılarla (ÇKA) gerçekleştirdikleri sınıflandırmalarda %97 doğruluk değerine ulaşmışlardır. ÇKA'nın doğrusal olmayan ilişkileri modelleme ve çeşitli veri türlerinden

öğrenme yeteneği bu modeli geleneksel makine öğrenmesi modellerine nazaran daha ideal bir çözüm haline getirmektedir.

Ramzan vd. [10], nesnelerin interneti teknolojisi kullanarak gerçekleştirdikleri çalışmalarında sensörlerden gerçek zamanlı olarak toplanan sıcaklık, iletkenlik, toprak su ihtiyacı ve dokusu gibi parametrelere ek olarak azot, fosfor, yağış bilgisi gibi parametreleri içeren ikincil bir veri seti kullanarak uygun mahsulün seçimine yönelik bir tahminleme çalışması gerçekleştirmişlerdir. Beş farklı algoritma kullanarak gerçekleştirdikleri analizlerde YSA ile %93.63 doğruluğa ulaşmışlardır.

Bhavani vd. [11], toprak mineral değerleri, pH, yağış, sıcaklık ve nem ölçümlerinden oluşan 7 özellik ve 2200 adet örnekli veri seti üzerinde Lojistik Regresyon (LR), DVM, KNN, Karar ağacı, RO, Naive Bayes (NB) algoritmaları yanında iki gizli katmanlı bir derin sinir ağı kullanmışlardır. En düşük doğruluğu %75 ile NB modeli ile en yüksek doğruluğu ise %97 ile derin sinir ağı ile elde etmişlerdir. Elde edilen bulgular derin sinir ağlarının toprak ve çevre arasındaki karmaşık ilişkileri geleneksel makine öğrenmesi yöntemlerine göre daha iyi ortaya koyduğunu göstermektedir.

Bu çalışmada çeşitli agro-klimatik parametreleri girdi olarak kullanarak en uygun mahsulü tahmin edebilen makine öğrenimi tabanlı bir sınıflandırma modelinin geliştirilmesi amaçlanmıştır. Bu doğrultuda, yaygın olarak kullanılan beş farklı sınıflandırma algoritması – Rastgele Orman, Lojistik Regresyon, Destek Vektör Makineleri, K-En Yakın Komşu ve Gradyan Artırma – uygulanmış, hiperparametreleri optimize edilmiş ve performansları kapsamlı şekilde değerlendirilmiştir. Çalışmanın sonuçları, benzer literatür çalışmalarıyla karşılaştırılarak mevcut araştırmanın tarım alanındaki makine öğrenimi uygulamalarına katkısı tartışılmıştır.

2. Materyal ve Metot (Material and Method)

2.1. Veri seti (Dataset)

Bu çalışmada kullanılan veri seti, çeşitli agro-klimatik parametrelere dayanarak en uygun mahsulün tahmin edilmesini amaçlayan bir yapıya sahiptir. Kamuya açık bir veri deposu olan Kaggle'dan elde edilmiştir. Veri seti, hassas tarım ve çiftçilere, tarım danışmanlarına ve politika yapıcılara destek olmayı amaçlayan makine öğrenimi uygulamalarında tipik olarak kullanılmaktadır. Veri setinde yer alan öznitelikler Tablo 1'de verilmiştir.

Tablo 1. Veri setinin öznitelikleri (Features of the dataset)

Öznitelik Adı	Açıklama	Birimi
N	Topraktaki azot içeriği	mg/kg
P	Topraktaki fosfor içeriği	mg/kg
K	Topraktaki potasyum içeriği	mg/kg
temperature	Ortalama sıcaklık	°C
humidity	Ortalama bağıl nem	%
ph	Toprak pH değeri	sayısal
rainfall	Yağış miktarı	Mm
label	Hedef değişken	22 mahsul türü

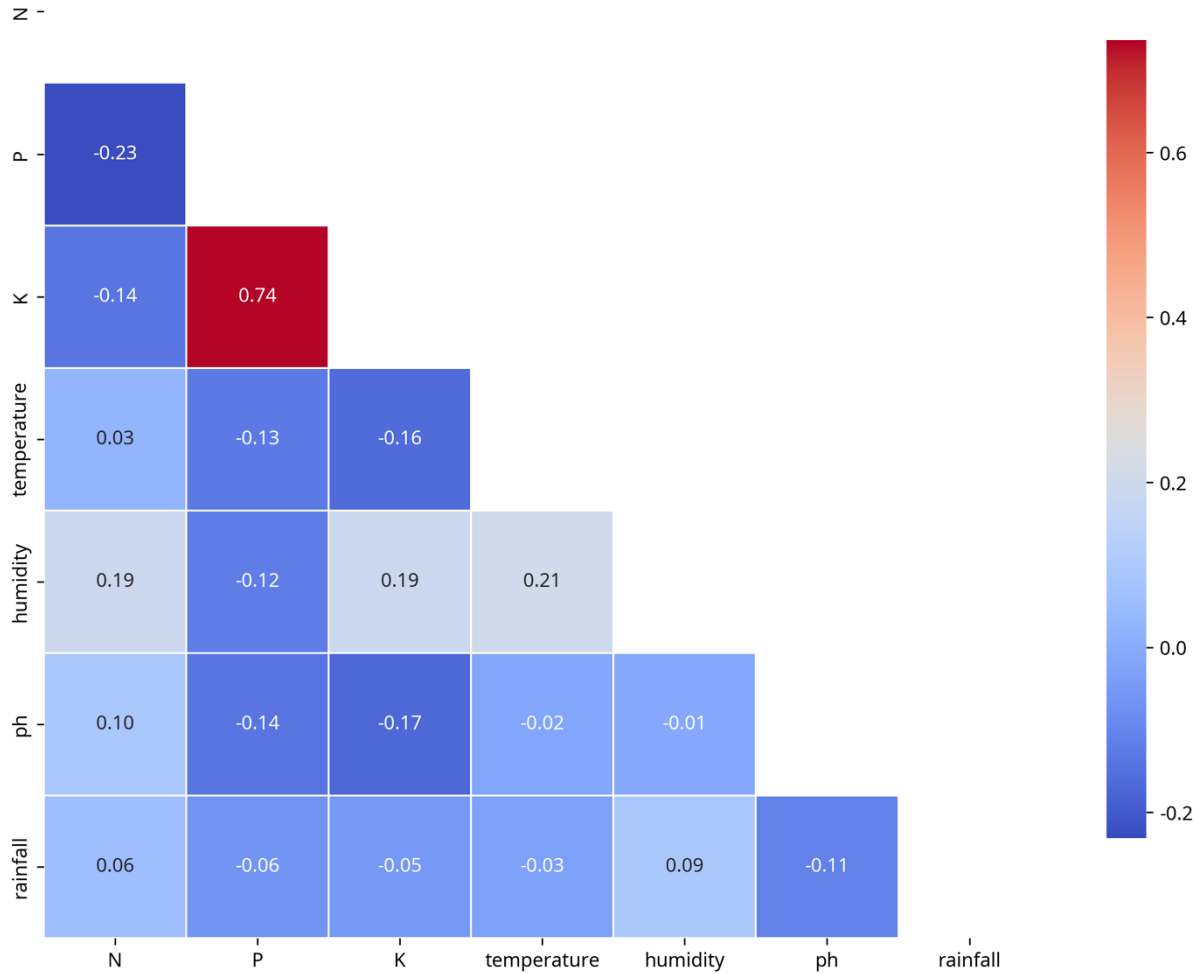
Özniteliklerden ilk 7 adedi girdi ve label özniteliği tahminlenecek hedef özellik olmak üzere toplam 8 öznitelik yer almaktadır. Girdi özniteliklerin istatistiki özeti Tablo 2'de yer almaktadır.

Tablo 2. Özniteliklerin istatistiki karakteristiği (Statistical characteristics of the features)

Öznitelik Adı	Minimum	Maximum	Ortalama	Standart Sapma
N	0	140	50.55	36.92
P	5	145	53.36	32.99
K	5	205	48.15	50.65
temperature	8.83	43.68	25.62	5.06
humidity	14.26	99.98	71.48	22.26
ph	3.5	9.94	6.47	0.77

rainfall 20.21 298.56 103.46 54.96

Veri setinde yer alan sayısal girdi öz nitelikler arasındaki korelasyon Şekil 1’de ısı haritası olarak verilmiştir.



Şekil 1. Korelasyon matrisi (Correlation matrix)

Şekil 1’de verilen korelasyon matrisi incelendiğinde öz nitelikler arasında genel olarak yüksek bir korelasyon söz konusu değildir. Yalnız toprakta bulunan fosfor (P) ve potasyum (K) oranları arasında 0.74’lük bir korelasyon olduğu görülmektedir.

Hedef özellik 22 farklı mahsul türünü içermektedir. Veri setinde yer alan mahsuller Tablo 3’te listelenmiştir.

Tablo 3. Veri setinde yer alan mahsul adları (Crop names in the dataset)

Mahsul adı	Mahsul adı
Apple (elma)	Maize (mısır)
Banana (muz)	Mango (mango)
Blackgram (siyah nohut)	Mothbeans (güve fasulyesi)
Chickpea (nohut)	Mungbean (maş fasulyesi)
Coconut (Hindistan cevizi)	Muskmelon (kavun)
Coffee (kahve)	Orange (portakal)
Cotton (pamuk)	Papaya (papaya)
Grapes (üzüm)	Pigeonpeas (güvercin bezelyesi)
Jute (jüt bitkisi)	Pomegranate (nar)
Kidneybeans (barbunya)	Rice (pirinç)
Lentil (mercimek)	Watermelon (karpuz)

Veri seti içerisinde 22 farklı mahsul yer almakta ve her bir mahsule ait 100 adet örnek bulunmakta olup, veri seti toplam 2200 örnekten oluşmaktadır. Hedef sınıfların her birine ait eşit sayıda örnek bulunmasından dolayı veri seti dengeli bir dağılıma sahiptir.

2.2. Veri ön işleme (Data preprocessing)

Veri ön işleme, makine öğrenmesi modellerinin performansını doğrudan etkileyen kritik bir adımdır. Bu çalışmada uygulanan ön işleme adımları şunlardır:

1. Eksik Veri Kontrolü: Veri seti, eksik değerler açısından incelenmiş ve herhangi bir eksik değere rastlanmamıştır.
2. Kategorik Veri Kodlaması: Hedef değişken olan 'label' (mahsul türleri) kategorik yapıda olduğundan, algoritmalarının işleyebilmesi için sayısal formata dönüştürülmüştür. Bu amaçla etiket kodlama (label encoding) kullanılmıştır.
3. Öznitelik Ölçeklendirme: Farklı ölçeklerdeki sayısal özniteliklerin model performansı üzerindeki olumsuz etkisini gidermek ve özellikle KNN ve DVM gibi mesafeye duyarlı algoritmaların performansını artırmak için öznitelik ölçeklendirme uygulanmıştır.

2.3. Kullanılan makine öğrenmesi modelleri (The machine learning models employed)

Bu çalışmadaki mahsul sınıflandırma problemi için literatürde de yaygın olarak kullanılan aşağıdaki beş makine öğrenmesi modeli kullanılmıştır:

Rastgele Orman (RO), birden fazla karar ağacının bir araya gelerek oluşturduğu bir topluluk (ensemble) öğrenme yöntemidir ve sınıflandırma ile regresyon problemlerinde sıklıkla kullanılır. Bu yöntem, her ağaç için farklı rastgele örnekler ve özellik alt kümeleri seçerek ağaçlar arasında çeşitlilik sağlar; bu da modelin genelleme yeteneğini artırır [12]. RO, özellikle yüksek boyutlu veri setlerinde, çok sayıda özellik içeren durumlarda etkili sonuçlar sunar ve genellikle aşırı öğrenmeye karşı dirençli yapısıyla bilinir. Ayrıca, modelin her bir özelliğe verdiği önemi belirleyebilmesi, özellikle açıklanabilirlik açısından değerlidir. Bu nedenle sağlık, finans ve biyoinformatik gibi pek çok alanda tercih edilmektedir [13].

Lojistik Regresyon (LR), bağımlı değişkenin kategorik olduğu durumlarda kullanılan temel istatistiksel modelleme yöntemlerinden biridir. Sıklıkla ikili sınıflandırma problemlerinde kullanılsa da, çok sınıflı genişletmeleri ile çoklu sınıflandırma problemlerinde de etkili şekilde uygulanabilir [14]. Modelin en önemli avantajı, tahmin edilen olasılıkların yorumlanabilir olmasıdır; bu, özellikle tıp ve sosyal bilimler gibi karar destek sistemlerinin önem taşıdığı alanlarda kullanışlıdır. LR, yapısı gereği doğrusal sınırlar çizse de uygun dönüşümlerle doğrusal olmayan problemler için de genişletilebilir. Basitliği ve hesaplama verimliliği sayesinde büyük veri setlerinde bile hızlı sonuçlar verebilir.

Destek Vektör Makineleri (DVM), örnekler arasındaki maksimum marjini sağlayacak hiperdüzlemi bulmayı amaçlayan, denetimli bir öğrenme algoritmasıdır. Bu özelliği sayesinde genelleme yeteneği oldukça yüksektir ve özellikle az sayıda örnekle yüksek boyutlu uzaylarda etkili sonuçlar verir [15]. DVM, doğrusal ayırıcıların yetersiz kaldığı durumlarda çekirdek (kernel) fonksiyonları sayesinde doğrusal olmayan ayırımlar da yapabilir. DVM'nin dezavantajlarından biri, büyük veri setlerinde eğitim süresinin uzun olabilmesidir; ancak doğruluk açısından çoğu zaman tatmin edici performans sağlar.

K-En Yakın Komşu (KNN), sınıflandırma ve regresyon problemlerinde kullanılan, örneklerin doğrudan özellik uzayındaki yakınlığına göre sınıflandırıldığı parametrik olmayan bir algoritmadır. Herhangi bir öğrenme süreci gerektirmemesi, yani “lazy learning” olması nedeniyle basit ve uygulaması kolaydır [16]. KNN, veri noktaları arasındaki mesafeye dayanarak karar verir ve bu nedenle özelliklerin uygun şekilde ölçeklendirilmesi oldukça kritiktir. Gürültülü verilere karşı duyarlıdır ve özellikle yüksek boyutlu veri setlerinde performansı düşebilir, bu da “boyutsallık laneti” olarak bilinir [17]. Bununla birlikte, doğru parametre (K değeri) seçildiğinde etkili ve güvenilir sonuçlar verebilir.

Gradyan Artırma (GA), zayıf öğrenicileri — genellikle karar ağaçlarını — ardışık olarak eğiterek hata değerini en aza indirmeye çalışan güçlü bir topluluk öğrenme yöntemidir. Her yeni model, bir öncekinin hata oranını azaltacak şekilde eğitilir ve bu süreç, modelin genel doğruluğunu artırır [18]. Bu teknik, sınıflandırma ve regresyon gibi birçok denetimli öğrenme probleminde yüksek performans sergiler. GA,

overfitting riskine karşı dikkatli hiperparametre ayarları gerektirir ancak düzenleme teknikleriyle bu risk azaltılabilir. Günümüzde XGBoost, LightGBM ve CatBoost gibi modern GA tabanlı kütüphaneler, makine öğrenmesi yarışmalarında sıklıkla birincilik alan modellerin temelini oluşturur [19].

2.4. Hiperparametre optimizasyonu (Hyperparameter optimization)

Modellerin performansını en üst düzeye çıkarmak için hiperparametre optimizasyonu uygulanmıştır. Her bir model için hiperparametre aralıkları tanımlanmış ve ızgara arama yöntemi kullanılarak en iyi hiperparametre kombinasyonları belirlenmiştir. Optimizasyon sürecinde 5 kat çapraz doğrulama ve doğruluk metriği esas alınmıştır.

3. Deneyisel Çalışma ve Bulgular (Experimental Study and Findings)

Çeşitli agro-klimatik değerlere bağlı olarak yetiştirilmeye uygun mahsul tahminlemesine yönelik 7 girdi ve 1 hedef olmak üzere 8 adet öznitelik ve 2200 örnekten oluşan veri seti üzerinde literatürde yaygın kullanılan Rastgele Orman (RO), Lojistik Regresyon (LR), Destek Vektör Makineleri (DVM), K-En Yakın Komşu (KNN) ve Gradyan Artırma (GA) algoritmaları kullanılarak sınıflandırmalar gerçekleştirilmiştir. Her bir algoritmanın ideal parametrelerine ulaşmak için gerçekleştirilen optimizasyon süreci sonrası tespit edilen hiperparametre değerleri Tablo 4'te verilmiştir.

Tablo 4. Modellerin optimize edilmiş hiperparametreleri (Optimized hyperparameters of the models)

Model	Optimize hiperparametreler
RO	'max_depth': 10, 'min_samples_leaf': 1, 'min_samples_split': 2, 'n_estimators': 200
LR	'C': 100, 'max_iter': 100, 'solver': 'lbfgs'
DVM	'C': 10, 'gamma': 'scale', 'kernel': 'linear'
KNN	'metric': 'manhattan', 'n_neighbors': 5, 'weights': 'distance'
GA	'learning_rate': 0.2, 'max_depth': 3, 'n_estimators': 200

Hiperparametre optimizasyonundan sonra elde edilen en iyi modellerin genelleme performansını daha güvenilir bir şekilde değerlendirmek amacıyla 5 kat çapraz doğrulama uygulanmıştır. Optimize edilmiş hiperparametrelerle yapılandırılan modellerin performansları farklı metrikler açısından değerlendirilmiş ve ortalama skorlar raporlanmıştır.

Model performanslarının değerlendirilmesinde temel olarak karışıklık matrislerinden (confusion matrix) yararlanılmıştır. Bir karışıklık matrisi kullanılan sınıflandırma modelinin sınıfları birbirinden ayırt etmedeki başarısını ölçmeye yardımcı olur. Genel olarak bir karışıklık matrisinin yapısı Şekil 2'de görülmektedir.

		Gerçek sınıf	
		Pozitif	Negatif
Tahmin edilen	Pozitif	Doğru Pozitif (DP)	Yanlış Pozitif (YP)
	Negatif	Yanlış Negatif (YN)	Doğru Negatif (DN)

Şekil 2. Karışıklık matrisi

Sınıflandırma modelinin Pozitif olarak tahmin ettiği değer gerçekte de Pozitif ise Doğru Pozitif (DP), gerçekte Negatif ise Yanlış Pozitif (YP) olarak adlandırılır. Benzer şekilde modelin Negatif olarak tahmin ettiği değer gerçekte de Negatif ise Doğru Negatif (DN), gerçekte Pozitif ise Yanlış Negatif (YN) olarak adlandırılır. Bu değerlere bağlı olarak üretilen doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 Skor değerleri Tablo 5'te verilen şekilde hesaplanır.

Tablo 5. Model performans metrikleri (Model performance metrics)

Metrik	Açıklama	Formül
Doğruluk (Accuracy)	Doğru sınıflandırılan örneklerin toplam örneklere oranıdır.	$(DP + DN) / (DP + DN + YP + YN)$
Kesinlik (Precision)	Pozitif olarak tahmin edilen örneklerden kaçının gerçekten pozitif olduğunun oranıdır.	$DP / (DP + YP)$
Duyarlılık (Recall)	Gerçekten pozitif olan örneklerden kaçının doğru bir şekilde pozitif olarak tahmin edildiğinin oranıdır.	$DP / (DP + YN)$
F1 Skoru (F1-Score)	Kesinlik ve duyarlılığın harmonik ortalamasıdır, iki metrik arası dengeyi gösterir.	$2 * (Kesinlik * Duyarlılık) / (Kesinlik + Duyarlılık)$

Optimize edilmiş RO modeli için 5 kat çapraz doğrulama işlemi ile elde edilen ortalama değerler Tablo 6’da verilmiştir.

Tablo 6. Rastgele Orman ile elde edilen sonuçlar (Results obtained using Random Forest)

Metrik	Ortalama Değer (%)
Doğruluk	99.32
Kesinlik	99.35
Duyarlılık	99.32
F1 Skoru	99.33

Optimize edilmiş LR modeli için 5 kat çapraz doğrulama işlemi ile elde edilen ortalama değerler Tablo 7’de verilmiştir.

Tablo 7. Lojistik Regresyon ile elde edilen sonuçlar (Results obtained using Logistic Regression)

Metrik	Ortalama Değer (%)
Doğruluk	98.41
Kesinlik	98.52
Duyarlılık	98.41
F1 Skoru	98.46

Optimize edilmiş DVM modeli için 5 kat çapraz doğrulama işlemi ile elde edilen ortalama değerler Tablo 8’de verilmiştir.

Tablo 8. Destek Vektör Makineleri ile elde edilen sonuçlar (Results obtained using Support Vector Machines)

Metrik	Ortalama Değer (%)
Doğruluk	98.86
Kesinlik	98.97
Duyarlılık	98.86
F1 Skoru	98.91

Optimize edilmiş KNN modeli için 5 kat çapraz doğrulama işlemi ile elde edilen ortalama değerler Tablo 9’da verilmiştir.

Tablo 9. K-En Yakın Komşu ile elde edilen sonuçlar (Results obtained using K-Nearest Neighbors)

Metrik	Ortalama Değer (%)
--------	--------------------

Doğruluk	98.18
Kesinlik	98.23
Duyarlılık	98.18
F1 Skoru	98.20

Optimize edilmiş GA modeli için 5 kat çapraz doğrulama işlemi ile elde edilen ortalama değerler Tablo 10'da verilmiştir.

Tablo 10. Gradyan Artırma ile elde edilen sonuçlar (Results obtained using Gradient Boosting)

Metrik	Ortalama Değer (%)
Doğruluk	98.64
Kesinlik	98.67
Duyarlılık	98.64
F1 Skoru	98.65

Kullanılan 5 farklı sınıflandırma modeli için elde edilen skorlar özet olarak Tablo 11'de verilmiştir.

Tablo 11. Tüm modeller ile elde edilen sonuçlar (Results obtained from all models)

Model	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru
RO	99.32	99.35	99.32	99.33
LR	98.41	98.52	98.41	98.46
DVM	98.86	98.97	98.86	98.91
KNN	98.18	98.23	98.18	98.20
GA	98.64	98.67	98.64	98.65

Bu çalışmada test edilen beş farklı modelin, mahsul tavsiyesi probleminde dikkate değer bir performans gösterdiği görülmektedir. Tüm metrikler açısından en yüksek skorları veren RO modeli, %99.32 doğruluk sağlamıştır. Bu durum, RO algoritmasının topluluk öğrenme yapısı sayesinde varyansı azaltma ve aşırı öğrenmeyi engelleme yeteneği ile açıklanabilir. Karar ağaçlarının birleşimi, farklı özellik etkileşimlerini yakalayarak karmaşık ilişkileri modellemede etkili olmuştur. DVM ve GA modelleri sırayla %98.86 ve %98.64 doğruluk sağlamışlardır. Bu modeller tüm metrikler açısından %98.5'in üzerinde oranlarla RO modeline yakın bir performans sergilemişlerdir. DVM'nin 'linear' çekirdek ile yüksek performans göstermesi, veri setindeki sınıfların doğrusal olarak ayırt edilebilirliğinin yüksek olduğuna işaret etmektedir. LR ve KNN modelleri diğerlerine nazaran daha az başarı gösterse de, %98'in üzerindeki skorları ile oldukça etkili oldukları söylenebilir. Özellikle LR'nin basitliği ve yorumlanabilirliği, bazı pratik uygulamalarda tercih sebebi olabilir.

Bu çalışmada elde edilen bulgular literatürdeki benzer güncel çalışmalarla karşılaştırılarak Tablo 12'de verilmiştir.

Tablo 12. Literatür karşılaştırması (Comparison with the literature)

Çalışma	Veri seti özellikleri	Kullanılan Model	Doğruluk
Mohapatra ve Kale (2021) [6]	Yağış, nem, toprak tipi, pH ve sıcaklık	KNN	%97.95
Madhuri ve Indiramma (2021) [7]	Sıcaklık, güneşlenme süresi, pH, çakıl durumu, erozyon, eğim, derinlik ve toprak drenajı	YSA	%96
Musanese vd. (2023) [20]	Azot, fosfor, potasyum, pH, yağış, sıcaklık, nem	YSA	%97
Acharya vd. (2024) [8]	Üre, fosfor, potasyum, pH, sıcaklık	Hibrit(NB+DVM+KA+RO)	%98.98
Deepika ve Andrew (2024) [9]	Azot, fosfor, potasyum, pH, yağış, sıcaklık, nem	Çok katmanlı algılayıcı	%97
Ramzan vd. (2024) [10]	Sıcaklık, iletkenlik, toprak su ihtiyacı ve dokusu, azot, fosfor, yağış bilgisi	YSA	%93.63
Bhavani vd. (2024) [11]	Toprak mineral değerleri, pH, yağış, sıcaklık ve nem	Naive Bayes	%75

Afzal vd. (2025) [2]	Azot, fosfor, potasyum, pH, yağış, sıcaklık, nem	RO+XGBoost	%98
Bu çalışma	Azot, fosfor, potasyum, pH, yağış, sıcaklık, nem	Optimize RO	%99.32

Tablo 12 incelendiğinde literatürdeki çalışmaların hemen hemen benzer özelliklere sahip veri setleri üzerinde çalıştığı görülmektedir. Literatürdeki çalışmalarda genel olarak hatırı sayılır başarıların elde edildiği görülmektedir. Bu çalışmaya en yakın doğruluk oranı %98 ile Afzal vd. [2] tarafından yapılan çalışmada elde edilmiştir. Afzal vd. [2] çalışmalarında RO ve XGBoost birleşimi hibrit bir algoritma kullanmışlardır. Bu çalışmada da RO Modeli ile en yüksek doğruluk değerine ulaşılmıştır. Hibrit bir model kullanılmamasına karşın model parametrelerinin optimize edilmesi geleneksel modellere nazaran daha yüksek bir başarı elde edilmesini sağlamıştır. Literatürdeki çalışmaların geneli ile karşılaştırıldığında bu çalışmada elde edilen %99.32 oranındaki doğruluk diğer çalışmalarla rekabet edebilecek düzeyde ve görece daha yüksektir.

Elde edilen yüksek başarının yanında bu çalışmanın sınırlılık olarak kabul edilecek bazı yanları da bulunmaktadır. Kullanılan veri seti belirli bir coğrafi bölgeye ait olduğu için daha geniş bir çeşitliliğe sahip toprak ve iklim koşullarına uygun olmayabilir. Bu nedenle modellerin farklı bölgelerdeki ölçümlerle test edilerek genelleme kabiliyetinin sınanması gerekebilir. Modellerin performansı mevcut veri setine bağlıdır ve veri setindeki olası gürültü veya yanlışlıklar sonuçları etkileyebilir. Bu çalışma statik bir veri seti üzerinde gerçekleştirilmiştir; dinamik olarak değişen iklim koşulları ve toprak özellikleri için modellerin zamanla güncellenmesi gerekebilir.

4. Sonuç (Conclusion)

Bu çalışmada, yedi farklı agro-klimatik özelliğe (N, P, K, sıcaklık, nem, pH, yağış) dayanarak 22 farklı mahsul türü için en uygun olanı tahmin etmek amacıyla beş farklı makine öğrenmesi sınıflandırma modeli (Rastgele Orman, Lojistik Regresyon, Destek Vektör Makineleri, K-En Yakın Komşu ve Gradyan Artırma) başarıyla uygulanmış ve karşılaştırılmıştır. Modellerin hiperparametreleri optimize edilmiş ve performansları doğruluk, kesinlik, duyarlılık ve F1 skoru metrikleri kullanılarak değerlendirilmiştir. Elde edilen sonuçlar, tüm modellerin yüksek performans gösterdiğini, özellikle Rastgele Orman modelinin %99.32 doğruluk oranı ile en iyi sonucu verdiğini ortaya koymuştur. Bu bulgular, makine öğrenmesinin tarımda karar verme süreçlerini desteklemek ve mahsul verimliliğini artırmak için güçlü bir araç olabileceğini göstermektedir. Çalışmanın sonuçları, literatürdeki benzer çalışmalarla da uyumludur ve Rastgele Orman gibi topluluk öğrenme yöntemlerinin bu tür sınıflandırma problemlerinde etkinliğini bir kez daha teyit etmektedir.

Geliştirilen modeller, çiftçilere kendi arazileri ve iklim koşulları için en uygun mahsulü seçme konusunda bilimsel temelli tavsiyeler sunarak, kaynakların daha verimli kullanılmasına ve tarımsal üretimin sürdürülebilirliğine katkıda bulunma potansiyeline sahiptir. Gelecekteki çalışmalar, daha kapsamlı veri setleri ve ek özellikler kullanarak bu modelleri daha da geliştirmeye odaklanabilir.

5. Kaynaklar (References)

- [1] H.C.J. Godfray, J.R. Beddington, I.R. Crute, L. Haddad, D. Lawrence, J.F. Muir, et al., Food security: the challenge of feeding 9 billion people, *Science* 327 (5967) (2010) 812–818, doi:10.1126/science.1185383.
- [2] H. Afzal, M. Ahmad, A. Aftab, et al., Incorporating soil information with machine learning for crop recommendation to improve agricultural output, *Sci. Rep.* 15 (1) (2025) 8560, doi:10.1038/s41598-025-88676-z.
- [3] R. Parvathi, Crop recommendation system by artificial neural network, (2021).
- [4] D.N. Varshitha, S. Choudhary, An artificial intelligence solution for crop recommendation, *Indones. J. Electr. Eng. Comput. Sci.* 25 (3) (2022) 1688–1695.
- [5] F.S. Prity, M.A. Islam, T. Rahman, et al., Enhancing agricultural productivity: a machine learning approach to crop recommendations, *Hum.-Centric Intell. Syst.* 4 (1) (2024) 497–510, doi:10.1007/s44230-024-00081-3.

- [6] B.N. Mohapatra, V. Kale, Crop recommendation system using machine learning, ITEGAM-JETIA 10 (48) (2024) 63–68.
- [7] J. Madhuri, M. Indiramma, Artificial neural networks based integrated crop recommendation system using soil and climatic parameters, Indian J. Sci. Technol. 14 (19) (2021) 1587–1597.
- [8] N. Acharya, P. Khatiwada, R. Pandey, S. Niroula, P. Chapagain, Crop recommendation system using machine learning: a comparative study, in: Proc. KEC Conf. 2024, (2024) 302–311, doi:10.3126/injet.v1i2.66708.
- [9] M.P. Deepika, G. Andrew, Artificial neural network for crop recommendation, Grenze Int. J. Eng. Technol. 10 (2024) 3933–3936.
- [10] S. Ramzan, A. Khan, A. Hussain, et al., An innovative artificial neural network model for smart crop prediction using sensory network based soil data, PeerJ Comput. Sci. 10 (2024) e2478.
- [11] R. Bhavani, V. Agalya, N. Kiruthika, C. Sugapriya, Deep neural network based crop recommendation system, Int. Res. J. Eng. Technol. (IRJET) 11 (4) (2024) 278–281.
- [12] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, doi:10.1023/A:1010933404324.
- [13] X. Chen, H. Ishwaran, Random forests for genomic data analysis, Genomics 99 (6) (2012) 323–329, doi:10.1016/j.ygeno.2012.04.003.
- [14] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, Applied Logistic Regression, third ed., Wiley, 2013, doi:10.1002/9781118548387.
- [15] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (3) (1995) 273–297, doi:10.1007/BF00994018.
- [16] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1) (1967) 21–27, doi:10.1109/TIT.1967.1053964.
- [17] K. Beyer, J. Goldstein, R. Ramakrishnan, U. Shaft, When is 'nearest neighbor' meaningful?, in: Int. Conf. Database Theory, (1999) 217–235, doi:10.1007/3-540-49257-7_15.
- [18] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. 29 (5) (2001) 1189–1232, doi:10.1214/aos/1013203451.
- [19] T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., (2016) 785–794, doi:10.1145/2939672.2939785.
- [20] C. Musanase, A. Vodacek, D. Hanyurwimfura, A. Uwitonze, I. Kabandana, Data-driven analysis and machine learning-based crop and fertilizer recommendation system for revolutionizing farming practices, Agriculture 13 (11) (2023) 2141.