



Teknoloji Fakültesi

BİLGİSAYAR MÜHENDİSLİĞİ BÖLÜMÜ

Mahsul Verimliliği Tahmini ve Planlama

VAHİT SEYİT BAŞA

170420024

Bitirme Projesi 2. Ara Raporu
Bilgisayar Mühendisliği Bölümü

DANIŞMAN
Dr. Öğr. Üyesi Ali Sarıkaş

İSTANBUL, 2025

ÖNSÖZ

Bu rapor, bitirme projesinin ikinci deęerlendirme dnemindeki alıřmaların sonularını kapsamaktadır. Projenin veri toplama ve n iřleme ařamalarında bana rehberlik eden danıřmanım Dr. Ali Sarıkaf'a teřekkrlerimi sunarım. Ayrıca, alıřmalarım boyunca destek olan laboratuvar arkadařlarıma ve aileme řkranlarımı ifade ederim.

İİNDEKİLER

1.GİRİř	1	1.1.Proje alıřmasının Amacı ve nemi	2	2.	
Veri Seti	2	2.1.İndeksler ve n iřleme	3	3.Modelleme	
Planı	4	4. SONULAR	5		

ZET

Tarımda verimlilik tahmini, srdrlebilir tarım uygulamaları ve planlaması iin kritik neme sahiptir. Uzaktan algılama (rneęin Sentinel-2 uydu verileri) ve makine ęrenimi yntemlerinin birleřimi, farklı coęrafi leklerde ve periyotlarda

mahsul verimliliği tahmininde yaygın olarak kullanılmaktadır

. Sentinel-2 uydularının sunduğu 10–20 m uzaysal çözünürlük ve sık frekanslı gözlem imkanı, tarımsal alanların ince ölçekli takibi için idealdir

. Aynı şekilde Google Earth Engine gibi platformlar, çok büyük uydu görüntüsü arşivlerini analiz etmeyi mümkün kılar

. Bu çalışma kapsamında, ABD'nin önemli mısır üretim merkezlerinden Iowa eyaleti seçilmiş; 2018–2023 dönemine ait Sentinel-2 görüntüleri ve USDA NASS verileri kullanılarak mısır verim tahmini amaçlanmıştır. Rapor, kullanılan veri seti kaynaklarını, bitki indekslerini ve ön işleme adımlarını detaylandıracak; henüz tamamlanmamış olan modelleme aşaması için önerilen yöntemleri açıklayacak ve elde edilen ara sonuçları ile gelecek çalışmaların yönünü sunacaktır.

Mayıs, 2025

Öğrenciler : VAHİT SEYİT BAŞA

ABSTRACT

Crop yield estimation is of critical importance for sustainable agricultural practices and planning.

The combination of remote sensing (e.g., Sentinel-2 satellite data) and machine learning methods is widely used for crop yield prediction across various

geographic scales and time periods.

The 10–20 meter spatial resolution and high-frequency observation capability offered by Sentinel-2 satellites make them ideal for fine-scale monitoring of agricultural lands.

Similarly, platforms like Google Earth Engine enable the analysis of vast satellite imagery archives.

In this study, the state of Iowa—one of the major corn-producing regions of the United States—was selected. Using Sentinel-2 imagery and USDA NASS data from the period of 2018 to 2023, corn yield estimation was targeted.

This report will detail the data sources, vegetation indices, and preprocessing steps used. It will also present the proposed methods for the yet-to-be-completed modeling phase, outline interim findings, and provide direction for future work.

May 2025

Students: VAHİT SEYİT BAŞA

SEMBOLLER

λ : aydınlatma ışığı dalga boyu

λ_s : aydınlatma deseni uzamsal frekansı

$\diamond \diamond \square$: optik transfer fonksiyonu kesim
 frekansı $\diamond \diamond$: aydınlatma deseni açısal
 yönelimi $\diamond \diamond$: aydınlatma deseni faz değeri
 $\diamond \diamond \square$: aydınlatma deseni ortalama ışık
 yoğunluğu $\diamond \diamond$: aydınlatma deseni
 modülasyon derinliği \otimes : konvolüsyon çarpımı
 operatörü $\diamond \diamond$: Fourier dönüşümü
 $\diamond \diamond^{\circ} \square$: ters Fourier dönüşümü
 $\diamond \diamond \square$: fiber kılıfı kırınım indisi
 $\diamond \diamond \square$: fiber çekirdeği kırınım indisi
 $\diamond \diamond$: Euler sabiti
 $\diamond \diamond \diamond \diamond$: milimetre
 $\mu \diamond \diamond$: mikrometre
 $\diamond \diamond \diamond \diamond$: saniye
 $\diamond \diamond$: aydınlatma deseni periyodu
 $(*)$: kompleks bir sayının kompleks eşleniği
 Σ : toplam sembolü
 $\diamond \diamond$: Dirac delta fonksiyonu
 \square : bir fonksiyonun frekans uzayındaki karşılığı

1. GİRİŞ

Dünya genelinde artan nüfus, iklim değişikliği ve doğal kaynakların sınırlı hale gelmesi, tarım sektöründe daha verimli, sürdürülebilir ve teknoloji destekli üretim

yöntemlerine olan ihtiyacı her geçen gün artırmaktadır. Özellikle temel gıda maddelerinden biri olan mısırın üretim miktarının doğru tahmini, tarımsal planlama, gıda güvenliği ve ekonomik kararlar açısından kritik bir öneme sahiptir. Bu bağlamda, uzaktan algılama teknolojileri ile yapay zekâ yöntemlerinin entegrasyonu, tarım alanlarında verim tahmini gibi karmaşık problemlerin çözümünde yeni olanaklar sunmaktadır.

Bu proje, Amerika Birleşik Devletleri'nin tarımsal üretimde önde gelen eyaletlerinden biri olan Iowa'da, 2018–2023 yılları arasında Nisan–Ekim aylarını kapsayan mısır üretim sezonuna ait uydu görüntüleri ile tarım istatistiklerini bir araya getirerek, ilçeler düzeyinde mahsul verimliliğini tahmin etmeye yönelik bir sistemin temelini atmayı amaçlamaktadır. Google Earth Engine (GEE) platformu üzerinden elde edilen Sentinel-2 uydu görüntüleri ile çeşitli bitki indeksleri (NDVI, SAVI, NDRE vb.) hesaplanmış; bu görüntüler, USDA (United States Department of Agriculture) tarafından sağlanan ilçe ve ay bazlı üretim verileriyle eşleştirilerek etiketli bir veri seti oluşturulmuştur.

Projenin bu ikinci aşamasında, veri seti oluşturma ve ön işleme adımları başarıyla tamamlanmış, test veri kümesi hazırlanmış ve kullanılabilecek potansiyel makine öğrenimi ve derin öğrenme algoritmalarına dair detaylı analizler yapılmıştır. Henüz model eğitimi gerçekleştirilmemekle birlikte, sonraki aşamalarda uygulanacak yöntemler ile yüksek doğrulukta verim tahmini yapılması hedeflenmektedir. Bu doğrultuda yapılan çalışmalar, hem tarımsal karar destek sistemlerinin geliştirilmesine katkı sağlamakta hem de uzaktan algılama ve yapay zekâ entegrasyonunun tarım sektöründeki uygulanabilirliğini ortaya koymaktadır.

1

1.1. Proje Çalışmasının Amacı ve Önemi

Bu karar destek sistemi, tarımda verimlilik ve sürdürülebilirliği artırmak amacıyla çok boyutlu verilerden faydalanarak ilerlemektedir. Görüntü işleme teknikleri, tarım arazilerindeki bitkilerin sağlık durumlarını izlemek ve erken aşamalarda

hastalıkları veya zararlıları tespit etmek için kullanılacaktır. Derin öğrenme algoritmalarının, bitki hastalıklarını tanımlamak ve büyüme süreçlerini modellemek için eğitilmesiyle, çiftçilere anlık ve doğru bilgi sağlanacaktır. Bu sayede, hastalıklar veya stres faktörleri erken tespit edilip, etkin bir şekilde müdahale edilerek, ürün kayıplarının önüne geçilebilecektir.

2. Veri Seti

Veri seti, iki ana kaynaktan oluşmaktadır: Sentinel-2 uydu görüntüleri ve USDA NASS istatistikleri. Özellikle şu bileşenler yer almaktadır:

- Sentinel-2 Uydu Verileri: Iowa eyaletinin tüm ilçelerini kapsayan Sentinel-2 L2A görüntüleri 2018–2023 yılları arasında Nisan–Ekim (büyüme dönemi) ayları için Google Earth Engine (GEE) üzerinden temin edildi. GEE, çok petabaytlık uydu görüntülerini küresel ölçekte işleyebilme olanağı sunar. Elde edilen görüntülerdeki ilgili bantlardan NDVI, SAVI, NDRE gibi bitki indeksleri hesaplanarak her ay için özet değerler çıkarıldı.
- USDA NASS Verileri: ABD Tarım Bakanlığı Ulusal Tarım İstatistik Servisi (NASS) tarafından yayımlanan aylık ve ilçe bazlı mısır üretim verileri kullanıldı. Bu veriler Iowa eyaletindeki her ilçe ve yıl için mısır üretim miktarını (örneğin bushel cinsinden) içermektedir. Uydudan hesaplanan indeks değerleri ile bu üretim verileri birleştirilerek etiketli bir veri seti oluşturuldu. Böylece her bir yıl-ay-ilçe örneği, ilgili aylarda hesaplanmış bitki indeksleri ve bu dönem sonundaki mısır verimi değeri ile temsil edildi.
- Veri Birleştirme: Uydudan elde edilen indeks zaman serileri ile USDA verileri senkronize edilip birleştirildi. Her örnek giriş vektörü olarak belirli bir yıl ve ay

(örneğin bulutlu alanlardan kaynaklanan gözlem boşlukları) interpolasyonla giderilmiş, ay bazında kesintisiz zaman serileri elde edilmiştir.

2.1. İndeksler ve Ön İşleme

Veri setindeki temel girdiler, Sentinel-2 görüntülerinden hesaplanan bitki indeksleri ve toprak düzeltme indeksi ile ilişkilidir. Kullanılan başlıca indeksler şunlardır:

- NDVI (Normalized Difference Vegetation Index): Bitki sağlığını ve yeşil örtü yoğunluğunu gösteren en yaygın indekstir. NDVI, kızılötesi (NIR) ve kırmızı (Red) bantlarının normalize farkı olarak hesaplanır:

$$NDVI = (NIR - Red) / (NIR + Red).$$

Sentinel-2 için bu formül, B8 (NIR) ve B4 (kırmızı) bantları kullanılarak $NDVI = (B8 - B4) / (B8 + B4)$ şeklinde uygulanır. NDVI değeri [-1, 1] aralığında olup, yüksek değerler sağlıklı ve sık bitki örtüsünü, düşük (veya negatif) değerler seyrek örtülü ya da diğer yüzey tiplerini işaret eder.

- SAVI (Soil Adjusted Vegetation Index): NDVI'nin toprak parıltısı etkisine duyarlılığını azaltmak amacıyla geliştirilmiş bir düzeltilmiş indekstir. USGS'e göre SAVI, "bitki örtüsü azaldığında NDVI'nin toprak parlaklığından etkilenmesini düzeltmek" için kullanılır. Tipik olarak $L=0.5$ toprak parlaklık faktörü ile $SAVI = ((NIR - Red) / (NIR + Red + L)) * (1 + L)$ formülüyle hesaplanır. SAVI, bitki örtüsü oranının düşük olduğu erken büyüme aşamalarında veya kuru toprak yüzeylerinde daha güvenilir geri dönüşler sağlar.

- NDRE (Normalized Difference Red Edge Index): NDRE, NIR ve kırmızı kenar (Red Edge) bantlarını kullanan bir indekstir. Bitkideki klorofil içeriğini ve olgunluk dönemindeki sağlıklı büyümeyi izlemek için özellikle faydalıdır. Hesaplama formülü $NDRE = (NIR - RedEdge) / (NIR + RedEdge)$ şeklindedir. Sentinel-2 verisi için NDRE, B7 (NIR, ~783 nm) ve B5 (RedEdge, ~705 nm) bantları kullanılarak $NDRE = (B7 - B5) / (B7 + B5)$ olarak elde edilir. NDRE değeri de -1 ila +1 aralığında değişir ve yüksek değeri olgun bitki örtüsünü gösterir.

Ön işleme aşamasında, uydu görüntülerindeki gürültü ve bozulmalar temizlendi. Sentinel-2 L2A verilerinde yer alan kalite bantları kullanılarak bulutlu ve gölgeli pikseller maskeleydi; böylece sadece temiz, açık piksel verisi işlendi. Kalan boşluklu gözlemler ise zaman serisi interpolasyonu (örneğin lineer veya kübik spline) dolduruldu. Ayrıca, farklı tarihlerdeki görüntüler coğrafi olarak hizalandı ve birim alan

üzerinden hesaplanan indeksler kıyaslanabilir hale getirildi. Sonuçta, her ilçe için aylık bazda eksiksiz bir indeks zaman serisi elde edildi. Bu hazırlıklar, ileride kullanılacak öğrenme modellerinin eğitiminde verinin kalitesini artırmayı hedeflemektedir.

3. Modelleme Planı

Bu aşamada hâlâ model eğitimi tamamlanmamıştır; ancak ileriki adımlar için farklı algoritma seçenekleri belirlenmiş ve test veri kümesi hazırlanmıştır. Planlanan yöntemler arasında hem klasik makine öğrenmesi hem de derin öğrenme teknikleri bulunmaktadır. Özet olarak önerilen yaklaşımlar şunlardır:Ansambl Öğrenme Yöntemleri (Random Forest, XGBoost): Random Forest (RF) ve XGBoost, çok sayıda karar ağacının bir araya gelmesiyle çalışan güçlü tahmin modelleridir. RF, rastgele örnekler ve değişken seçimleriyle oluşturulan çoklu karar ağaçları sayesinde veri集中的 karmaşık ilişkileri öğrenebilir. XGBoost ise gradyan artırma yöntemini kullanarak ardışık ağaçlar oluşturur ve hataları minimize eder. Literatürde bu yöntemlerin, uydu bitki indeksleriyle entegre edildiğinde yüksek doğruluk sağladığı gösterilmiştir. Örneğin RF ve XGBoost, verim tahmini problemlerinde yaygın olarak tercih edilmektedir (RF literatürde “non-parametrik gelişmiş ağaç temelli bir yöntem” olarak tanımlanmıştır. Zaman Serisi Modelleri (LSTM): LSTM (Uzun Kısa Süreli Bellek), ardışık verideki zaman bağımlılıklarını yakalamak için özel olarak tasarlanmış tekrarlayan sinir ağıdır. Bitki indekslerinin yıllık fenolojik döngülerine bağlı olarak gelişimi, LSTM modellerinde saklanan hücre belleğiyle öğrenilebilir. Literatürde, bitki indekslerinden elde edilen zaman serilerini kullanarak mahsul verimini tahmin etmek için LSTM ve LSTM türevlerinin en yaygın derin öğrenme yaklaşımlarından olduğu vurgulanmıştır. Bu modeller, geçmiş

aylardaki indeks verilerini değerlendirerek gelecek ürün verimini tahmin edecek şekilde yapılandırılacaktır.

4

Konvolüsyonel Sinir Ağları (1D-CNN): 1D-CNN, zaman serilerine konvolüsyon filtresi uygulayan derin ağlardır. Zaman içindeki ardışık indeks değerlerindeki yerel paternlere duyarlıdır. Örneğin bitki gelişim evresinde indeks değerlerindeki ani değişimler veya dalgalanmalar, CNN filtreleriyle yakalanabilir. Hazırlanan veri setindeki her örnek, zamana göre dizilmiş indeks değerleri olarak CNN'e beslenebilir. Bu sayede model, uzamsal değil zamansal ilişkileri öğrenir. Literatürde CNN ve LSTM modellerinin sıklıkla birlikte de kullanıldığı, bazı çalışmaların 3D-CNN veya CNN-LSTM hibrit yaklaşımları test ettiği görülmüştür.

Model Değerlendirme: Geliştirilecek modellerin performansı için RMSE (Kök Ortalama Kare Hata), R^2 (Determinasyon Katsayısı) gibi metrikler kullanılacaktır. Eğitim-veri ve test-veri ayrımı (örneğin zaman bazlı kronolojik bölme veya çapraz doğrulama) uygulanarak model genelleme yeteneği incelenecek, aşırı öğrenme kontrolü sağlanacaktır. Hiperparametre optimizasyonu için ızgara arama veya rasgele arama yöntemleri değerlendirilecek, model karmaşıklığı ile tahmin başarısı arasındaki denge gözetilecektir. Bu kapsamlı planla, önerilen tüm algoritmaların en uygun koşullarda sınanması hedeflenmektedir.

4. SONUÇLAR

Şu ana kadar projenin veri hazırlama ve planlama aşamaları tamamlanmıştır. Oluşturulan veri seti eksiksiz ve etiketli hale getirilmiş, test veri kümesi belirlenmiştir. Henüz model eğitimi gerçekleşmemiş olmakla birlikte, raporda literatür desteğiyle bir dizi algoritma önerisi sunulmuştur. İkinci dönem itibarıyla elde edilen ara sonuçlar, kaynak verilerle ön işleme adımlarının başarılı olduğunu göstermektedir. Örneğin hesaplanan bitki indekslerinin mevsimsel değişimlere uygun dağıldığı ve USDA verileriyle uyumlu çıktığı gözlemlenmiştir (detaylı analizler ileriki raporlarda yer alacaktır).

Gelecekteki çalışmalar, bu veri seti üzerinde önerilen modellerin eğitimi ve

değerlendirmesini içerecektir. İlk olarak RF, XGBoost, LSTM ve 1D-CNN modelleri üzerinde denemeler yapılarak karşılaştırmalar yapılacaktır; ardından en başarılı model(ler) seçilerek ileri optimizasyon adımları yürütülecektir. Ayrıca veri seti kapsamı genişletilerek (daha fazla yıl veya farklı eyaletler eklenerek), ek indekslerin hesaplanması (örneğin EVI, SAVI varyasyonları) ve iklim verilerinin entegrasyonu gibi iyileştirmeler planlanmaktadır.

5

Sonuçların yorumlanması ve görselleştirme aşamalarına da ağırlık verilerek, elde edilen bulgular haritalar veya grafiklerle desteklenecektir. Bu sayede, model çıktıları tarımsal planlama ve karar destek mekanizmaları için daha anlaşılır bir formatta sunulacaktır. Rapor sonunda önerilen kaynaklardan elde edilen teorik bilgiler ışığında, çalışmanın sonraki aşamalarında saha doğrulama ve gerçek verim uygulamaları ile geçerlilik testi yapılması hedeflenmektedir.

Kaynakça

- Google Earth Engine – Birçok petabayt uydu görüntüsünü ve mekansal veri setini içeren, bulut tabanlı bir analiz platformudur earthengine.google.com.
- Panzeri et al., “A Method to Determine the Optimal Period for Field-Scale Yield Prediction Using Sentinel-2 Vegetation Indices”, Land, 2024 [mdpi.com](https://www.mdpi.com).
- SmartAfriHub, “Normalized Difference Vegetation Index (NDVI)” ng.smartafrihub.com ve “Normalized Difference Red Edge Index (NDRE)” ng.smartafrihub.com, bitki indeksleri açıklamaları (erişim: 2025).
- USGS Landsat Missions, “Landsat Soil Adjusted Vegetation Index” açıklaması usgs.gov.
- Iowa State Univ. Extension, “USDA NASS Corn and Soybean County

Yields for Iowa, 2015-2024”extension.iastate.edu.

- Muruganantham et al., “A Systematic Literature Review on Crop Yield Prediction with Deep Learning and Remote Sensing”, Remote Sensing, 2022mdpi.com.
- Dhillon et al., “Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape”, Frontiers in Remote Sensing, 2022frontiersin.org.