

TÜRKİYE'DE İKLİM VE RAKIM FAKTÖRLERİNE DAYALI BUĞDAY VERİMLİLİK ANALİZİ VE OPTİMUM ÜRÜN ÖNERİ SİSTEMİ

YÜCEL GÜL / 210541091
github.com/yucellgull
linkedin.com/in/yucellgull
yucellgull@hotmail.com

Özet

Bu çalışmada, Türkiye'nin çeşitli bölgelerinde tarımsal verimliliği etkileyen iklim, toprak ve yükseklik gibi çevresel faktörler incelenmiştir. Verimlilik tahminleri ve optimum ürün önerileri için beş farklı makine öğrenmesi modeli kullanılmıştır: Rastgele Orman, Destek Vektör Makineleri, Gradyan Artırma, K En Yakın Komşu ve AdaBoost. Modeler GridSearchCV ile optimizasyon edilmiş ve en iyi performans gösteren modeller belirlenmiştir. Sonuçlar, tarımsal üretim planlaması ve sürdürülebilir tarım uygulamaları açısından değerlendirilmiştir.

1. Projenin Amacı

Bu projenin amacı, Türkiye'nin çeşitli bölgelerinde tarımsal verimliliği etkileyen çevresel faktörlerin (rakım, güneş ışığı, nem, toprak türü, yağış miktarı ve sıcaklık gibi) analiz edilerek, bu faktörlere dayalı olarak en uygun buğday öneri sisteminin geliştirilmesidir. Proje kapsamında aşağıdaki hedeflere ulaşılması amaçlanmaktadır:

Çevresel Faktörlerin Etikelerinin Belirlenmesi: Tarımsal verimlilik üzerindeki iklim ve coğrafi faktörlerin etikelerini belirlemek. Türkiye'nin farklı bölgelerinde buğday verimliliği üzerinde hangi çevresel faktörlerin daha belirleyici olduğunu analiz etmek.

Makine Öğrenmesi Modelleri Kullanarak Tahminler Yapmak: Farklı makine öğrenmesi modelleri (Rastgele Orman, Destek Vektör Makineleri, Gradyan Artırma, K En Yakın Komşu ve AdaBoost) kullanarak tarımsal verimlilik tahminleri yapmak. Bu modellerin performansını karşılaştırarak en iyi tahmin modelini belirlemek.

Hiperparametre optimizasyonu yaparak modellerin doğruluğunu artırmak.

Elde edilen veriler ışığında, tarımsal verimliliği artırmak ve sürdürülebilir tarım uygulamalarını teşvik etmek amacıyla bir ürün öneri sistemi oluşturmak.

Tarım Politikalarına ve Uygulamalarına Katkıda Bulunmak: Çalışma sonuçlarını bölgesel tarım politikalarının geliştirilmesine ve uygulanmasına katkı sağlayacak şekilde sunmak. Sürdürülebilir tarım uygulamalarının yaygınlaştırılmasına yönelik önerilerde bulunmak.

Bu amaçlar doğrultusunda, proje sonuçlarının Türkiye'deki tarımsal üretim planlamasında ve sürdürülebilir tarım uygulamalarının geliştirilmesinde önemli bir rol oynaması hedeflenmektedir. Çalışmanın bulguları, tarım sektörü için yenilikçi yaklaşımların benimsenmesine katkıda bulunacaktır.

2. Projenin Kapsamı

Bu proje, Türkiye'nin farklı coğrafi bölgelerinde tarımsal verimliliği etkileyen çevresel faktörleri analiz etmeye ve bu faktörlere dayanarak en uygun ürünlerin belirlenmeye hedeflemektedir. Projeyi ilk adımı olarak, Türkiye'nin çeşitli bölgelerinden rakım, güneş ışığı, nem, toprak türü, yağış miktarı ve sıcaklık gibi çevresel veriler toplanmıştır. Bu veriler, Meteoroloji Genel Müdürlüğü, Wikipedia ve TÜİK gibi çeşitli kaynaklardan elde edilmiştir.[6],[7],[8]

Toplanan veriler üzerinde eksiksiz verilerin doldurulması, aykırı değerlerin temizlenmesi ve verilerin standartize edilmesi gibi veri ön işleme adımları gerçekleştirilmiştir. Bu adımlar, verilerin makine öğrenmesi modelleri için uygun hale getirilmesini sağlamıştır. Veriler, farklı makine öğrenmesi modelleri kullanılarak analiz edilmiştir.

Kullanılan modeller arasında Rastgele Orman (Random Forest) Regressor^[1], Destek Vektör Makineleri (SVM)^[2], Gradyan Artırma (Gradient Boosting) Regressor^[3], K En Yakın Komşu (KNN) Regressor^[4] ve AdaBoost Regressor^[5] bulunmaktadır. Modelerin performansını artırmak amacıyla, GridSearchCV^[6] kullanılarak hiperparametre optimizasyonu yapılmıştır.

Bu süreçte, her model için en iyi parametre kombinasyonları belirlenmiş ve modellerin performansı optimize edilmiştir. Modelerin performansı, Ortalama Kare Hata (MSE) ve R² (belirleme katsayısi) değerleri ile değerlendirilmiştir ve karşılaştırılmıştır. En iyi performans gösteren model, tarımsal verimlilik tahminlerinde kullanılmak üzere seçilmiştir.

Proje kapsamında, çevresel faktörlerin tarımsal verimlilik üzerindeki etkileri detaylı olarak analiz edilmiştir. Özellikle iklim ve coğrafi koşulların buğday gibi temel tarım ürünlerinin verimliliği üzerindeki etkileri incelenmiştir. Bu analizler sonucunda, her bölge için en uygun tarımsal ürünlerin belirlenmesi amacıyla bir ürün öneri sistemi geliştirilmiştir.

Son olarak, çalışma sonuçlarının tarımsal üretim planlaması ve bölgesel tarım politikalarının geliştirilmesinde kullanılabileceği değerlendirilmiştir. Bu kapsama, proje Türkiye'de tarımsal verimliliği artırmayı ve sürdürülebilir tarım uygulamalarının yaygınlaştırılmasına yönelik önerilerde bulunmak ve tarım sektörü için yenilikçi yaklaşımlar geliştirmek için kullanılacaktır.

3. Veri Görselleştirme ve Veri Analizi (EDA)

3.1 Buğday Türleri ve Üretim Şekillerine Göre Ortalama Verimlilik

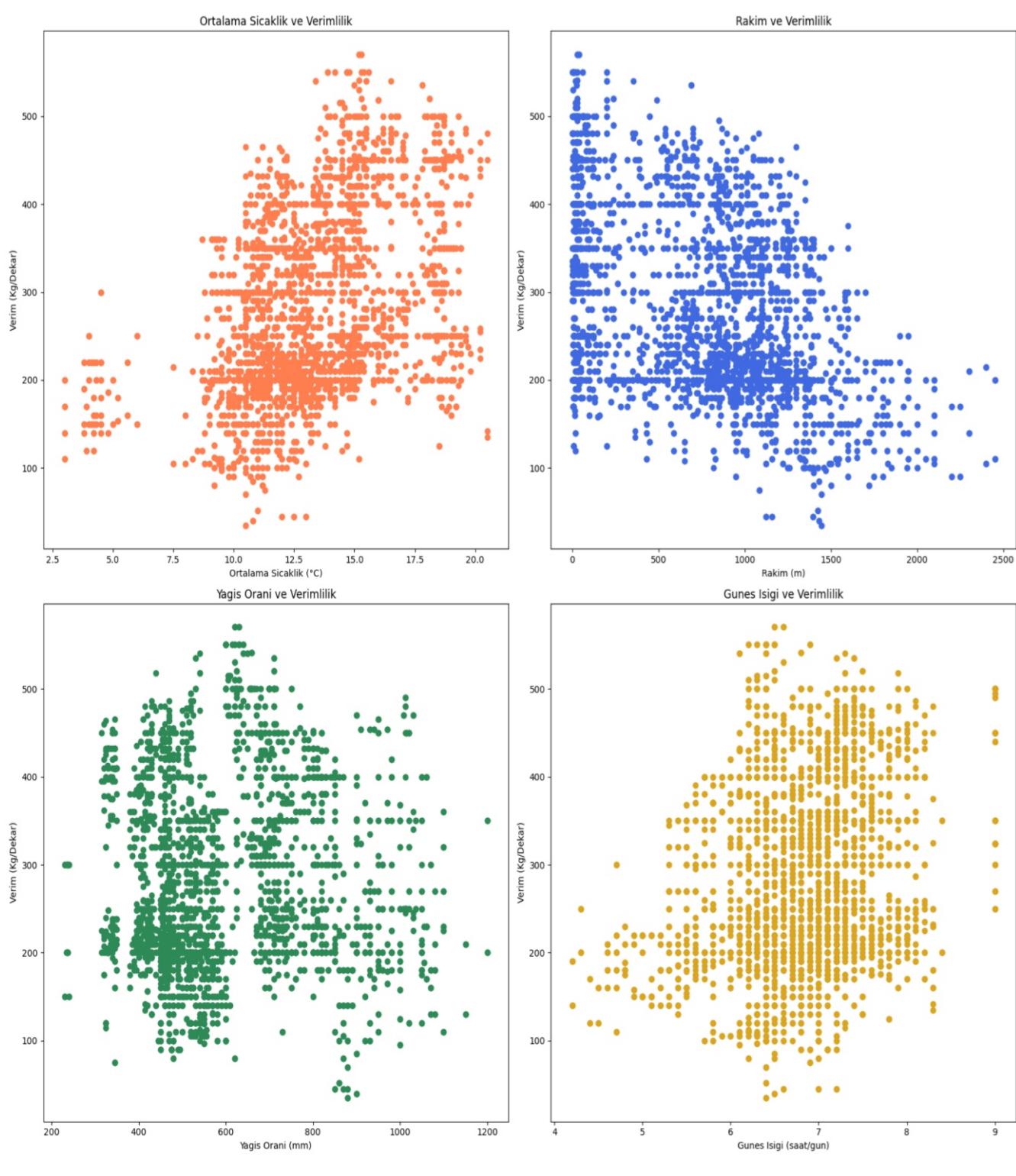
Yandaki grafikte (Şekil 3.1), farklı buğday türleri ve üretim şekillerine göre ortalama verimlilik (kg/dekar) karşılaştırılmaktadır.

Sulama Faktörü: Grafik, sulamış, hem diğer buğday türlerinden hem de durum buğdayında verimliliği önemli ölçüde arttığını göstermektedir.

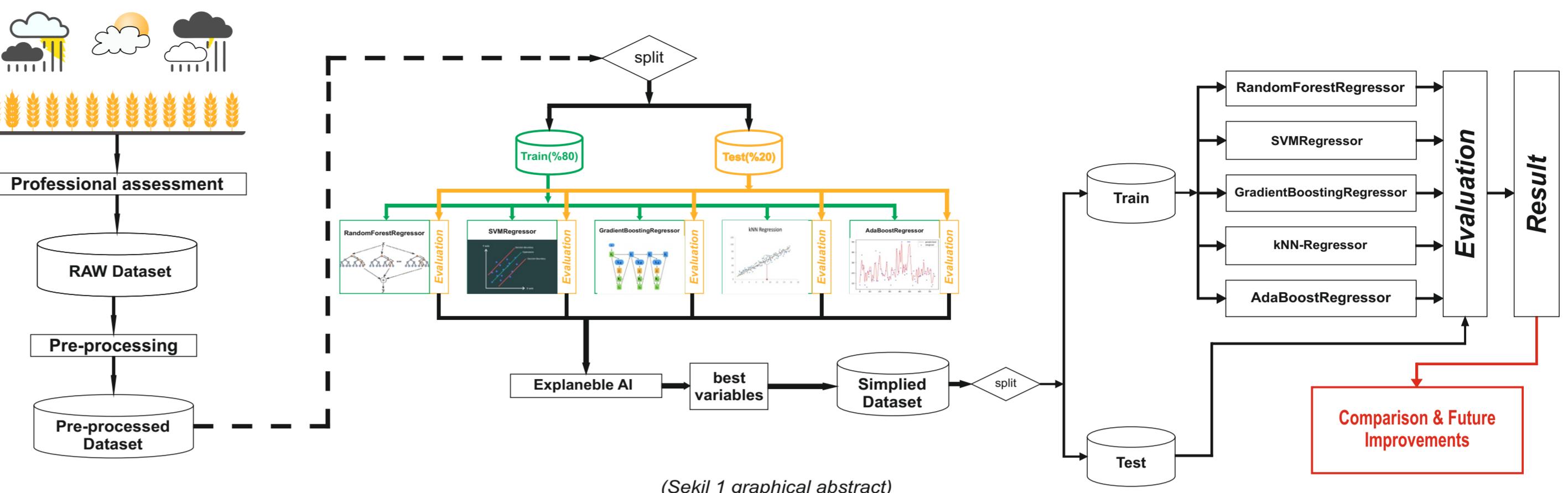
Bağışıklık Faktörleri: Kuru ve sulu tarım koşulları arasında belirgin bir fark olduğu, ancak buğday türleri arasında çok büyük bir verimlilik farkı olmadığı görülmektedir.

3.2 Çevresel Faktörler ve Verimlilik

Aşağıdaki grafikte (Şekil 3.2) ortalama sıcaklık, rakım, yağış oranı ve güneş ışığının tarımsal verimlilik üzerindeki etkileri aşağıdaki grafiklerde gösterilmiştir.



Graphical Abstract

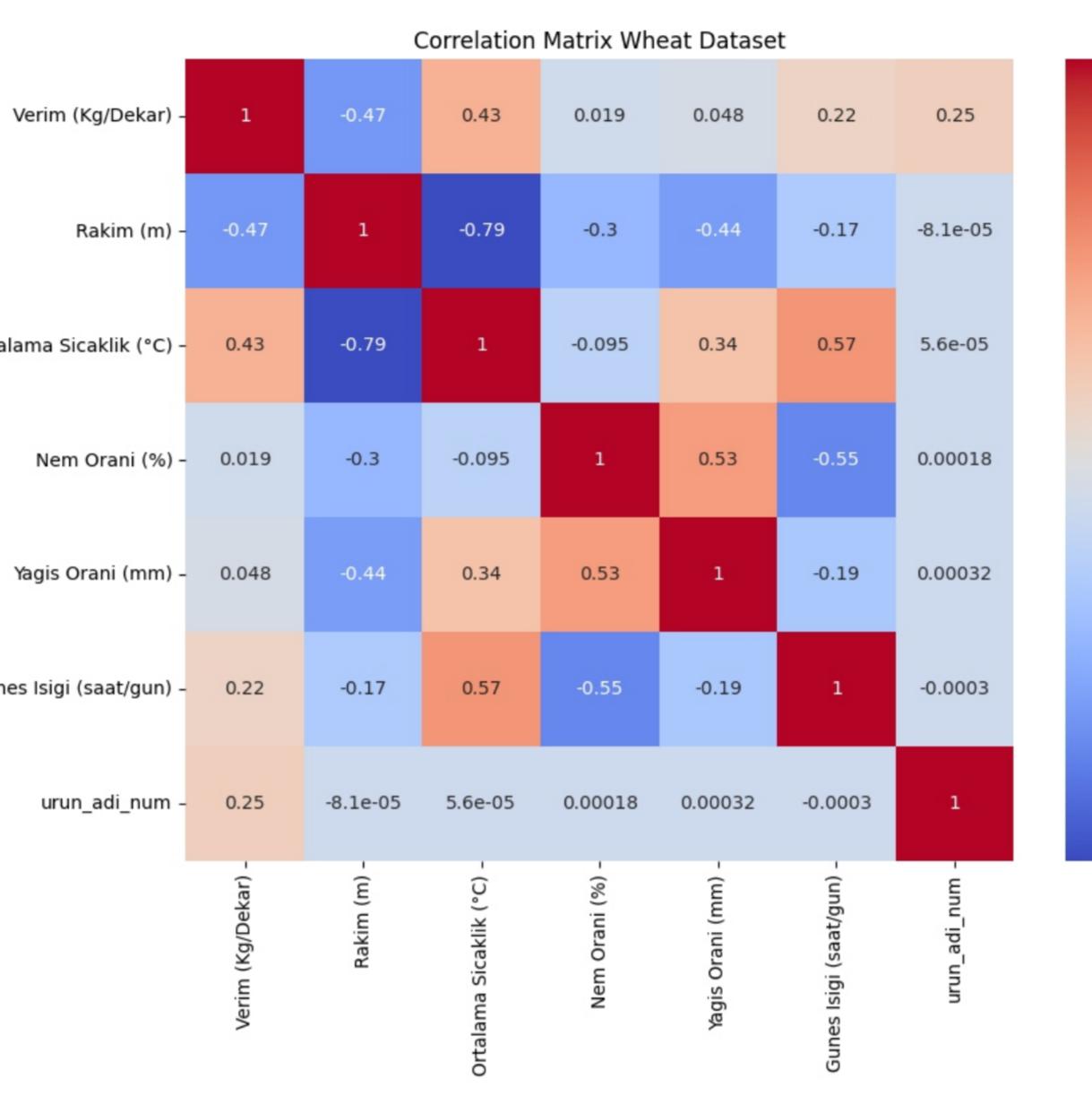


(Şekil 1 graphical abstract)

3. Veri Görselleştirme ve Veri Analizi (EDA)

3.3 Korelasyon Matrisi

Aşağıdaki grafik (Şekil 3.3) buğday verimliliği ve çeşitli çevresel faktörler arasındaki korelasyonları göstermektedir. Korelasyon katsayıları, değişkenler arasındaki ilişkinin gücünü ve yönünü ifade eder. Değerler -1 ile 1 arasında değişir, pozitif değerler pozitif ilişkiyi, negatif değerler negatif ilişkiyi gösterir. Renk skali, korelasyon katsayılarının büyüklüğünü görsel olarak belirtir.



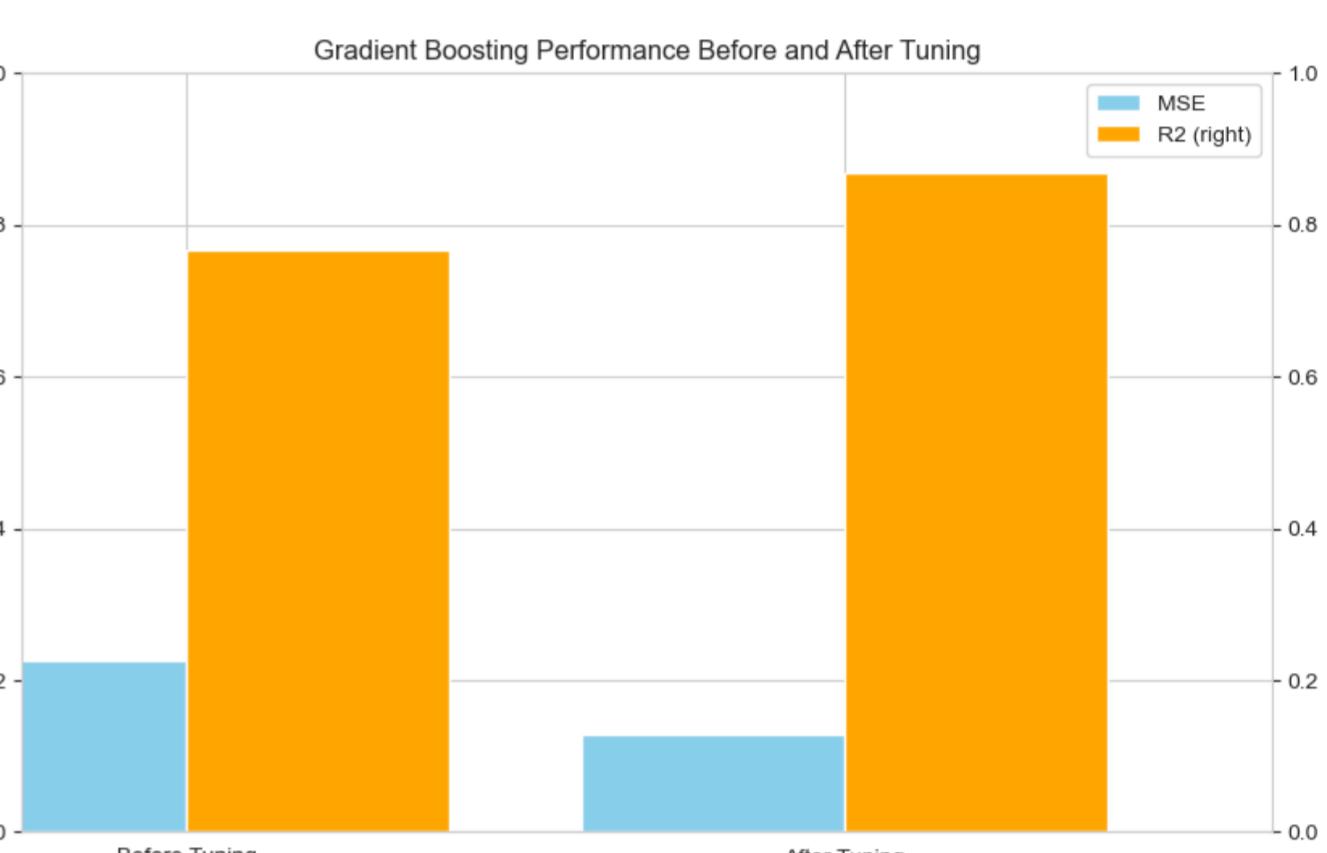
(Şekil 3.3 Buğday Veriseti Korelasyon Matrisi)

5. Sonuçlar

5.4 Hiperparametre optimizasyonu:

Gradient Boosting modelinin hiperparametre optimizasyonu öncesi ve sonrası performansı aşağıdaki grafikte gösterilmiştir. Optimizasyon, modelin doğruluğunu artırarak daha düşük MSE ve daha yüksek R² değerleri sağlamıştır.

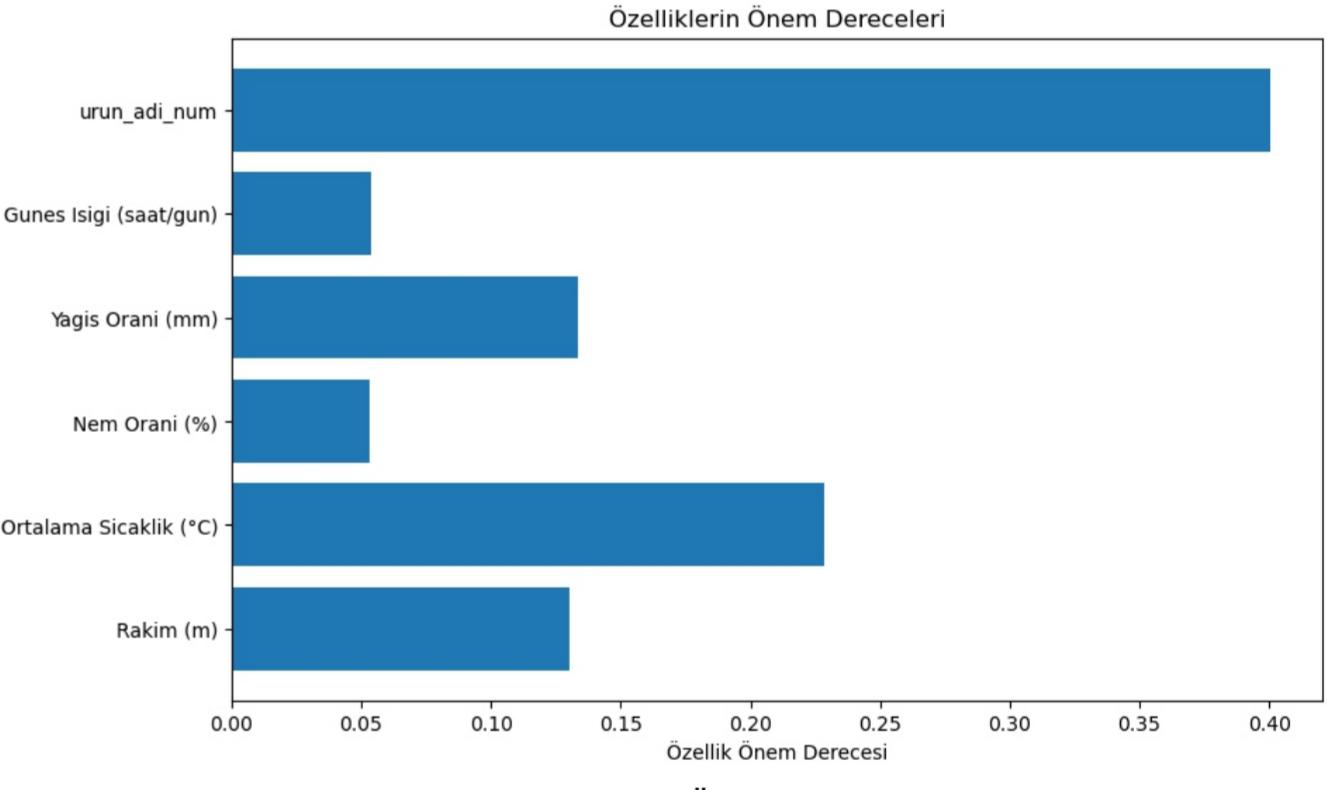
(Optimizasyon Öncesi): MSE: 0.2266, R²: 0.7674
(Optimizasyon Sonrası): MSE: 0.1287, R²: 0.8679



(Şekil 5.2 Gradient Boosting Modeli Performansının Optimizasyon Öncesi ve Sonrası)

5.5 Özneliliklerin Önem Derecesi:

Aşağıdaki grafikte, çeşitli özneliliklerin önem dereceleri gösterilmektedir. Bu önem dereceleri, her bir öznelliğin modelin tahmin gücüne ne kadar katkıda bulunduğu gösterir.



6. Öneriler

Çevresel Faktörlerin Daha Detaylı Analizi: Tarımsal verimliliği etkileyen çevresel faktörlerin daha detaylı bir şekilde incelenmesi, daha doğru ve güvenilir tahminler yapılmasını sağlayabilir. Örneğin, mikro iklim verileri, toprak kimyası ve su tutma kapasitesi gibi daha spesifik veriler toplanabilir.

Farklı Bölgeler İçin Veri Toplama: Farklı iklim ve coğrafi özelliklere sahip bölgelerden daha fazla veri toplanarak, modellerin genelleme yeteneğini artırılabilir. Bu, farklı bölgelerde uygulanabilir tarım politikalarının geliştirilmesine yardımcı olabilir.

Diğer Tarımsal Ürünler İçin Modeller: Sadece buğday değil, aynı zamanda mısır, arpa, pamuk gibi diğer önemli tarım ürünlerini de benzer modellerle yapılmalıdır. Bu, tarımsal üretimi çeşitlendirmeye ve genel verimliliği artırma yardımcı olabilir.

Uzun Vadeli Veri Setleri: Uzun vadeli iklim ve tarımsal veriler kullanılarak modellerin dayanıklılığı ve doğruluğu artırılabilir. Bu, iklim değişikliğinin tarımsal verimlilik üzerindeki etkilerini anlamak için de faydalı olabilir.

Polisilik ve Uygulamalı Araştırmalar: Elde edilen bulguların bölgeliksel tarım politikalarının geliştirilmesinde kullanılması önerilmektedir. Ayrıca, çiftçilerle işbirliği yaparak uygulamalı araştırmalar yürütmek, teorik bulguların pratikte uygulanabilirliğini test etmek açısından önemlidir.

Kaynakça

- [1] Random Forest (Regresyon): Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
- [2] Destek Vektör Makineleri (SVM): Vapnik, V. (1995). The nature of statistical learning theory. Springer, 156-171.
- [3] GradientBoostingRegressor: Friedman, J. H. (2000). Greedy boosting and additive logistic regression. Annals of Statistics, 28(2), 387-406.
- [4] KNN: Fix, E., & Hodges, J. L. (1951). Nonparametric discrimination: Distribution free classification. Annals of Mathematical Statistics, 22(3), 361-374.
- [5] AdaBoostRegressor: Schapire, R. E., & Freund, Y. (1997). A boosting framework for predictive learning. Machine learning, 37(2-3), 235-263.
- [6] Türkiye il ve ilce buğday verimliliği: <https://data.tuik.gov.tr>
- [7] Türkiye'deki iklim ve hava durumu verileri: <https://mgm.gov.tr/>
- [8] Türkiye'deki rakım verileri: <https://tr.wikipedia.org/>
- [9] GridSearchCV: <https://scikit-learn.org/>
- [10] Ek kaynaklar: <https://medium.com/>.

