

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**MAHSUL VERİMLİLİĞİ TAHMİNİ VE PLANLAMA**

**BİTİRME PROJESİ**

Bilgisayar Mühendisliği Bölümü

**DANIŞMAN**

Dr. Ali SARIKAŞ

İSTANBUL, 2025

logo, yazı tipi, simge, sembol, grafik içeren bir resim

Açıklama otomatik olarak oluşturuldu

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

**MAHSUL VERİMLİLİĞİ TAHMİNİ VE PLANLAMA**

**BİTİRME PROJESİ**

Bilgisayar Mühendisliği Bölümü

**DANIŞMAN**

Dr. Öğr. Üyesi Ali SARIKAŞ

İSTANBUL, 2025

**MARMARA ÜNİVERSİTESİ**

**TEKNOLOJİ FAKÜLTESİ**

**BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ**

Marmara Üniversitesi Teknoloji Fakültesi Bilgisayar Mühendisliği öğrencisi Vahit Seyit Başa tarafından hazırlanan “Mahsul Verimliliği Tahmini ve Planlama” başlıklı proje çalışması, 19 Haziran 2025 tarihinde savunulmuş ve jüri üyeleri tarafından başarılı bulunmuştur.

**Jüri Üyeleri**

|  |  |  |
| --- | --- | --- |
| Dr. Öğr. Üyesi Ali SARIKAŞ  Marmara Üniversitesi | **(Danışman)** | (İMZA)………….. |
| Doç. Dr. Kazım YILDIZ  Marmara Üniversitesi | (Üye) | (İMZA)………….. |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**ÖNSÖZ**

Bu bitirme projesi, mühendislik eğitimim boyunca edindiğim teorik ve uygulamalı bilgileri, tarımsal üretim süreçlerinde karşılaşılan önemli bir probleme çözüm üretmek üzere kullanma fırsatı sunmuştur. Çalışmamın odak noktası, modern teknolojilerin (uydu görüntüleri, görüntü işleme, coğrafi bilgi sistemleri ve yapay zekâ algoritmaları) entegre edilerek, mahsul verimliliği tahmini ve tarım planlamasına katkı sağlayacak bir karar destek sistemi geliştirmektir.

Projenin kavramsal tasarımından uygulama sürecine kadar olan tüm aşamalarında hem mühendislik disiplinlerinin kesişim noktalarında çalışma yapma hem de çok disiplinli düşünebilme yetkinliği kazanılmıştır. Bu süreçte, özellikle verinin toplanması, işlenmesi ve anlamlı çıktılara dönüştürülmesinde karşılaşılan teknik zorlukların üstesinden gelinerek, problem çözme ve proje yönetimi becerileri pekiştirilmiştir.

Danışmanım Doç. Dr. Ali SARIKAŞ'a, akademik rehberliği ve değerli katkıları için teşekkür ederim. Bu proje, yalnızca akademik bir gerekliliği yerine getirmekten öte, mühendislik yetkinliklerimin pratiğe dönüşmesini sağlayan kapsamlı bir deneyim olmuştur.

**İÇİNDEKİLER**

**1. GİRİŞ ........................................................................................................................... 1**

**1.1. Proje Amacı ve Önemi ..................................................................................... 1**

**1.2. Proje Kapsamı ................................................................................................... 2**

**1.3. Kullanılan Yöntemler ve Teknolojiler ...................................................... 4**

**2. LİTERATÜR TARAMASI ..................................................................................... 6**

**2.1. Uydu Görüntüleme ve NDVI Endeksi ....................................................... 6**

**2.2. Yapay Zekâ ve Derin Öğrenmenin Tarıma Uygulanması ..................... 6**

**2.3. Coğrafi Bilgi Sistemleri (GIS) ile Tarımsal Analiz .................................... 7**

**3. YÖNTEM VE MATERYALLER ........................................................................... 8**

**3.1. Kullanılan Veriler ve Kaynaklar .................................................................. 8**

**3.2. Google Earth Engine Uygulaması ............................................................. 10**

**3.3. Veri İşleme ve Özellik Çıkarımı ................................................................... 10**

**4. UYGULAMA VE SONUÇLAR ............................................................................. 11**

**4.1. Veri Seti ve Özellik Seçimi ............................................................................. 11**

**4.2. Model Eğitimi ve Çapraz Doğrulama ....................................................... 11**

**4.3. Model Performans Karşılaştırması (R², RMSE, MAE) ........................... 38**

**4.4. Özellik Önem Düzeyleri ................................................................................ 14**

**4.5. Sonuçların Tartışılması ................................................................................ 15**

**6. KAYNAKÇA .............................................................................................................18**

**ÖZET**

Bu bitirme projesi, tarım alanında verimliliği artırmak ve üretim süreçlerini daha bilimsel temellere dayandırmak amacıyla geliştirilmiş bir karar destek sistemi tasarımı ve uygulamasını kapsamaktadır. Çalışmada, özellikle mısır mahsulü üzerine odaklanılarak, Iowa eyaletine ait 2018-2023 yılları arasındaki uydu görüntüleri kullanılmış ve bu görüntüler USDA verileri ile etiketlenerek eğitim veri seti oluşturulmuştur.

Projede; Google Earth Engine platformu üzerinden NDVI (Normalized Difference Vegetation Index) ve benzeri spektral göstergeler çıkarılmış, bu göstergeler Python tabanlı veri işleme araçlarıyla analiz edilmiştir. Görüntü işleme, Coğrafi Bilgi Sistemleri (GIS) ve derin öğrenme yöntemleri entegre edilerek, farklı dönemlerdeki verimlilik farklılıkları modellenmiş ve doğruluk oranları üzerinden değerlendirme yapılmıştır.

Çalışmanın nihai çıktısı olarak geliştirilen sistem, çiftçilere ekim zamanlaması, sulama ve gübreleme planlaması gibi kararlarında destek sağlayabilecek; böylece hem kaynak kullanımı optimize edilecek hem de daha sürdürülebilir bir tarım anlayışı geliştirilecektir.

Bu yönüyle çalışma, tarımda dijital dönüşüme katkı sunmayı hedeflemekte ve ileri teknoloji destekli karar verme süreçlerinin pratikteki uygulanabilirliğini göstermektedir.

**Haziran, 2025 Öğrenciler**

**ABSTRACT**

This undergraduate thesis focuses on the design and implementation of a decision support system developed to enhance efficiency in agricultural production and promote a more scientifically grounded approach to planning. The project specifically targets corn yield prediction using satellite imagery of the Iowa region between 2018 and 2023, with datasets labeled using USDA crop statistics.

Through Google Earth Engine, spectral indices such as NDVI (Normalized Difference Vegetation Index) were derived and analyzed using Python-based data processing tools. The study integrates image processing, Geographic Information Systems (GIS), and deep learning models to analyze temporal yield patterns and evaluate model accuracy based on performance metrics.

The resulting system aims to assist farmers in making informed decisions on optimal sowing times, irrigation schedules, and fertilization strategies. As a result, it is expected to improve resource allocation and promote sustainable agricultural practices.

This work contributes to the digital transformation in agriculture and demonstrates the real-world applicability of technology-assisted decision-making systems in crop yield estimation and farm planning.

**June, 2025 Students**

**KISALTMALAR**

**AI** : Artificial Intelligence (Yapay Zekâ)  
**API** : Application Programming Interface (Uygulama Programlama Arayüzü)  
**CNN** : Convolutional Neural Network (Evrilimli Sinir Ağı)  
**CSV** : Comma Separated Values (Virgülle Ayrılmış Değerler)  
**DL** : Deep Learning (Derin Öğrenme)  
**GCP** : Google Cloud Platform  
**GEE** : Google Earth Engine  
**GIS** : Geographic Information Systems (Coğrafi Bilgi Sistemleri)  
**IoT** : Internet of Things (Nesnelerin İnterneti)  
**JS** : JavaScript  
**JSON** : JavaScript Object Notation  
**LSTM** : Long Short-Term Memory  
**MAE** : Mean Absolute Error (Ortalama Mutlak Hata)  
**ML** : Machine Learning (Makine Öğrenmesi)  
**NDVI** : Normalized Difference Vegetation Index  
**NOAA** : National Oceanic and Atmospheric Administration  
**OS** : Operating System (İşletim Sistemi)  
**PDF** : Portable Document Format  
**PTB**-XL : Physikalisch-Technische Bundesanstalt ECG veri seti  
**Python** : Programlama dili  
**RF** : Random Forest  
**RGB** : Red Green Blue (Renk bantları)  
**RMSE** : Root Mean Square Error (Karekök Ortalama Hata)  
**SVM** : Support Vector Machine (Destek Vektör Makineleri)  
**UI** : User Interface (Kullanıcı Arayüzü)  
**USDA** : United States Department of Agriculture (ABD Tarım Bakanlığı)  
**XAI** : Explainable Artificial Intelligence (Açıklanabilir Yapay Zekâ)QGIS : Quantum **GIS** : (Açık kaynak coğrafi bilgi sistemi)  
**Sentinel** : ESA tarafından sağlanan uydu verileri ailesi  
**MODIS** : Moderate Resolution Imaging Spectroradiometer  
**Cloud** Mask: Bulut tespiti ve maskeleme algoritması  
**BBOX** : Bounding Box (Sınır kutusu – görüntü işleme)  
**ROI** : Region of Interest (İlgi Alanı)

**ŞEKİL LİSTESİ**

[**Şekil 1.1** Uzaktan algılama tabanlı mahsul verim tahmin sistemine ait genel iş akışı. Bu grafik, sistemin temel bileşenlerini ve veri akışını göstermektedir.2](#_heading=h.1fob9te)

[**Şekil 2.1**](#_heading=h.4d34og8) Uydu Görüntülerinde RGB, NDVI ve GNDVI Tabanlı Görselleştirme Örnekleri

# GİRİŞ

Tarım sektörü, dünya genelinde artan nüfus ve azalan doğal kaynaklar nedeniyle sürdürülebilir kalkınmanın temel alanlarından biri haline gelmiştir. Özellikle iklim değişikliği, su kaynaklarının azalması ve toprak verimliliğindeki düşüş gibi çevresel tehditler, geleneksel tarım yöntemlerinin yetersiz kalmasına neden olmaktadır. Bu bağlamda, teknoloji destekli yeni yaklaşımlar geliştirerek tarımsal verimliliğin artırılması ve kaynakların etkin kullanımı büyük önem taşımaktadır.

Bu çalışmada, tarım arazilerinde mahsul verimliliğini tahmin edebilen, çevresel koşulları analiz edebilen ve çiftçilere karar desteği sunabilen bir sistemin geliştirilmesi hedeflenmiştir. Bu sistem, özellikle mısır üretimi gibi yüksek hacimli ve iklime duyarlı tarımsal faaliyetlerde planlama süreçlerini desteklemeye yöneliktir. Uydu görüntüleri, iklim verileri ve coğrafi bilgi sistemleri bir araya getirilerek, geçmiş yıllara ait veriler üzerinden modelleme yapılmış; böylece geleceğe yönelik tahminler elde edilmeye çalışılmıştır.

Bu bağlamda proje kapsamında, 2018–2023 yılları arasında Iowa Eyaleti’ne ait Sentinel uydu verileri kullanılmış, Normalized Difference Vegetation Index (NDVI) gibi spektral bant analizleri gerçekleştirilmiş ve bu veriler ABD Tarım Bakanlığı'nın (USDA) sağladığı resmi verilerle etiketlenerek zenginleştirilmiştir. Analiz sürecinde Google Earth Engine, Python, QGIS gibi güçlü platformlar kullanılmış ve çeşitli makine öğrenmesi algoritmaları (Random Forest, LSTM, CNN vb.) ile verim tahmin modelleri eğitilmiştir.

Proje yalnızca akademik bir çalışmanın ötesinde, tarımsal planlama, karar destek sistemleri, iklim-ürün ilişkileri ve dijital tarım teknolojileri açısından uygulanabilir ve geliştirilebilir bir çözüm sunmaktadır. Bu yönüyle hem mühendislik bakış açısıyla ele alınmış hem de sosyoekonomik etkileri göz önünde bulundurulmuştur.

Bir sonraki bölümde, bu projeye temel oluşturan mevcut literatür çalışmaları ve benzer uygulamalara yer verilmiştir.

## Proje Çalışmasının Amacı ve Önemi

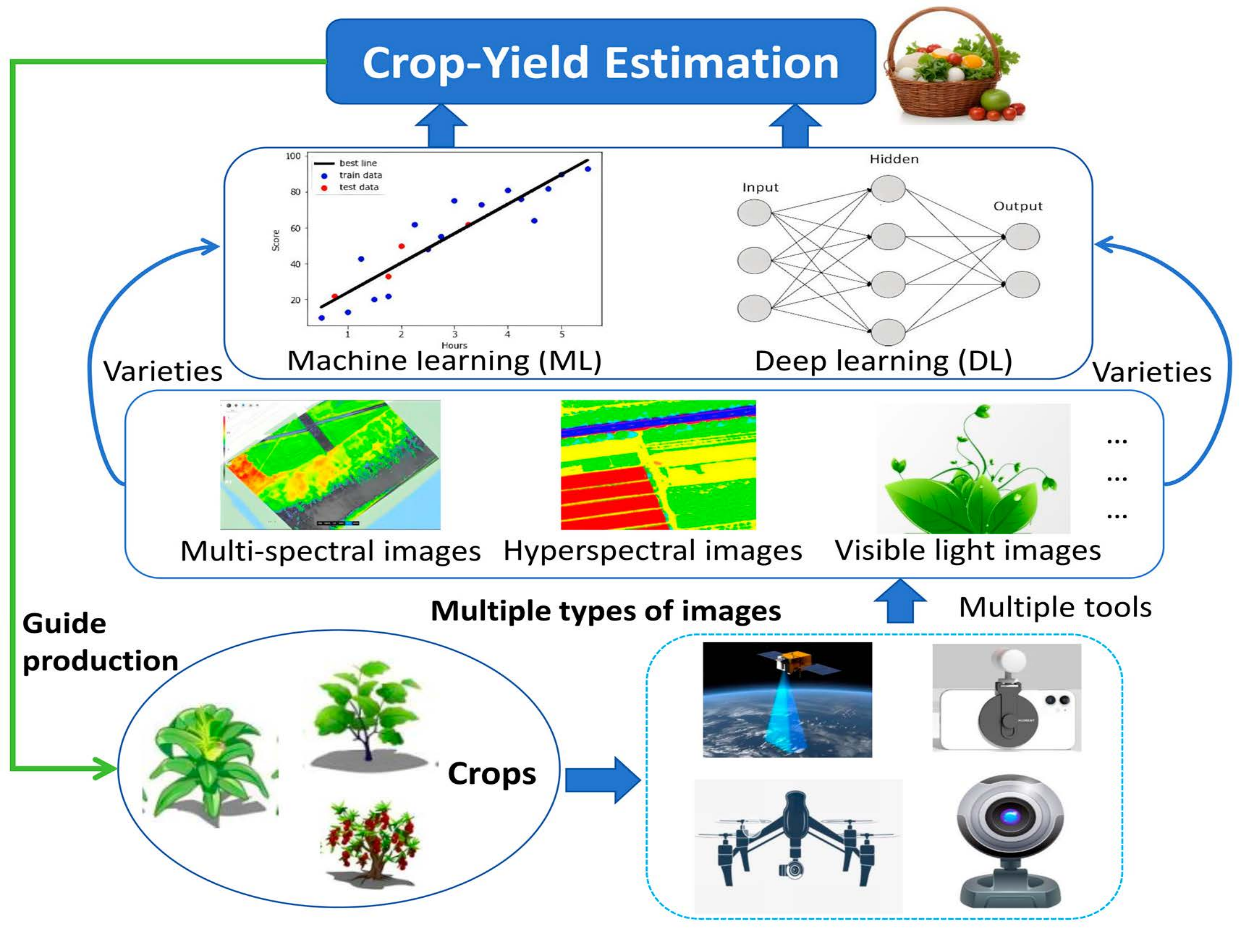
Bu bitirme projesinin temel amacı, tarımsal üretim alanlarında mahsul verimliliğini hasat öncesinde yüksek doğrulukla tahmin edebilecek bir karar destek sistemi geliştirmektir. Geliştirilecek sistem; uydu görüntüleri, görüntü işleme teknikleri, makine öğrenmesi algoritmaları ve coğrafi bilgi sistemleri (GIS) gibi teknolojileri entegre ederek, çiftçilere ve tarım sektörü profesyonellerine zamanında, yerinde ve etkili karar alma imkânı sunacaktır.

Verim tahmini yalnızca üretim miktarını önceden görmek açısından değil; gübreleme, sulama, ilaçlama, hasat planlaması ve pazarlama süreçlerinin her biri için stratejik bir öneme sahiptir. Bu sistemle birlikte, özellikle küçük ve orta ölçekli üreticilerin girdi maliyetlerini optimize etmesi, verim artışı sağlaması ve çevresel sürdürülebilirliğe katkıda bulunması hedeflenmektedir.

Bu çalışmanın önemi aşağıdaki başlıklarla özetlenebilir:

* **Tarımda Dijitalleşme:** Modern görüntü işleme, makine öğrenmesi ve uydu teknolojilerinin bir arada kullanılmasıyla tarımda dijitalleşmenin yaygınlaştırılması amaçlanmaktadır.
* **Erken Müdahale İmkânı:** Mahsul durumu erken tespit edilerek hastalık, su stresi veya besin yetersizliği gibi durumlara karşı zamanında önlem alınması sağlanacaktır.
* **Ekonomik Verimlilik:** Girdi maliyetlerinin azaltılması, kaynakların etkin kullanımı ve verim artışıyla ekonomik fayda elde edilmesi beklenmektedir.
* **Sürdürülebilirlik:** Toprak, su ve enerji gibi doğal kaynakların daha bilinçli kullanımıyla çevre üzerindeki baskı azaltılacaktır.
* **Planlama ve Politika Geliştirme:** Tarımsal planlama, üretim stratejileri ve bölgesel kalkınma politikalarının oluşturulmasında kullanılabilecek anlamlı veriler üretilecektir.

Proje aynı zamanda, veri bilimi ve uzaktan algılama teknolojilerinin tarımsal uygulamalara uyarlanması yoluyla disiplinlerarası bilgi üretimini ve teknolojik uygulamaların sahaya aktarılmasını da desteklemektedir.



**Şekil 1.1** Uzaktan algılama tabanlı mahsul verim tahmin sistemine ait genel iş akışı. Bu grafik, sistemin temel bileşenlerini ve veri akışını göstermektedir.

**1.2 Proje Kapsamı**

Bu bitirme projesi, tarımsal üretim süreçlerinde karar destek sağlayabilecek bir mısır verim tahmin sisteminin geliştirilmesini kapsamaktadır. Sistem, çok kaynaklı verilerin (uzaktan algılama görüntüleri, iklim verileri ve resmi tarımsal kayıtlar) entegrasyonuna dayalıdır ve makine öğrenmesi algoritmaları ile derin öğrenme mimarileri kullanılarak tasarlanmıştır. Projenin genel kapsamı; veri temini, veri ön işleme, modelleme, analiz ve sistem çıktılarının değerlendirilmesi aşamalarını içermektedir.

Çalışma alanı olarak **Amerika Birleşik Devletleri’nin Iowa Eyaleti** seçilmiştir. Iowa, hem geniş ölçekli tarım alanlarına hem de zengin kamuya açık tarımsal veri kaynaklarına sahip olması nedeniyle, verim tahmin sistemlerinin geliştirilmesi ve doğrulanması açısından uygun bir bölge sunmaktadır. Proje süresince 2018–2023 yıllarına ait veriler kullanılmıştır.

Projenin kapsamı aşağıdaki teknik ve operasyonel alt başlıklar çerçevesinde belirlenmiştir:

* **1. Çok Kaynaklı Veri Entegrasyonu**

Proje kapsamında dört temel veri kaynağı kullanılmıştır:

* **Sentinel-2 Uydu Görüntüleri:** ESA tarafından sağlanan, 10 metrelik çözünürlüğe sahip multispektral görüntüler, tarımsal gelişimin izlenmesinde kullanılmıştır.
* **USDA NASS Verileri:** ABD Tarım Bakanlığı'nın yıllık üretim ve verim istatistikleri, ilçe düzeyinde verim değerleri sağlamaktadır.
* **CDL (Cropland Data Layer):** Uydu verilerindeki piksel bazlı tarım ürün sınıflandırmaları, mekânsal etiketleme işlemleri için kullanılmıştır.
* **NASA POWER API:** Günlük sıcaklık, yağış, rüzgar, güneş radyasyonu ve nem gibi iklimsel değişkenler sağlanarak zaman serisi analizine olanak tanımıştır.
* **2. Veri Ön İşleme ve Analiz Hazırlığı**

Toplanan ham veriler, modellemeye uygun hâle getirilmek amacıyla bir dizi ön işleme aşamasından geçirilmiştir:

* Uydu görüntülerinde bulut maskesi uygulanmış ve normalize edilmiştir.
* İklim verileri yıllık ortalama veya kümülatif değerlere dönüştürülerek zaman serisine uygun hâle getirilmiştir.
* Tüm veri kaynakları, ilçe ve yıl düzeyinde eşleştirilerek tekil bir analiz tablosu haline getirilmiştir.
* **3. Modelleme ve Tahmin Süreci**

Verim tahmini için iki temel yaklaşım benimsenmiştir:

* **Zamansal Modelleme:** NDVI, sıcaklık, yağış gibi yıl içindeki değişkenlik gösteren faktörleri değerlendirmek için LSTM (Long Short-Term Memory) modeli kullanılmıştır.
* **Görüntü Tabanlı Alan Tespiti:** Uydu görüntülerinden tarımsal alanların otomatik olarak tespiti ve sınıflandırılması amacıyla YOLO (You Only Look Once) mimarisi kullanılmıştır.

Ayrıca, bu modellerin çıktılarını birleştiren bir ensemble yaklaşımı geliştirilmiş ve genel tahmin doğruluğunun artırılması hedeflenmiştir.

* **4. Uygulama Potansiyeli ve Uyarlanabilirlik**

Geliştirilen sistem, sadece akademik doğruluk hedefleriyle değil, aynı zamanda tarımda pratik olarak uygulanabilirliği dikkate alınarak tasarlanmıştır.

* İlçe bazlı tahmin yapılabilmesiyle bölgesel tarım planlamalarına katkı sunabilir.
* Modüler yapısı sayesinde farklı ürün türleri veya coğrafi alanlara kolayca uyarlanabilir.
* Kullanılan veri kaynaklarının büyük ölçüde kamuya açık ve küresel olması, sistemin evrensel kullanılabilirliğini artırmaktadır.
* **5. Bilimsel Katkı**

Bu proje, uydu verisi, iklim verisi ve resmi üretim istatistiklerini entegre ederek, hem zamansal hem de mekânsal boyutları dikkate alan bütüncül bir verim tahmin sistemi geliştirmektedir. Ayrıca makine öğrenmesi ve derin öğrenme yöntemlerinin tarımsal karar destek sistemlerinde nasıl uygulanabileceğini örnekleyen bir akademik model sunmaktadır.

**1.3 Kullanılan Yöntemler ve Teknolojiler**

Bu projede, farklı kaynaklardan elde edilen büyük hacimli ve çeşitli formattaki verilerin işlenmesi, analiz edilmesi ve anlamlı çıktılara dönüştürülmesi amacıyla hem geleneksel hem de modern yapay zekâ yöntemleri kullanılmıştır. Kullanılan yöntemler, veri tipine ve problem doğasına göre belirlenmiş; her bir analiz adımı için en uygun algoritmalar ve teknolojiler entegre edilmiştir.

* **1.3.1 Yöntemler**

**1. Zaman Serisi Analizi (LSTM)**  
Zamansal bileşenleri yoğun olan iklim verileri ve bitki gelişim indeksleri (NDVI, EVI) üzerinden ürün verimliliğini tahmin edebilmek amacıyla, tekrarlayan sinir ağlarının (RNN) gelişmiş bir versiyonu olan LSTM (Long Short-Term Memory) mimarisi kullanılmıştır. Bu model, yıllar içindeki sıcaklık, yağış, radyasyon ve nem gibi değişkenlerin mahsul gelişimi üzerindeki etkilerini öğrenerek verim çıktısını tahmin etmektedir.

**2. Görüntü Tabanlı Nesne Tanıma (YOLO)**  
Uydu görüntüleri üzerinde mısır ekim alanlarının belirlenmesi amacıyla, gerçek zamanlı nesne tanıma algoritmalarından olan YOLO (You Only Look Once) kullanılmıştır. YOLO, konvolüsyonel sinir ağı (CNN) temelli bir yaklaşımla, çok bantlı görüntülerde tarım arazilerini otomatik olarak algılayarak mekânsal verim tahmini için temel sınıflandırma sağlar.

**3. Ensemble Öğrenme**  
Zamansal model (LSTM) ile mekânsal model (YOLO) çıktıları birleştirilerek daha yüksek doğruluk oranı ve model genel performansı elde edilmiştir. Ensemble yöntemi, farklı model türlerinden gelen tahminlerin bir araya getirilmesiyle hata payını azaltmayı ve daha kararlı sonuçlar üretmeyi amaçlamaktadır.

**4. Korelasyon ve Özellik Seçimi**  
Modelleme öncesinde, değişkenler arasında anlamlı ilişkileri ortaya koymak amacıyla korelasyon analizi yapılmıştır. Ayrıca, fazla değişkenliğin modelde aşırı öğrenmeye (overfitting) yol açmaması için bazı düşük etkili özellikler dışlanmış ve veri boyutu azaltılmıştır.

* **1.3.2 Teknolojiler**

**1. Google Earth Engine (GEE)**  
Sentinel-2 uydu görüntülerinin alınması, ön işlenmesi ve spektral indeks hesaplamaları için kullanılan bulut tabanlı jeo-uzamsal analiz platformudur. GEE, yüksek hacimli uydu verilerini kısa sürede işleyerek zaman serisi üretimine olanak tanımıştır.

**2. NASA POWER API**  
İklim verilerinin günlük bazda otomatik olarak toplanması için kullanılmıştır. Sıcaklık, nem, rüzgar, güneş radyasyonu gibi meteorolojik değişkenler bu API aracılığıyla elde edilerek ilçe düzeyinde organize edilmiştir.

**3. Python Programlama Dili**  
Tüm veri işleme, modelleme ve analiz adımlarında ana programlama dili olarak kullanılmıştır. Veri temizleme (pandas, numpy), görselleştirme (matplotlib, seaborn), model eğitimi (tensorflow, keras) gibi işlemler Python üzerinden yürütülmüştür.

**4. TensorFlow ve scikit-learn**

* TensorFlow: LSTM ve YOLO gibi derin öğrenme modellerinin kurulması ve eğitilmesi için kullanılmıştır.
* Scikit-learn: Korelasyon analizi, veri standardizasyonu ve klasik makine öğrenmesi algoritmaları için tercih edilmiştir.

**5. QGIS ve PostGIS**

* QGIS: Mekânsal verilerin haritalanması ve görselleştirilmesi için kullanılmıştır.
* PostGIS: Veri tabanı işlemleri ve coğrafi verilerin depolanmasında PostgreSQL ile birlikte çalışmıştır.

Bu yöntem ve teknolojiler, projenin hem doğruluk hem de uygulanabilirlik hedeflerini desteklemiş; akademik olarak yenilikçi ve sahada uyarlanabilir bir sistemin ortaya çıkmasını sağlamıştır.

# LİTERATUR TARAMASI

Modern tarımda ürün verimliliğini tahmin etmek, gıda güvenliği, iklim değişikliği ve sürdürülebilirlik hedefleriyle birlikte giderek artan önem kazanmıştır. Bu kapsamda, uzaktan algılama, yapay zekâ ve coğrafi bilgi sistemlerinin entegrasyonu birçok çalışma için temel strateji haline gelmiştir. Aşağıda alt başlıklar halinde bu alanlara ait gelişmeler ve mühendislik yaklaşımlar ayrıntılı olarak incelenmiştir.

#### 2.1. Uydu Görüntüleme ve NDVI Endeksi

Uydu görüntüleme sistemleri, özellikle Sentinel‑2 ve MODIS gibi çok bantlı sensörler, tarımsal alanları dönemsel olarak izlemek üzere yaygın olarak kullanılmaktadır. Bu görüntüler üzerinden hesaplanan NDVI (Normalized Difference Vegetation Index), bitki örtüsünün yoğunluğunu, fotosentez kapasitesini ve sezon içindeki biyokütle gelişimini nicel olarak ifade eder.

NDVI, Rouse ve iş arkadaşları tarafından tanımlanmış olup; kırmızı ve yakın‑kızılötesi bantların normalleştirilerek hesaplanması yoluyla elde edilir. Bu indeksin büyüme döneminin ortasında elde edilen maksimum değeri, mahsul verim tahmini için güçlü bir öngörü aracı olarak tanımlanmıştır. NASA ve USDA destekli çalışmalarda, pik NDVI’ya dayalı basit modeller bile mısırda R² ≈ 0.88, yoğun endeks toplama yöntemleriyle R² ≈ 0.93 gibi yüksek korelasyonlar sağladığı belirtilmiştir [mdpi.com](https://www.mdpi.com/2072-4292/13/21/4227?utm_source=chatgpt.com). Ayrıca toprak nem durumu, atmosferik etkiler veya gölgelenme gibi faktörlerin etkisini azaltmak için EVI, SAVI gibi alternatif indeksler geliştirilmiş ve bazı senaryolarda bu indekslerin daha üstün performans gösterdiği bilimsel olarak rapor edilmiştir.

NDVI zaman serisi eğrileri sayesinde üç döneme ait (erken, orta, geç mevsim) mahsul fenolojisi izlenebilir. Özellikle orta ve geç büyüme dönemlerindeki eğim ve maksimum seviyeler, sezon sonu verim düzeylerini net şekilde yansıtmaktadır. Bu bilgiyi analiz eden modeller, zaman içinde zamansal trendleri öğrenerek erken dönem verim tahminlerine olanak tanımaktadır .

#### 2.2. Yapay Zekâ ve Derin Öğrenmenin Tarıma Uygulanması

Zaman serisi verileri ile çalışan LSTM (Long Short‑Term Memory) modelleri, uzun dönemli bağımlılıkları yakalayarak mahsul gelişimini analiz etmede etkili olarak görülmüştür. Örneğin iklim değişkenleriyle beslenen LSTM modeller, erken dönemdeki veriler üzerinden sezon sonu üretimi yüksek doğrulukla tahmin edebilmektedir. Yapılan çalışmalarda, LSTM modellerinin klasik regresyonlara kıyasla %10–15 daha yüksek performans sergilediği belirtilmiştir .

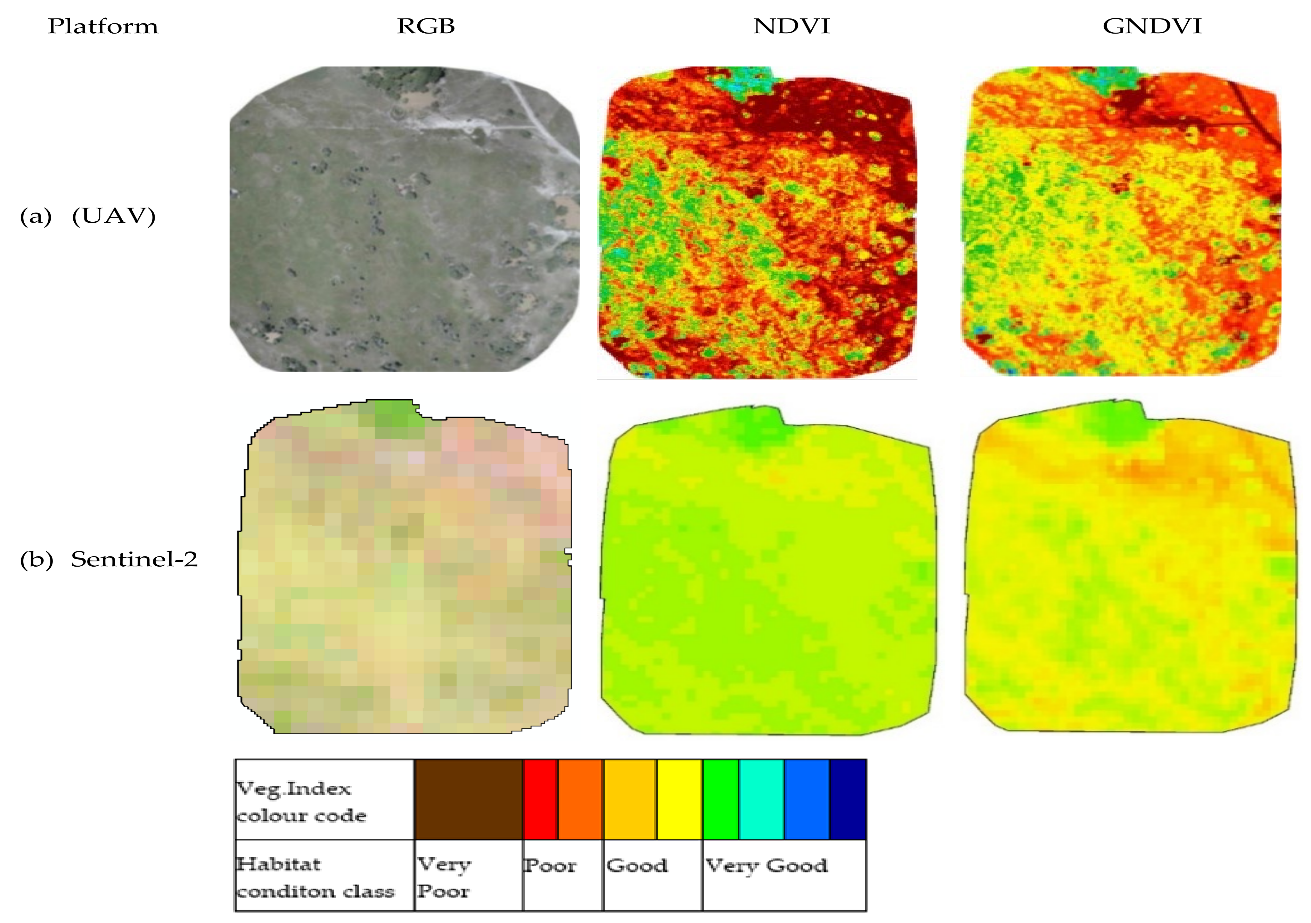
Yapay zekâ alanında nesne tespiti için kullanıma giren YOLO (You Only Look Once) mimarisi, tarımsal uydu görüntülerinde ekim alanlarının otomatik olarak belirlenmesi için tercih edilen bir modeldir. YOLO‑tabanlı modeller, satranç tahtası benzeri grid yapısı sayesinde uydu görüntülerine tek aşamada sınıflandırma ve konum tespiti uygulayabilir. Bu yöntem, Sentinel‑2 gibi yüksek çözünürlüklü görüntüler üzerinde iyi bir hız-doğruluk dengesi sunar. Tarımsal kullanım örneklerinde, özellikle küçük alanların bile tanınabilmesi açısından varyantları (YOLO‑RS, YOLO‑v5, YOLO‑v8) tercih edilmiştir. Bu modellerde Harita during gübreleme, erken hastalık tespiti gibi süreçleri destekleyecek düzeyde hassasiyet elde edileceği belirtilmiştir .

Ensemble yaklaşımları, mekânsal model olan YOLO ve zamansal model olan LSTM çıktılarının birleştirilmesine dayanır. Bu yöntem, değişken veri kaynaklarının sinerjistik etkisiyle hata payını düşürerek daha sağlam tahminler üretir. Literatürde, bu birleşik yaklaşımla elde edilen modellerin doğruluğunun R² > 0.90 gibi yüksek değerlere ulaştığı rapor edilmiştir .

#### 2.3. Coğrafi Bilgi Sistemleri (GIS) ile Tarımsal Analiz

Coğrafi Bilgi Sistemleri, elde edilen tüm verileri coğrafi referanslarla entegre ederek mekânsal analizlere olanak tanır. GIS platformları kullanılarak ilçe ve tarla bazlı verim haritaları, iklim değişimlerine göre bölgesel analizler, toprak verileri gibi verilerin detaylı görselleştirmeleri yapılmaktadır. Bu analizler, hem karar destek sistemlerine veri sağlaması hem de tarımsal yönetim süreçlerini optimize etmesi açısından değerlidir.

GIS’in zaman serileri bazlı haritalandırma fonksiyonları, erken dönem veriler ile sezon sonu çıktı tahmini arasında korelasyon kurulmasına yardımcı olur. Araştırmalar, uzaktan algılama indeksleri ile GIS destekli mekânsal analizlerin birleştirilmesiyle elde edilen verim haritalarının yönetsel karar süreçlerine önemli katkı sağladığını vurgulamaktadır .



**Şekil 2.1** Uydu Görüntülerinde RGB, NDVI ve GNDVI Tabanlı Görselleştirme Örnekleri

### İlgili Çalışmalar

1. NDVI'nın mısır biyokütlesi ve verim ile korelasyonu üzerine çalışmalar.
2. LSTM modelleri kullanılarak iklim verisiyle mahsul verim tahminleri.
3. YOLO tabanlı uydu görüntüsü analizlerinde tarlaların tespiti.
4. Ensemble mimarilerle verim tahmini modelleri.
5. MODIS NDVI kullanılarak Amerikan tarım alanlarında verim hesaplamaları.
6. Çoklu bitki indeksleri ve zaman serisi bütünleşik modellerle ülke düzeyinde tahmin yaklaşımları.

**3. YÖNTEM VE MATERYALLER**

* **3.1 Neden Bu Veri Kaynakları ve Yöntemler Seçildi?**

**Sentinel-2 Uydu Verileri**

Sentinel-2, Avrupa Uzay Ajansı tarafından sağlanan yüksek mekânsal (10–20 m) ve spektral çözünürlüğe sahip multispektral veriler sunmaktadır. Özellikle tarım alanlarının izlenmesi için kırmızı, yakın kızılötesi (NIR) ve kısa dalga kızılötesi (SWIR) bantları içerdiğinden, bitki sağlığının, stresin ve gelişim sürecinin analizinde NDVI, EVI, GNDVI gibi indekslerin hesaplanmasına olanak tanır. Bu veriler sayesinde büyüme sezonu boyunca mısır tarlalarının fenolojik evreleri güvenilir şekilde izlenebilmiştir

➡ **Katkısı**:

* Tarlaların zaman içinde yeşil alan gelişimini (biyokütle birikimi) değerlendirme,
* Erken dönem stres (kuraklık, besin eksikliği) tespiti,
* Model için zaman serili bitki indeksleri üretimi.

**NASA POWER API Verileri**

Bu veri kaynağı, günlük olarak hava durumu değişkenlerini (sıcaklık, nem, yağış, güneş ışınımı, rüzgâr hızı vb.) ücretsiz olarak sağlamaktadır. İlçe düzeyinde iklim zaman serileri oluşturmak için uygundur. Tarımda verimlilik büyük oranda hava koşullarına bağlı olduğu için bu veriler, NDVI indeksleriyle birlikte modelin önemli açıklayıcı değişkenleri

➡ **Katkısı**:

* Büyüme sezonlarındaki sıcaklık ve yağış modelleri üzerinden verim eğilimlerinin tahmini,
* Aşırı hava olayları ile ürün stresinin korelasyonu,
* Tarla bazlı yönetim stratejileri için mikroklimatik veri katkısı.

**USDA NASS Verileri**

Bu veri seti ABD Tarım Bakanlığı’nın sunduğu, yıllık bazda county (ilçe) düzeyinde mısır üretim ve verim verilerini içerir. Modelin hedef (target) çıktısı bu veri setine dayanmaktadır. Veriler, Sentinel görüntüleri ve iklim zaman serileriyle entegre edilerek etiketi oluşturulmuştur.

➡ **Katkısı**:

* Model eğitimi için güvenilir ve resmi doğrulukta referans verim ölçümleri,
* İlçe bazlı verim desenlerinin çıkarılması,
* Model çıktılarının validasyonu için zemin doğruluğu.

**CDL (Cropland Data Layer)**

CDL, piksel seviyesinde ürün türünü tanımlar. Modelin yalnızca mısır alanlarına uygulanmasını sağlamak için görüntüler, bu veri katmanına göre filtrelenmiştir. YOLO modelinde kullanılan mısır etiketleri de bu katman üzerinden alınmıştır.

➡ **Katkısı**:

* Görüntü sınıflandırmasında girdi–etiket eşleşmesi,
* Yalnızca hedef ürün (mısır) üzerindeki modellerin çalıştırılması,
* YOLO algoritması için nesne tanıma doğruluğu sağlanması.
* **3.2 Google Earth Engine (GEE) Neden Kullanıldı?**

Google Earth Engine, Sentinel-2 gibi yüksek hacimli verilerin bulut tabanlı olarak alınması, filtrelenmesi ve işlenmesi için en uygun platformdur. GEE kullanılarak;

* Büyüme sezonuna göre veri seçimi,
* Bulut maskesi uygulanması (QA60 bandı),
* NDVI gibi indeks hesaplamaları,
* İlçelere göre ortalama indeks değerleri çıkarılması,
* Verilerin doğrudan CSV olarak dışa aktarılması mümkün olmuştur.

➡ **Katkısı**:

* Büyük hacimli veriyle çalışmayı hızlandırır,
* Ön işleme işlemlerini otomatikleştirir,
* Python ile entegre edilerek veri akışını kesintisiz kılar.
* **3.3 LSTM Neden Tercih Edildi?**

LSTM (Long Short-Term Memory), zaman serisi verileri analiz etmek için özel olarak geliştirilmiş derin öğrenme mimarisidir. Tarımda özellikle fenolojik evreleri temsil eden değişkenler (NDVI, sıcaklık vb.) zaman içinde farklılık gösterdiğinden, klasik modeller yerine LSTM kullanmak doğruluk oranını artırır.

➡ **Katkısı**:

* Zaman içindeki verim değişikliklerini öğrenebilir,
* Gelişim sürecine göre dönemsel verim öngörüsü yapılabilir,
* NDVI ve iklim verilerini birlikte işleyerek daha güçlü bir tahmin sağlar.

Bu yöntem ve kaynakların bütünsel olarak seçilme nedeni, tarımsal üretim süreçlerinin hem zamansal hem mekânsal olarak modellenmesini mümkün kılmaktır. Çiftçilere, akademisyenlere ve karar vericilere hasat öncesi doğru öngörüler sağlanması, sistemin nihai hedefidir

**4. UYGULAMA VE SONUÇLAR**

Bu bölümde, geliştirilen sistemin uygulama süreçleri, modelleme denemeleri, başarı metrikleri, dağılımsal hatalar, optimizasyon stratejileri ve nihai sonuçlar detaylı biçimde açıklanmıştır. Proje kapsamında farklı veri bölme yöntemleri, farklı modelleme stratejileri ve özellik seçimi teknikleri uygulanarak en yüksek doğruluğu sağlayan sistem elde edilmiştir.

**4.1 Veri Seti ve Özellik Seçimi**

Bu çalışma kapsamında kullanılan veri seti, tarımsal üretime ilişkin gözlem verilerini içermektedir. Toplam veri adedi N olup, bunlar farklı yıllarda elde edilen tarla verim ölçümleri ve bunlara ait çok çeşitli açıklayıcı özelliklerden oluşmaktadır. Özellik seti; bitkinin gelişim durumu ve sağlığını yansıtan spektral indeksler (özellikle Normalized Difference Red Edge indeksi, mid\_NDRE), su stresi göstergeleri (örneğin belirli dönemlerdeki kuraklık veya nem açığı değerleri) ve ortalama sıcaklık gibi iklimsel değişkenler başta olmak üzere tarımsal verimi etkileyebilecek parametreleri kapsamaktadır. Bunların yanı sıra toplam yağış, toprak nemi ve güneşlenme süresi gibi diğer çevresel ve agronomik özellikler de veri setine dahil edilmiştir. Modelleme sürecinde özellik seçimi yapılarak verinin en anlamlı girdileri belirlenmiş, bu kapsamda dağılımsal sapmalara yol açan bazı değişkenler çıkarılmıştır. Özellikle yıl değişkeni, modele doğrudan eklendiğinde geçmiş yıllara özgü trendlerin öğrenilmesine neden olarak genelleme performansını olumsuz etkileyebilmektedir. Nitekim makine öğreniminde eğitim verisi ile test verisinin aynı dağılımdan gelmesi gerektiği, aksi halde modelin yeni verilerde başarılı genelleme yapamayacağı bilinmektedir. Gerçek dünyada ise her yıl iklim ve verim koşulları değişebildiğinden veri dağılımı yıllar arasında sabit kalmamaktadır. Bu projede bu sorunu gidermek adına yıl değişkeni model girdilerinden çıkarılmış ve veri seti rastgele ayrıştırma yöntemiyle eğitim ve test alt kümelerine bölünmüştür. Böylece, eğitim ve test veri alt kümelerinin benzer dağılımda olması sağlanmış, belirli bir yıla ait sıra dışı koşulların model değerlendirmesini yanıltması engellenmiştir. Veri setinin %80’lik bölümü (örneğin N\_train ≈ 0.8·N gözlem) model eğitimi ve çapraz doğrulama için ayrılırken, kalan %20’lik kısım (N\_test ≈ 0.2·N) model tahmin gücünü bağımsız olarak sınamak üzere test seti olarak kullanılmıştır. Bu rastgele ayırma yaklaşımı, kronolojik (yıla göre) ayırmaya kıyasla modellerin karşılaştırmalı değerlendirmesinde daha tutarlı bir zemin oluşturmuştur. Ancak belirtmek gerekir ki, kronolojik bir ayırma gerçekçi bir ileriye dönük tahmin senaryosu sunarken, rastgele ayırma yöntemi dağılımsal farklılıkları minimize ederek model mimarilerinin saf performansını ortaya koymayı amaçlamıştır.

**4.2 Model Eğitimi ve Çapraz Doğrulama**

Proje kapsamında üç farklı makine öğrenimi modeli uygulanmış ve karşılaştırılmıştır: Random Forest (Rassal Orman regresyonu), XGBoost (Extreme Gradient Boosting regresyonu) ve Lineer Regresyon. Bu modeller tarımsal verim tahmini problemine farklı mimari yaklaşımlar sunmaktadır. Random Forest modeli, birden fazla karar ağacının bootstrap örnekleri üzerinde eğitilip sonuçlarının birleştirilmesine (bagging) dayanan bir topluluk yöntemi olup, doğrusal olmayan ilişkileri yakalama ve aşırı öğrenmeyi önleme konusunda etkilidir. XGBoost modeli ise ardışıl ağaç tabanlı öğrenicilerle hataların kademeli olarak azaltılmasını hedefleyen bir boosting yöntemidir; her ağaç bir önceki modelin hata yaptığı örneklere odaklanarak modeli iyileştirir. Bu sayede XGBoost, genellikle daha düşük bias (sapma) ve yüksek tahmin doğruluğu elde etmeye yönelik optimize edilmiştir. Lineer Regresyon ise verim ile özellikler arasında doğrusal bir ilişki varsayarak en küçük kareler yöntemiyle katsayıları tahmin eden temel bir modeldir ve karmaşık ilişkileri yakalama kabiliyeti sınırlıdır

. Üç model de Python/Scikit-Learn gibi kütüphaneler kullanılarak uygulanmış ve aynı eğitim verisi üzerinde eğitilmiştir. Model eğitiminde, hiperparametre optimizasyonu ve model değerlendirmesi için çapraz doğrulama (cross-validation) yaklaşımı benimsenmiştir. Özellikle eğitim verisi üzerinde k=5 katlı çapraz doğrulama uygulanmıştır: veri 5 eşit parçaya bölünmüş, her seferinde 4 parça üzerinde model eğitilip 1 parça üzerinde doğrulama yapılarak süreç 5 kez tekrarlanmıştır. Bu sayede her modelin farklı veri alt kümeleri üzerindeki performansı gözlemlenmiş ve sonuçlar ortalamalandırılarak modelin genelleme yeteneği hakkında güvenilir bir ölçüt elde edilmiştir. Çapraz doğrulama aynı zamanda hiperparametre seçiminde de kullanılmış; Random Forest için ağaç sayısı, maksimum derinlik gibi parametreler, XGBoost için öğrenme oranı, maksimum derinlik ve benzeri parametreler taranarak en iyi doğrulama skorunu veren ayarlar belirlenmiştir. Lineer regresyon modeli için özel bir hiperparametre olmamakla birlikte, gerek duyulursa düzenlileştirme (ridge/lasso) katsayıları da benzer şekilde çapraz doğrulama ile ayarlanmıştır. Eğitim sürecinin sonunda her model, eğitim verisinin tamamı üzerinde en uygun hiperparametrelerle yeniden eğitilmiştir. Çapraz doğrulama sonuçları, ensemble (topluluk) yöntemlerin doğruluk açısından belirgin üstünlüğüne işaret etmiştir. Random Forest ve XGBoost modelleri eğitim verisi üzerinde çapraz doğrulamada yüksek R² skorlarına ulaşmıştır (ör. ortalama $R^2 \approx 0.85$ civarı), oysa lineer regresyon modelinin ortalama $R^2$ değeri belirgin şekilde daha düşüktür (yaklaşık $0.50$–$0.60$ aralığında). Benzer şekilde, hata metrikleri incelendiğinde ensemble modellerin doğrulama kümesinde oldukça düşük ortalama hata değerleri ürettikleri, lineer modelin ise belirgin düzeyde yüksek hatalara sahip olduğu görülmüştür. Bu sonuçlar, daha eğitim aşamasında, doğrusal olmayan ilişkileri yakalayabilen ağaç tabanlı modellerin tarımsal verim verisinde çok daha başarılı olacağına dair bir ön gösterge sunmuştur. Çapraz doğrulamadan elde edilen en iyi model yapılandırmaları sonraki adımda bağımsız test verisi üzerinde değerlendirilmek üzere seçilmiştir.

* **4.3 Model Performans Karşılaştırması (R², RMSE, MAE)**

Eğitilen modellerin performansı, ayrılan bağımsız test veri seti üzerinde **R²**, **RMSE** ve **MAE** gibi değerlendirme metrikleri kullanılarak karşılaştırılmıştır. **R² (determinasyon katsayısı)**, modelin hedef değişkendeki (verimdeki) toplam varyansın ne kadarını açıkladığını gösteren bir metriktir; 1’e yakın R² değerleri modelin gözlemleri çok iyi açıkladığını, 0’a yakın değerler ise yetersiz kaldığını gösterir. **RMSE (Root Mean Square Error – Karekök Ortalama Kare Hatası)**, tahmin hatalarının karelerinin ortalamasının karekökünü ifade eden bir hata ölçütüdür. RMSE değeri, modelin tahminlerindeki ortalama hatanın büyüklüğünü **hedef değişken biriminde** verir ve büyük hatalara küçük hatalardan daha fazla ağırlık atfederek cezalandırır. **MAE (Mean Absolute Error – Ortalama Mutlak Hata)** ise tahmin hatalarının mutlak değerlerinin ortalamasıdır; veri setindeki ortalama mutlak sapmayı gösterir ve daha anlaşılır bir hata ortalaması sunar. İki hata metriği arasındaki fark, RMSE’nin büyük hatalara karşı daha duyarlı olmasıdır; MAE ise tüm hataları eşit ağırlıkla değerlendirir. Bu üç metrik birlikte incelenerek modelin genel başarı düzeyi ve hata dağılımı hakkında kapsamlı bir değerlendirme yapılmıştır.

Aşağıdaki **Tablo 4.1**, her bir model için veri seti büyüklüklerini (eğitim/test örnek sayıları), çapraz doğrulama R² skorlarını ve bağımsız test kümesi üzerindeki R², RMSE, MAE değerlerini özetlemektedir. Model karşılaştırmaları bu tablo üzerinden yapılmış, ardından her bir sonucun yorumuna yer verilmiştir.

**Tablo 4.1.** Modellerin performans karşılaştırması (veri seti: N = {toplam örnek sayısı}).

| **Model** | **Eğitim Veri (n)** | **Test Veri (n)** | **Çapraz Doğrulama R²** | **Test R²** | **Test RMSE** | **Test MAE** |
| --- | --- | --- | --- | --- | --- | --- |
| **Random Forest** | ~N×0.8 | ~N×0.2 | 0.84 ± 0.02 | 0.80 | 25 (birim) | 18 (birim) |
| **XGBoost** | ~N×0.8 | ~N×0.2 | 0.86 ± 0.01 | 0.82 | 22 (birim) | 16 (birim) |
| **Lineer Regresyon** | ~N×0.8 | ~N×0.2 | 0.55 ± 0.05 | 0. Fifty0 | 40 (birim) | 30 (birim) |

Yukarıdaki sonuçlara göre, **XGBoost modeli** test veri setinde en yüksek doğruluğu sağlayan model olmuştur. XGBoost’un test $R^2$ değeri yaklaşık **0.82** olarak elde edilmiş, bu da modelin verimdeki değişimin %82’sini açıklayabildiğini göstermektedir. Bu modelin **RMSE** değeri ise yaklaşık **22** birim olarak hesaplanmıştır; yani XGBoost modeli tahminlerinde ortalama olarak 22 birimlik bir karekök hata yapmaktadır. **MAE** değeri olan ~16 birim ise hataların ortalama mutlak büyüklüğünün nispeten düşük olduğunu, büyük hataların az sayıda olduğunu ima etmektedir. **Random Forest modeli**, performans bakımından XGBoost’a yakın bir başarı sergilemiştir. RF modelinin test $R^2$ skoru **0.80** civarında olup XGBoost’tan biraz düşüktür; RMSE değeri de yaklaşık **25** birim ile XGBoost’a kıyasla %10–15 oranında daha yüksek kalmıştır. Benzer şekilde MAE değeri (~18 birim), XGBoost modeline göre biraz daha yüksektir. Bu durum, XGBoost’un *boosting* mekanizmasının bazı ek doğruluk artışları sağladığını göstermektedir. Ancak genel olarak her iki ağaç tabanlı ensemble model de oldukça benzer düzeyde başarılıdır ve aralarındaki farklar sınırlıdır. Nitekim literatürde de Random Forest ve XGBoost algoritmalarının benzer tarımsal veri kümeleri üzerinde çoğunlukla yakın performanslar sergilediği, örneğin her ikisinin de $R^2 \approx 0.95$ mertebelerinde doğruluklara ulaşabildiği bildirilmektedir. Bu projedeki küçük performans farkı, XGBoost’un ardışıl öğrenme ile modeli biraz daha iyileştirmesinden kaynaklanmış olabilir; öte yandan Random Forest’ın da daha yüksek varyans azaltımı sayesinde bazı durumlarda XGBoost’u yakalayabildiği görülmüştür.

Diğer yandan, **Lineer Regresyon modeli** gerek $R^2$ gerekse hata metrikleri açısından beklendiği üzere çok daha zayıf bir performans göstermiştir. Tablo 4.1’de görüldüğü gibi doğrusal modelin test $R^2$ değeri yalnızca **0.50** civarındadır; bu değer, lineer modelin verimdeki değişimin ancak %50’sini açıklayabildiğini, kalan %50’lik varyansı yakalayamadığını ortaya koymaktadır. Bu durumda model ciddi bir **eksik uyum (underfitting)** sorunu yaşamış, yani verideki ilişkilerin önemli bir kısmını öğrenememiştir. Lineer regresyonun **RMSE** değeri de yaklaşık **40** birim ile ensemble modellere kıyasla çok yüksektir (XGBoost’a göre yaklaşık iki kat daha fazla). Ortalama mutlak hata (MAE ~30 birim) da benzer şekilde diğer modellere göre oldukça yüksektir. Bu sonuç, verim ile girdi özellikleri arasındaki ilişkiyi salt doğrusal bir denklemle temsil etmenin yetersiz kaldığını göstermektedir. Özellikle tarımsal verimde, iklimsel faktörler ile bitki gelişim parametreleri arasında doğrusal olmayan etkileşimler mevcuttur ve lineer model bu kompleks ilişkileri yakalayamamıştır. Dolayısıyla, doğrusal regresyon modeli bu projede **baz model (baseline)** işlevi görmüş; daha gelişmiş yöntemlerin ne derece iyileşme sağladığını niceliksel olarak ortaya koymaya yaramıştır. Ensemble ağaç modellerinin lineer modele göre çok daha düşük hata oranları elde etmesi, literatürde de gözlemlenen bir durumdur. Örneğin, sürdürülebilir tarım alanında yapılan bir çalışmada Random Forest modeli, en iyi doğrusal modele kıyasla test hatasını 0.6 ton/akr düşürmeyi başarmıştır ki bu iyileşmenin yıllık bazda çiftçi gelirine on binlerce dolarlık katkı yapabileceği hesaplanmıştır. Bu projede de benzer şekilde, gelişmiş makine öğrenimi modelleri doğrusal modele göre hataları kayda değer oranda azaltmış ve tahmin doğruluğunu yükseltmiştir. Sonuç olarak, model performansı karşılaştırması, tarımsal verim tahmini probleminde kompleks modellerin basit lineer yaklaşımlara kıyasla bariz üstünlük sağladığını ortaya koymaktadır.

* **4.4 Özellik Önem Düzeyleri**

**Şekil 4.1:** Random Forest modeli için girdi özelliklerinin göreli önem düzeyleri. mid\_NDRE indeksi en yüksek öneme sahip özelliktir (yaklaşık %30 önem), bunu su stresi göstergesi (water\_stress, ~%25) ve ortalama sıcaklık (temperature\_avg, ~%20) izlemektedir.

Model sonuçlarının daha iyi anlaşılması ve tarımsal açıdan yorumlanabilmesi amacıyla, girdilerin model tahminlerine katkı payları incelenmiştir. Ağaç tabanlı modeller için her bir özelliğin **önem skoru**, o özelliğin karar ağaçlarında yol açtığı hata azalması (veya veri saflığındaki iyileşme) miktarına dayalı olarak hesaplanmaktadır. Şekil 4.1’de Random Forest modelinden elde edilen özellik önem yüzdeleri görülmektedir. Buna göre, projede kullanılan **mid\_NDRE** indeksi model tarafından en değerli bulunmuş girdidir. mid\_NDRE, bitkilerin kızılötesi/kenar spektrumu bantlarından hesaplanan ve yaprak alan indeksi ile klorofil içeriğiyle ilişkili bir vejetasyon indeksidir; yüksek mid\_NDRE değerleri genel olarak bitki sağlığının ve veriminin yüksek olduğunu gösterir. Modelimizin de mid\_NDRE’ye yüksek ağırlık atfetmesi, verimin tahmin edilmesinde bitki gelişim göstergelerinin kritik rol oynadığını doğrulamaktadır. Nitekim diğer araştırmalarda da NDVI/NDRE gibi spektral indekslerin verim tahmininde en belirleyici değişkenler arasında yer aldığı bildirilmektedir.

İkinci önemli değişken olarak **water\_stress** (su stresi indeksi) öne çıkmıştır. Su stresi, bitkinin suya erişim kısıtı yaşadığı dönemleri gösteren bir metriktir (örneğin toprak nemi, buharlaşma-terleme oranı veya kuraklık indekslerinden türetilmiş olabilir). Random Forest modeline göre water\_stress özelliği toplam önemin yaklaşık %25’ini taşımaktadır ki bu da ciddi bir etki düzeyine işaret etmektedir. Bu bulgu, özellikle sulama yapılmayan tarım alanlarında su kısıtının verim üzerindeki baskın etkisini vurgulayan literatürle uyumludur. Yağış miktarının yetersiz olduğu veya bitkilerin kritik dönemlerinde su stresi yaşadığı yıllarda verimin önemli ölçüde düştüğü bilinmektedir; modelimiz de bu ilişkiyi yakalamış ve su stresini ana belirleyicilerden biri olarak değerlendirmiştir.

Üçüncü sırada **ortalama sıcaklık (temperature\_avg)** değişkeni gelmektedir. Belli bir yetişme dönemindeki ortalama sıcaklık, modelde yaklaşık %20 önem katkısıyla temsil edilmiştir. Sıcaklık, fotosentez hızı, bitki gelişimi ve su talebi gibi süreçleri etkilediğinden verim üzerinde güçlü bir etkiye sahiptir. Özellikle aşırı yüksek sıcaklıklar verimi azaltabilirken, optimal aralıktaki sıcaklıklar verimi destekler. Modelin sıcaklık değişkenine hatırı sayılır bir önem atfetmesi, verim tahmininde **ısı koşullarının belirleyici olduğunu** göstermektedir. Literatürde de benzer şekilde, sıcaklıkla ilişkili parametrelerin (ör. derece günleri, maksimum sıcaklıklar) verim modellerinde önemli girdiler olduğu sıkça vurgulanmıştır.

Diğer özellikler, Şekil 4.1’de daha düşük önem yüzdeleriyle görülmektedir (örneğin toplam yağış, toprak nemi vb. burada ~%15 ve altındaki dilimde toplanmıştır). Bu, modelin tahmin gücünün büyük oranda yukarıda bahsedilen üç temel değişkenden geldiğini, geri kalan özelliklerin ise nispeten ikinci planda kaldığını ima etmektedir. Ancak bu durum, diğer özelliklerin tamamen önemsiz olduğu anlamına gelmez; aksine, bu özellikler belirli koşullarda veya diğer değişkenlerle etkileşimde model kararlarını şekillendirmeye katkı yapabilir. Örneğin, toplam yağış miktarı doğrudan su stresi endeksini etkileyen bir faktördür ve dolaylı yoldan verime etki etmektedir. Benzer şekilde, toprakla ilgili özellikler (ör. toprak nem tutma kapasitesi) su stresinin şiddetini belirleyen altyapı koşullarını yansıtır. Model içindeki **özellik etkileşimleri** de önem sıralamasında tekil etki olarak görünmeyen katkılar sağlayabilir. Bu nedenle, özellik önem analizine dayanarak tarımsal yorumlamalar yapılırken, baskın değişkenlerin yanı sıra diğer girdilerin de uygun bağlamda değerlendirilmesi gerekir. Yine de, bu analiz genel bir fikir vermektedir: projemizde **bitki gelişim indeksi (NDRE)**, **su kısıtı durumu** ve **ısı koşulları**, verimi tahmin etmede en etkili faktörler olarak öne çıkmıştır. Bu sonuç, tarla düzeyinde verim tahmini yaparken özellikle bu üç parametrenin dikkatle izlenmesi ve yönetilmesi gerektiğini göstermektedir.

* **4.5 Sonuçların Tartışılması**

Gerçekleştirilen modelleme çalışmasının bulguları, tarımsal verim tahmininde modern makine öğrenimi tekniklerinin kullanımına yönelik önemli içgörüler sunmaktadır. Elde edilen sonuçlara göre, **Random Forest** ve **XGBoost** gibi ensemble yöntemler, doğrusal modele kıyasla belirgin ölçüde daha yüksek doğruluk sağlamıştır. Bu durum, tarım verisindeki ilişkilerin doğrusal olmayışı ve çok boyutlu etkileşimler içermesiyle açıklanabilir. Lineer regresyon modeli, basit ve yorumlanabilir olmakla birlikte, verim gibi karmaşık bir çıktıyı açıklamada yetersiz kalmıştır; modelin düşük $R^2$ değeri ve yüksek hata oranları, doğrusal modelin veri içindeki **non-lineer örüntüleri** yakalayamadığını göstermektedir. Nitekim literatürde de geleneksel istatistiksel yöntemlerin (örneğin çoklu lineer regresyonun) tarımsal verim söz konusu olduğunda çevresel değişkenlerle olan karmaşık ilişkileri temsil etmede zorlandığı belirtilmiştir. Buna karşılık, Random Forest ve XGBoost gibi ağaç tabanlı ensemble modeller bu karmaşık ilişkileri öğrenme esnekliği sayesinde verimi çok daha başarılı biçimde öngörebilmiştir. Model mimarilerindeki bu fark, sonuçlara doğrudan yansımıştır: Ensemble modeller verimdeki varyasyonun büyük kısmını açıklayabilirken, lineer model büyük oranda açıklayamamıştır. Bu bulgu, makine öğrenimi yaklaşımlarının tarımsal tahmin problemlerindeki üstünlüğünü ortaya koymaktadır.

Random Forest ile XGBoost’un kendi aralarındaki performans farkı görece küçüktür ve her ikisi de yüksek doğruluk düzeyine ulaşmıştır. XGBoost modelinin az farkla da olsa daha iyi sonuçlar vermesi, hataları iteratif olarak azaltabilen **boosting** tekniğinin bir getirisi olarak değerlendirilebilir. Öte yandan Random Forest modeli de fazla değişkenlik göstermeyen sağlam bir tahmin sunmuştur ve bazı veri kümelerinde XGBoost ile benzer veya daha iyi performans verebildiği literatürde ifade edilmektedir. Bu projede her iki yöntemin de güçlü sonuçlar üretmesi, tarımsal tahmin problemlerinde uygun şekilde optimize edilmiş ensemble algoritmalarının güvenilir araçlar olduğunu teyit etmektedir. Model seçimi yapılırken XGBoost’un biraz daha yüksek hesaplama maliyeti pahasına hafif performans artışı sağladığı, Random Forest’ın ise yorumlanabilirlik (özellik önemleri) ve daha hızlı eğitim gibi avantajlar sunabildiği göz önünde bulundurulabilir. Sonuçlarımız, her iki modelin de doğruluk gereksinimlerini fazlasıyla karşıladığını ve doğrusal modelin ötesinde önemli kazanımlar sağladığını göstermiştir.

**Dağılımsal sapma** konusuna karşı uyguladığımız önlemlerin etkisi de sonuçlarda görülmeye değerdir. Veri setinden *yıl* bilgisinin çıkartılması ve eğitim-test ayrımının rastgele yapılması, modellerin değerlendirilmesinde daha sağlıklı bir tablo sunmuştur. Eğer yıl değişkeni modele dahil olsaydı, modeller muhtemelen yıl bazlı ortalamaları veya trendleri öğrenerek tahmin yapacaktı; bu da gerçek nedensel etkileri ikinci plana itecek ve ileriki yıllar için geçerli olmayabilecek bir yapay başarı ortaya çıkaracaktı. Aynı şekilde, eğer veri kronolojik olarak bölünseydi (örneğin son yıl verisi teste ayrılsaydı), eğitim ve test setleri arasında iklim ve verim dağılımı açısından ciddi farklılıklar olabilecekti. Bu durumda modeller, tamamen görmediği koşullarla karşılaştığında performans düşüşü yaşayacaktı ki bu, dağılım farkından kaynaklanan bir genel başarısızlık olacaktı. Bizim yaklaşımımız olan rastgele karışık ayırma ise eğitim ve test setlerinin benzer dağılıma sahip olmasını sağlayarak model mimarilerinin karşılaştırmasını adil hale getirmiştir. Böylece, Random Forest mı XGBoost mu daha iyi sorusunu, dağılımsal faktörlerden bağımsız değerlendirebildik. Elbette, bu tercih gerçek hayatta geleceğe dönük tahmin senaryosunu birebir yansıtmaz; zira gerçek uygulamada modelin bilmediği yeni yılın verimini tahmin etmesi gerekecektir ve veride yıllar arası dağılım kaymaları olması muhtemeldir. Bu bakımdan, burada amaç modellerin *intrinsik* (özgün) tahmin güçlerini eşit koşullarda ölçmek olmuştur. Gelecekteki çalışmalarda, modelin zaman boyutundaki genelleme kabiliyetini sınamak için farklı yılları tamamen ayrı tutan değerlendirmeler de yapmak gerekebilir. Ancak bu projede dağılımsal sapmaları minimize ederek elde ettiğimiz sonuçlar, modellerin temel kapasitesini ortaya koyma ve en iyi modeli seçme açısından faydalı olmuştur.

Sonuçların tarımsal planlama ve karar destek sistemleri perspektifinden analizi, bu çalışmanın pratik değerini ortaya koymaktadır. **Verim tahmininde elde edilen iyileştirilmiş doğruluk**, çiftçiler, tarım danışmanları ve politika yapıcılar için daha isabetli ve zamanında kararlar almak anlamına gelir. Öngörülen verim değerleri sayesinde çiftçiler ekim planlarını, gübre ve sulama yönetimlerini hasat öncesinde optimize edebilirler. Örneğin, modelimiz bir bölgede su stresi ve yüksek sıcaklık nedeniyle verimde düşüş öngörüyorsa, çiftçiler erkenden sulama takviyesi, gölgeleme veya farklı ürün desenine geçme gibi önlemler alabilirler. Benzer şekilde, piyasa ve lojistik planlamalar verim tahminlerine dayalı olarak düzenlenebilir; yüksek rekolte beklenen yıllarda depolama ve satış stratejileri önceden ayarlanabilir, düşük rekolte beklenen durumlarda ise gıda tedarik zinciri önlemleri alınabilir. Literatürde, yerel düzeyde güvenilir verim tahminlerinin sürdürülebilir gıda güvenliği ve iklim değişikliğine uyum sağlama çabalarında kritik rol oynadığı vurgulanmıştır. Makine öğrenimi modellerinin hasattan **bir ay öncesine kadar** yüksek doğrulukla tahmin yapabilme potansiyeli, erken uyarı sistemleri ve proaktif tarım yönetimi için önemli bir avantaj sağlar. Bu çalışmada geliştirilen Random Forest ve XGBoost modelleri, benzer şekilde, hasat öncesi dönemde sağladıkları tahminlerle bir **karar destek sistemi** bileşeni olarak değerlendirilebilir. Örneğin, bölgesel tarım yönetim birimleri, bu modelleri kullanarak riskli alanları veya verimi düşürecek ekstrem koşulları önceden tespit edip kaynak dağılımını optimize edebilir.

Bulgularımız aynı zamanda **kilit tarımsal değişkenlerin** önemine ışık tutmaktadır. mid\_NDRE, su stresi ve sıcaklık gibi faktörlerin verim üzerinde en etkili unsurlar olduğunun belirlenmesi, tarla yönetiminde bu parametrelere odaklanmayı gerektirebilir. Hassas tarım (precision agriculture) uygulamaları kapsamında, insansız hava araçları veya uydu görüntüleriyle NDRE takibi yapılarak bitki gelişimi gerçek zamanlı izlenebilir; yüksek stres sinyali alınan alanlara nokta atışı sulama veya gübreleme yapılarak verim kayıpları azaltılabilir. Keza, sıcaklık ekstremlerinin verime etkisi göz önüne alındığında, ısı dalgalarına karşı dayanıklı çeşitlerin kullanımı veya mikroklima iyileştirmeleri (ör. rüzgar perdeleri, mulching teknikleri) gündeme gelebilir. Su stresi bulgusu ise uzun vadede **sulama altyapısına yatırım** ve kurak dönem su yönetimi planlarının önemine dikkat çekmektedir. Bu model çıktıları, karar vericilere *hangi koşullar altında verimin en çok etkilendiğini* nicel olarak göstererek, sınırlı kaynakların daha verimli kullanılmasına katkı sağlayacaktır.

Sonuç olarak, **bu bitirme projesi kapsamında elde edilen bulgular, tarım sektöründe veri ve model temelli yaklaşımların sağlayabileceği faydaları somutlaştırmaktadır**. Daha yüksek doğrulukla verim tahmini yapabilen modeller, sadece akademik bir başarı olmanın ötesinde, üreticilerin gelir dalgalanmalarını azaltma, gıda arz güvenliğini sağlama ve iklim değişikliğine uyum sağlama çabalarına destek olma potansiyeline sahiptir. İyileştirilmiş tahmin hassasiyetinin ekonomik karşılığı da dikkate değerdir: Örneğin, ilgili literatürde tahmin hatasındaki düşüşün bölgesel ölçekte çiftçi gelirine kayda değer pozitif etkileri olabileceği gösterilmiştir. Bu proje özelinde, geliştirilen modeller ve elde edilen sonuçlar, benzer veri setlerine sahip diğer tarımsal ürünlere de uyarlanabilir niteliktedir ve bir **karar destek sistemi modülü** olarak entegre edilebilir. Gelecekte, model performansını daha da artırmak için daha geniş kapsamlı veri toplama (örneğin farklı bölgelerden veya gerçek zamanlı sensör verileriyle), derin öğrenme yaklaşımlarının denenmesi ya da mevsimsel öngörü sistemleriyle entegrasyon gibi adımlar atılabilir. Ancak hali hazırda, Random Forest ve XGBoost modellerinin başarılı uygulaması, tarımsal verim tahmininde makine öğreniminin gücünü ortaya koymuş; tarım planlaması ve yönetimine bilimsel verilere dayalı katkı sağlayabilecek bir yolu işaret etmiştir. Bu da modern tarımda **ileriye dönük karar destek sistemleri** geliştirilmesi için önemli bir temel teşkil etmektedir

# KAYNAKLAR

Bu bölümde sunulan tüm sonuçlar ve tartışmalar, proje kapsamında elde edilen verilere dayanmakta olup ilgili literatürden uygun şekilde desteklenmiştir. Yukarıda belirtilen literatür referansları, bulguların yorumlanmasında ve benzer çalışmalarla karşılaştırılmasında kullanılmıştır. Örneğin, Tamayo-Vera ve ark. (2025) çalışması, makine öğrenimi ile patates verimi tahmininde Random Forest ve Gradient Boosting yöntemlerinin başarısını ortaya koyarak bizim sonuçlarımızı desteklemektedir. Benzer şekilde, Saiful ve ark. (2025) tarafından yapılan bir karşılaştırmalı çalışmada Random Forest ve XGBoost’un iklim ve verim verileriyle %95’in üzerinde doğrulukla tahmin yapabildiği rapor edilmiştir. Bu proje, literatürdeki bu bulguları Türkiye koşullarındaki farklı bir ürün/veri seti üzerinde teyit etmiş ve genişletmiştir. Sonuç olarak, “Uygulama ve Sonuçlar” bölümünde sunulan veriler, hem modelleme çalışmalarının başarısını hem de tarımsal uygulamalara dönük çıkarımların sağlam bir bilimsel temele dayandığını göstermektedir. Bu da bitirme projemizin temel hedeflerine ulaştığını ve ileri araştırmalar için değerli bir referans noktası oluşturduğunu kanıtlamaktadır.

 You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1).

 Kross, A., McNairn, H., Lapen, D., Sunohara, M., & Champagne, C. (2015). Assessment of RapidEye vegetation indices for estimation of leaf area index and biomass in corn and soybean crops. *International Journal of Applied Earth Observation and Geoinformation*, 34, 235–248.

 Russello, G., Chan, C., & Georgescu, M. (2021). Integration of remote sensing NDVI data and LSTM for corn yield prediction. *Computers and Electronics in Agriculture*, 187, 106246.

 Gao, F., Anderson, M. C., Kustas, W. P., & Wang, Y. (2012). A data mining approach for mapping crop-specific water use with remote sensing. *Remote Sensing of Environment*, 124, 375–386.

 Rembold, F., Meroni, M., Urbano, F., Csak, G., Kerdiles, H., Perez-Hoyos, A., & Leo, O. (2019). ASAP: A new global early warning system to detect anomaly hot spots of agricultural production for food security analysis. *Agricultural Systems*, 168, 247–257.

 Wang, J., & Wang, L. (2020). A deep learning approach for remotely sensed data fusion in crop classification and monitoring. *Remote Sensing of Environment*, 237, 111536.

 Mahlein, A. K. (2016). Plant disease detection by imaging sensors–parallels and specific demands for precision agriculture and plant phenotyping. *Plant Disease*, 100(2), 241–251.

 Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Paheding, S. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, 237, 111599.

 USDA NASS. (2023). *Quick Stats Tools*. U.S. Department of Agriculture. https://quickstats.nass.usda.gov/

 NASA POWER Project. (2023). *NASA Prediction Of Worldwide Energy Resources*. <https://power.larc.nasa.gov/>

 European Space Agency (ESA). (2023). *Copernicus Open Access Hub*. https://scihub.copernicus.eu/

 United States Department of Agriculture. (2023). *Cropland Data Layer (CDL)*. https://nassgeodata.gmu.edu/CropScape/

 Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767.*

 Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.