

GAZİ

JOURNAL OF ENGINEERING SCIENCES

Agricultural Product Recommendation System for Isparta Province Using Machine Learning Techniques

Gamze Yakut^a, Rabia İlknur Çay^b, Hasan Hüseyin Öztürk^c

Submitted: 06.12.2023 Revised: 15.12.2023 Accepted: 18.12.2023 doi:10.30855/gmbd.0705S17

ABSTRACT

Keywords: soil fertility, artificial intelligence, randomforest, bagging algorithm

^a Isparta Uygulamalı Bilimler Üniversitesi, OSB MYO, Mekatronik Programı, Isparta, Türkiye
Orcid: 0000-0002-8614-304X
e mail: gamzeyakut@isparta.edu.tr

^b Isparta Uygulamalı Bilimler Üniversitesi, Teknoloji Fakültesi, Mekatronik Mühendisliği, Isparta, Türkiye
Orcid: 0000-0002-2894-3540

^c Isparta Uygulamalı Bilimler Üniversitesi, Teknoloji Fakültesi, Mekatronik Mühendisliği, Isparta, Türkiye
Orcid: 0009-0006-7082-1543

^{*}Corresponding author:
gamzeyakut@isparta.edu.tr

In rural areas, the failure to cultivate crops that are suitable for the soil's nutritional values and unwise practices like improper fertilization/irrigation can have negative consequences for both soil quality and crop yield. Soil analyses are important for determining the specific nutrient value ratios that agricultural areas possess. Farmers deciding what to plant in their fields without using soil analysis methods can lead to reduced crop yields in the following years. When choosing the crop to be harvested, parameters such as organic matter, pH, salinity, saturation, depth, potassium, phosphorus, and lime content should be examined. In addition, factors related to climatic and geographical characteristics such as sunlight, humidity, rainfall and altitude should also be taken into account when deciding which crop to grow in a particular region. This study aims to suggest the most productive crops to be grown in the desired region, taking into account the parameters analysing the soil structure and location-based climatic characteristics. To create a better model, exploratory data analysis methods were employed. Algorithms such as Random Forest, XG Boost and Bagging were used to make crop recommendations based on soil values. Among these models, it was observed that Bagging algorithm made the most accurate prediction with 63% accuracy rate. Additionally, a complexity matrix was used to evaluate the accuracy of our classification system.

Makine Öğrenimi Teknikleri Kullanılarak Isparta İli İçin Tarımsal Ürün Önerme Sistemi

ÖZ

Kırsal bölgelerde, toprağın besin değerlerine uygun ürünlerin yetiştirilmemesi ve bilinçsiz gübreleme/sulama gibi uygulamalar, toprak kalitesi ve ürün verimi açısından olumsuz sonuçlar doğurmaktadır. Toprak analizleri, tarım alanlarının sahip olduğu özgün besin değeri oranlarını belirlemek için önemlidir. Çiftçilerin, toprak analiz yöntemlerini kullanmadan tarlalarında ne ekeceklerine karar vermeleri, ilerleyen yıllarda hasat edilecek ürünlerin verimini düşürebilir. Hasat edilecek ürünü seçerken organik madde, pH, tuzluluk, saturasyon, derinlik, potasyum, fosfor ve kireç gibi parametreler incelenmelidir. Ayrıca belirli bir bölgede hangi ürünün yetiştirileceğine karar verirken, güneş ışığı, nem, yağış ve rakım gibi iklim ve coğrafi özelliklerle ilgili faktörler de dikkate alınmalıdır. Bu çalışma, toprak yapısını analiz eden parametreleri ve konum bazlı iklim özelliklerini göz önüne alarak istenen bölgede yetiştirilecek en verimli ürünleri önermeyi amaçlamaktadır. Daha iyi bir model oluşturmak için keşifsel veri analizi yöntemleri kullanılmıştır. Toprak değerlerine dayalı olarak ürün önerileri yapmak için Random Forest, XG Boost ve Bagging gibi algoritmalar kullanılmıştır. Bu modeller arasında Bagging algoritmasının %63 doğruluk oranı ile en doğru tahminlemeyi yaptığı görülmüştür. Ayrıca, sınıflandırma sistemimizin doğruluğunu değerlendirmek için bir karmaşıklık matrisi kullanılmıştır.

Anahtar Kelimeler: toprak verimi, yapay zekâ, randomforest, bagging algoritması

1. Giriş (Introduction)

İnsanlık var olduğu günden beri doğrudan veya dolaylı olarak tarımla iç içe olmuştur. Temel ihtiyaçlarını karşılamının yanı sıra yaşam kalitesini daha da arttırmak için tarımı sürekli olarak kullanmıştır. Yaşadığı zamanın imkân tanıdığı teknolojik gelişmeleri çağın ihtiyaç ve gereklerine uygun bir şekilde tarımsal faaliyetlere entegre etmiştir. Böylece yaşadığı çevreyi sürekli olarak geliştirmiştir.

Tarım; bitkisel ve hayvansal mamullerin üretilmesi, işlenmesi, depolanması ve pazarlanması süreçlerini kapsayan bir bilim dalıdır [1],[2],[3]. İnsanlığın ilk zamanından bu yana tarım kültürü; üretim-tüketim faaliyetleri, tarım arazisi kullanım şekli gibi parametrelerde çeşitli evrelerden geçmiştir. İlk çağlarda ilkel yollarla ve tesadüflerle başlayan tarımsal üretim yöntemlerinin yerini günümüzde modern teknolojinin kullanıldığı teknikler almıştır [4]. Çağımızın en yeni gelişimi olan yapay zekâ sistemleri sağlık, eğitim, tarım, ulaşım, savunma sanayi gibi pek çok alanda öncelikli olarak kullanılmaktadır. Spesifik olarak tarımda; bitki hastalıkları, üretim planlamaları, zirai ürün- gübre kullanma miktarları ve periyotları, verimlilik tespiti, tarım robotları uygulamaları, sulama zamanı gibi konularda kısacası ekimi yapılan ürünün, tüketicinin eline geçene kadarki olan tüm süreçlerinde yapay zekâ kullanılmaktadır [5].

Günümüzde insan nüfusu geçmiş yıllara göre hızla artmaktadır [6]. Doğal kaynakların sürekli ve dikkatsizce kullanımı çevrenin tahrip edilerek bozulmasına ve hem insanlığın hem de diğer canlıların gelecek nesillerini tehdit etme noktasına ulaşmasına neden olmuştur [7]. Dünya üzerindeki tarıma elverişli alanların bu denli hızlı ve geri dönüşü zor bir şekilde tükeniyor olması, kullanılabilir su kaynaklarının azalması tarım kaynaklı yaşanan sorunlara çözüm arayışı konusunda ulusları birleştirmiştir [8],[9],[10]. Bu amaçla teknolojinin tarımla birlikte çalışmasını sağlayan sürdürülebilir tarım uygulamaları kullanılmaktadır. Böylece yeryüzü kaynakları akıllıca kullanılmış olacak ve uzun vadede doğal kaynaklar korunacaktır. Ayrıca üst seviyede tarımsal verimlilik de sağlanacaktır [5],[10],[11],[12]. Sürdürülebilir tarım uygulamalarının amaçlarına bakılacak olursa; insanların temel ihtiyaçlarını karşılamak, doğayı koruyup devamlılığını sağlamak, su ve toprak kalitesinin düşmesini engellemek, yenilenemeyen kaynakları daha nitelikli bir şekilde kullanmak, biyolojik çeşitliliği korumak, insanlığın yaşam kalitesini arttırmak gibi maddeler sayabilmektedir [11].

Her tarım alanının kendine ait karakteristik özellikleri vardır. Bu ayrıştırıcı özellikler sayesinde toprakta yetişen ürünlerin verimliliğinde de farklılıklar meydana gelmektedir. Toprak yapısına uygun ekim yapmamak, daha çok gübrelemeye neden olmakla birlikte istenilen verimin alınmadığı sonuçlar da doğurabilmektedir. Üstelik bu durum zaten kısıtlı olan su gibi doğal kaynakların tükenmesine neden olmaktadır [13]. Bu sorunların önüne geçebilmek için toprak uygun aralıklarda analiz ettirilmeli ve analiz sonuçları iklimsel bağlamda da değerlendirilmelidir. Böylece toprağa en uygun ürünün üretimine karar vermek sezonda elde edilen ürünün kalitesini ve miktarını arttırmaktadır. Artan ürün miktarı ve kalitesi çiftçiye ticareti sırasında dikkate değer ekonomik getiri sağlamaktadır.

Tüm bunlara rağmen yapılan bir araştırmaya göre toprak analiz sonuçlarında gübre seçimini yapan üreticilerin oranı %12.24'tür [14]. Üreticilerin %33.67'si gübre miktarını belirlerken kendi bilgi ve tecrübesine göre karar vermektedir. Üreticilerin düzenli olarak toprak analizi yaptırmamalarının nedenleri incelendiğinde; % 30.14' ü toprak analizini önemsemedikleri için, %20.15'i masraflı olduğu için, %19.18'i analiz sonuçlarına güvenmediği için, %16.44'ü toprak örneği nasıl alınır bilmediği için ve %13.70' i toprak analizini yaptırmamanın uzun sürdüğü gerekçesiyle toprak analizi yaptırmadıklarını belirtmişlerdir [14]. Bu istatistikî gerekçeler dikkate alındığında çiftçi ne kadar toprak yapısına uygun uygulamalar yaparsa hasatta elde edeceği verimi arttırmaktadır. Aynı zamanda tarımsal süreçte gübreleme-sulama gibi ekstra masraf olarak yaptığı harcamaların da önüne geçmiş olmaktadır. Bu amaçla bu çalışmada çiftçi tarafından ilgili kurumlarca analizi yapılan toprağın bünyesinde bulunan organik madde, pH, tuzluluk, satürasyon, derinlik, potasyum, fosfor ve kireç parametrelerini uygun aralıklarda kategorize ederek değerlendirilmektedir. Ek olarak lokasyon-iklim özellikleri de incelenip tarım alanının bulunduğu coğrafi koşulların iklim parametreleri de dikkate alınmaktadır. Bu sayede üretici mevcut tarım alanından en yüksek verimi alabileceği ürünü belirleyebileceği gibi aynı zamanda tercih edebileceği ürünler arasından seçim de yapabilmektedir. Bu seçim ürün kalitesi ve miktarını artırmanın yanı sıra toprak bünyesinin kalitesini de bozmadığı hatta desteklediği için gelecek yıllarda yapılacak olan tarım faaliyetlerine de yatırım sağlamaktadır. Ayrıca tarımı yapılacak ürün, toprağın yapısına göre seçileceği için gübreleme-sulama gibi ürün elde etme sürecinde çiftçiye ek masraf olan kalemlerin miktarında azalma sağlamaktadır. Çiftçi bu sayede ticaret yaparken elde ettiği ürünlerin kalitesindeki ve verimli ürün sayısındaki artış ile kazanmaktadır. Aynı zamanda ek giderlerinde azalma meydana geldiği için masraflardan kâr etmiş olmaktadır. Bir başka bakış açısıyla da toprak kendine yakın

olan ürünü yetiştirirken zorlanmamış ve diğer besin değerlerince de sömürülmemiş olduğu için gelecek yıllar için kendini hazırlayabilecektir.

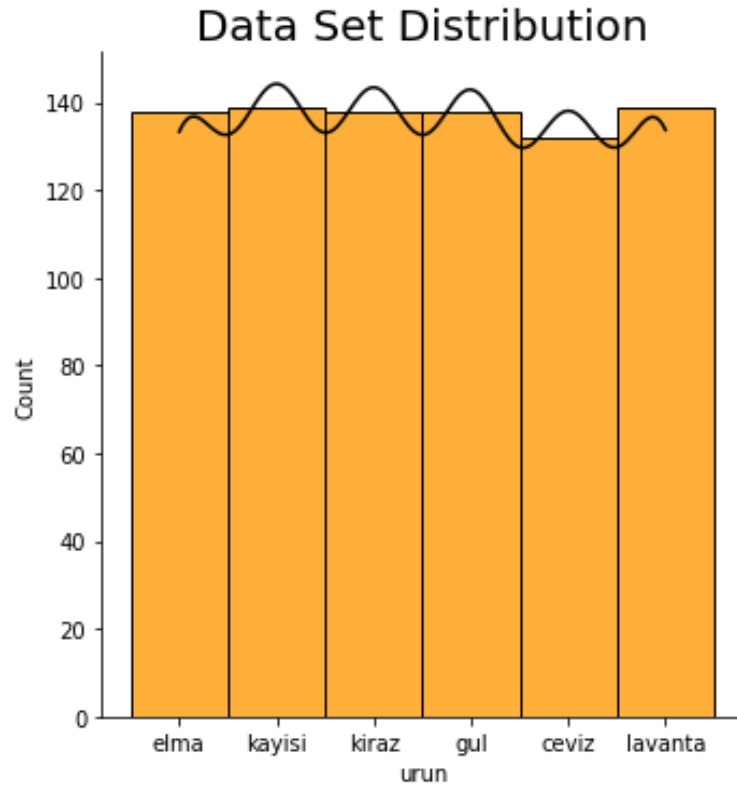
2. Materyal-Metot (Material-Method)

Bu çalışma ile topraktan alınan verilerin yapay zekâ yardımıyla mahsule uygunluğunun araştırılması gerçekleştirilmektedir.

2.1. Materyal (Material)

2.1.1. Toprak analiz veriseti (Soil analysis dataset)

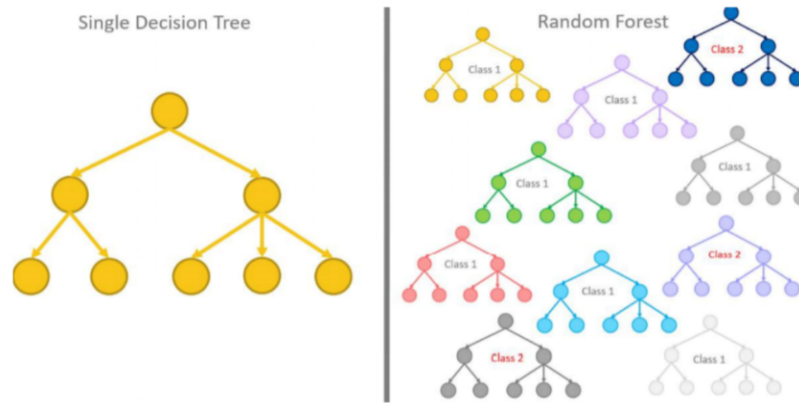
Isparta İl Tarım ve Orman Müdürlüğü ve Isparta İl Özel İdare'si tarafından sağlanan 7500+ gerçek toprak verisi analizi yapılan tarım alanlarının karakteristik özelliklerini yansıtan parametreleri (il, ilçe, köy, tarım şekli, derinlik, potasyum, fosfor, kireç, organik madde, toplam tuz, pH, saturasyon, ürün) içerecek şekilde Excel tablosu haline getirilmiştir. Veri seti dağılımı Şekil 1'de verilmiştir.



Şekil 2. Veri seti dağılımı (Dataset distribution)

2.1.2. Random Forest algoritması (Random Forest algorithm)

Random Forest en popüler makine öğrenimi yöntemlerinden birisidir. Random Forest algoritması, rastgele karar ağaçlarının birleşiminden oluşan denetimli sınıflandırma algoritmalarındandır. Hem sınıflandırma hem de regresyon problemlerinin çözümünde kullanılmaktadır. Random Forest algoritmasının temel amacı sınıflandırma işlemleri esnasında birden fazla karar ağacı üreterek sınıflandırma değerini yükseltmektir. Bir başka deyişle birbirinden bağımsız olarak çalışan birçok karar ağacının bir araya gelerek aralarından en yüksek puan alan değerini seçilmesi işlemidir [15][16].



Şekil 2. Random Forest algoritması (Random Forest algorithm)

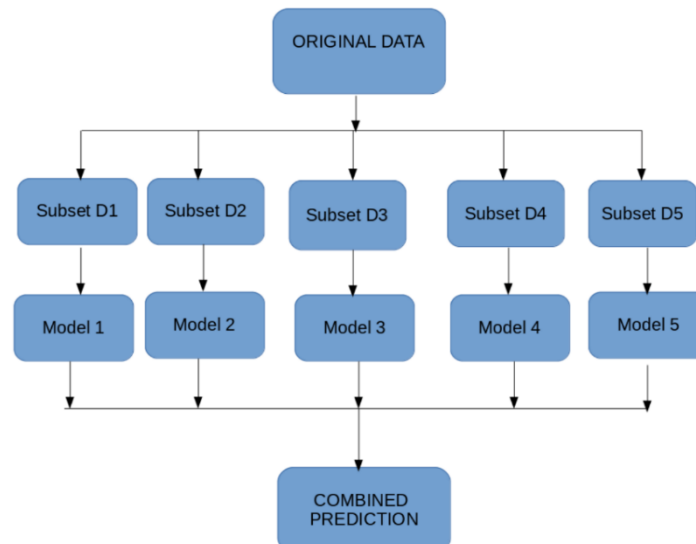
2.1.3. XG Boost algoritması (XG Boost algorithm)

XG Boost, diğer araştırma alanlarında karmaşık süreçlerin modellenmesinde çok güçlü olduğu kanıtlanmış bir makine öğrenimi algoritmasıdır. XG Boost (eXtreme Gradient Boost) yöntemi ilk olarak 2016 yılında Chen tarafından tanıtılmış ve otomotiv üretimi, bina soğutma yükünün tahmin edilmesi ve HVAC sistemleri için arıza tespiti gibi diğer birçok uygulamada kullanılmıştır. Temelde XG Boost algoritması Gradient Boosting algoritmasının optimize edilmiş bir türüdür. XG Boost, ağacı oluştururken maksimum derinlik değerini kullanır. Oluşturulan ağaç, aşağı yönde aşırı ilerleme gösterirse, budama gerçekleştirilir. Böylece aşırı öğrenmenin önüne geçilir. Gradient Boosting algoritması, kayıp fonksiyonun hesaplanmasında birinci dereceden fonksiyon kullanırken, XG Boost bu hesaplamaları ikinci dereceden fonksiyonlar kullanarak gerçekleştirir. Ayrıca paralel çalışma özelliği, diğer algoritmalara göre sonuca daha kısa sürede ulaşılmasını sağlar [17] [18].

2.1.4. Bagging algoritması (Bagging algorithm)

Bagging yöntemi Breiman tarafından 1996 yılında geliştirilmiştir. Bagging, orijinal eğitim veri kümesinin rastgele örneklenmiş alt kümeleri üzerinde bireysel öğrencileri eğitmek için kullanılan bir yöntemdir. Çalışma mantığı,

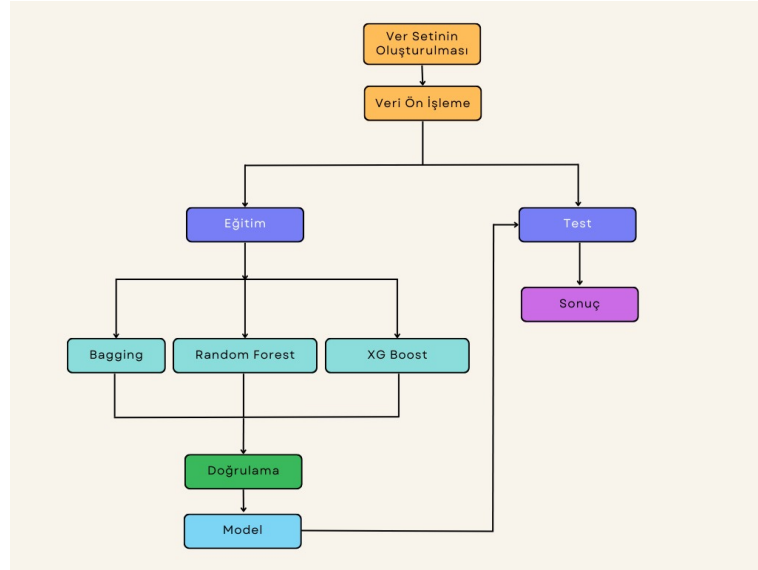
- Orijinal veri kümesinden çoklu alt kümeler oluşturulur,
- Bu alt grupların her birinde bir temel model (zayıf model) oluşturulmuştur,
- Modeller paralel olarak çalışır ve birbirinden bağımsızdır,
- Nihai tahminler, tüm modellerden gelen tahminler birleştirilerek belirlenir, şeklindedir [19][20].



Şekil 3. Bagging algoritması (Bagging algorithm)

2.2. Metot (Method)

Bu çalışmanın geliştirilme aşaması; veri setinin oluşturulması ve etiketlenmesi, arayüz tasarımı ve yapay zekâ yazılımı olarak üç kısımda incelenmektedir. Arayüz tasarımı için Qt Designer programı, yapay zekâ yazılımının gerçekleştirilmesinde Python programlama dili kullanılmıştır. Çalışmaya ait iş akış diyagramı Şekil 4'te verilmiştir.



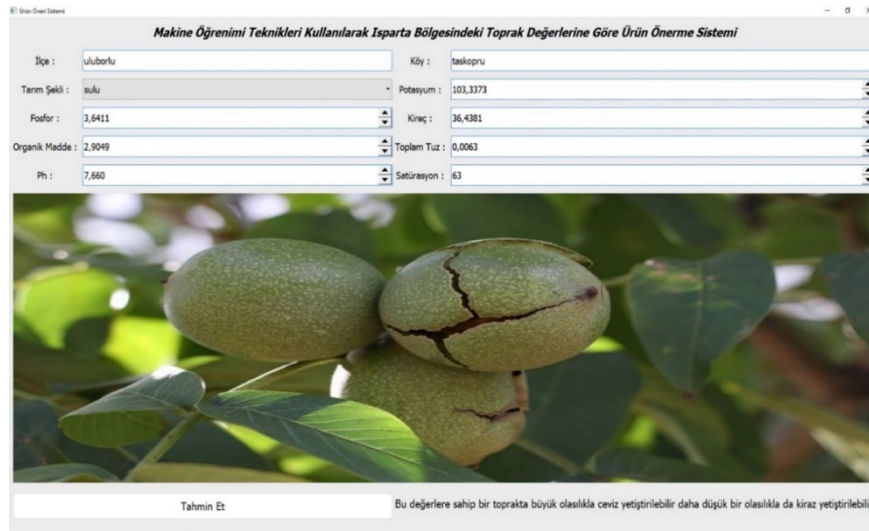
Şekil 4. İş akış şeması (Workflow diagram)

2.2.1. Veri setinin oluşturulması (Creation of the data set)

Gerçek tarım alanlarında yapılmış olunan analiz sonuçlarından Excel tablosu oluşturularak kullanıma uygun bir hale getirilmiştir. Veri setinde en çok bulunan 5 ürün rastgele 139'ar tane ve 1 ürün rastgele 130 tane seçilmiştir. Veri setinde sırası ile “elma, kayısı, kiraz, gül, ceviz ve lavanta” verileri yer almaktadır.

2.2.2. Qt designer ile arayüz tasarımı (Interface design with Qt designer)

Çalışmanın kişiler tarafından rahatça kullanılabilmesi amacıyla Qt Designer programında veri setine uygun değerlerin girilebildiği bir arayüz gerçekleştirilmiştir. Veri setinde yer alan “ilçe, köy, tarım şekli, potasyum, fosfor, kireç, organik madde, toplam tuz ve pH” değerlerinin girilebildiği uygun “line edit” yapıları kullanılmıştır. Girilen değerlerin tahmini yapıldığında en uygun sonuç görseli arayüzde gösterilmektedir. Arayüz Şekil 5'te verilmiştir.



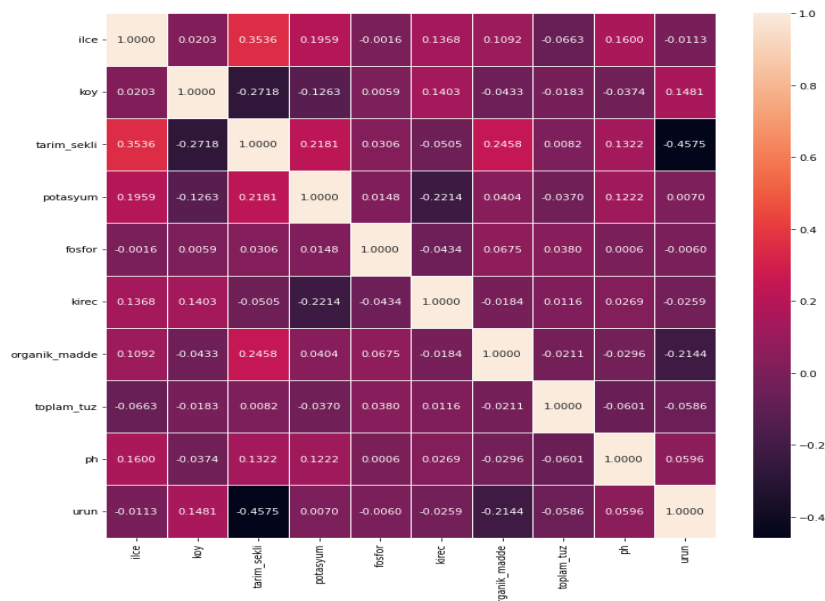
Şekil 5. Kullanıcı arayüzü (User interface)

2.2.3. Yapay zekâ algoritmaları ve denemesi (Artificial intelligence algorithms and experimentation)

Çalışmada, korelasyonu artırmak ve daha iyi bir tahmin modeli geliştirmek için veri seti üzerinde keşifsel veri analizleri (aykırı değerlerin çıkarılması, eksik verilerin tamamlanması, vb.) uygulanmıştır. Random Forest, XG Boost ve Bagging algoritmaları kullanılarak toprak değerlerine göre (ilçe, köy, tarım şekli, potasyum, fosfor, kireç, organik madde, toplam tuz ve pH) toprağın yatkın olduğu bitki (ürün) tahmini gerçekleştirilmiştir.

3. Bulgular (Findings)

Çalışmaya ait korelasyon grafiği Şekil 6'da verilmiştir. Korelasyon grafiği incelendiğinde verilerin "ürün" sütunu ile benzerliği ya çok düşük ya da negatiftir. Bu, modelin öğrenimini zorlaştıran bir durumdur. Bu tarz veriler ile eğitim yapılırken ensemble yöntemler kullanmak daha güçlü tahminler elde edilmesine olanak tanır.



Şekil 6. Veri seti korelasyon grafiği (Data set correlation chart)

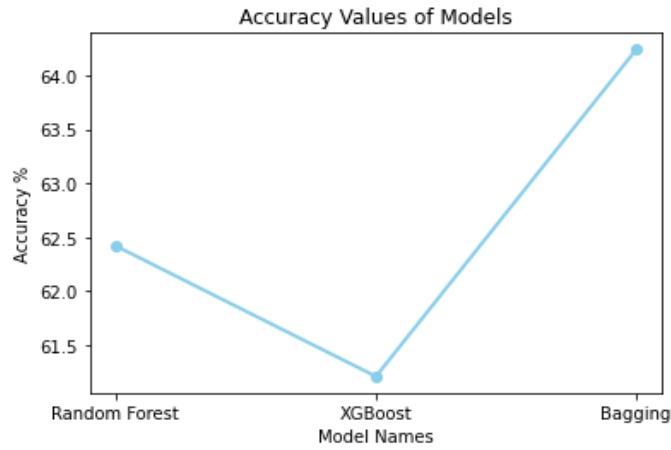
3.1. Model performanslarının metriklerle karşılaştırması (Comparison of model performances with metrics)

Çalışmada, Isparta bölgesindeki toprak verilerine ait "ilçe, köy, tarım şekli, potasyum, fosfor, kireç, organik madde, toplam tuz ve pH" değerleri giriş ve "ürün" değerleri de çıkış parametresi olarak belirlenmiştir. Python programlama dilinde kodlanan yazılım ile veri seti; Random Forest, XG Boost ve Bagging sınıflandırma algoritmaları ile tahminlenmiştir. Bu tahminleme sonucunda accuracy değeri en iyi iki algoritma Random Forest ve Bagging olarak bulunmuştur. F1_score'a bakıldığında Bagging algoritması daha iyi iken Random Forest biraz daha başarısız olmuştur. Şekil 6'daki korelasyon için gerçekleştirilen tahminlemeye ait sonuçlar tablo 1'de verilmiştir.

Tablo 1. Korelasyon için gerçekleştirilen tahmin sonuçları (Estimation results for correlation)

Performans ölçütü	Random Forest	XG boost	Bagging
Accuracy	0.6242	0.6121	0.6424
F1_score	0.6113	0.6071	0.6368
R ²	0.0401	0.0004	0.061

Şekil 6'daki korelasyon için model değerlendirme grafiği Şekil 7'de verilmiştir.

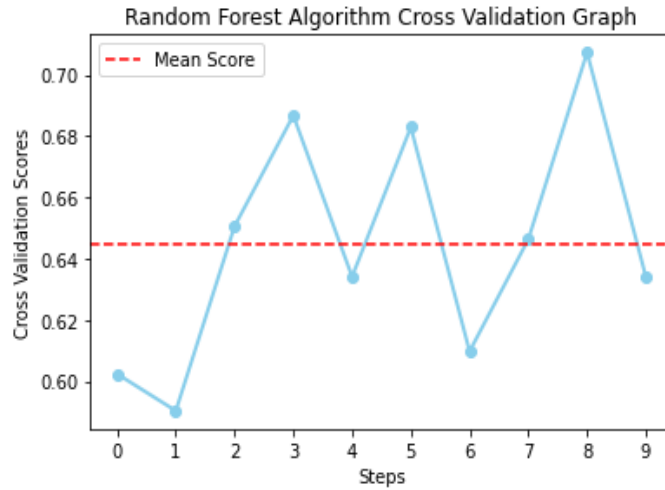


Şekil 7. Model değerlendirme grafiği (Model evaluation chart)

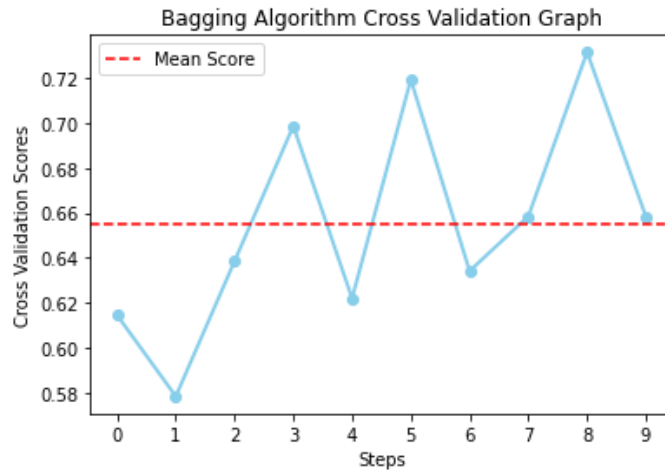
Yukarıdaki tablo ve grafiklerde görüldüğü gibi korelasyonun çok düşük olduğu veri setinde, modeller iyi bir tahminde bulunmuşlardır. Modeller F1_score performans değerlendirme ölçütüne göre karşılaştırıldığında Random Forest %61, XG Boost %60 ve Bagging algoritmasının %63 ile en doğru tahminlemeyi yaptığı görülmektedir. Bagging algoritmasındaki hiper parametre olarak kullanılan model karar ağacı, ağaç sayısı 100, maksimum örneklem 0.8, ve rastgelelik 1 olarak ayarlanmıştır.

3.1.1. Çapraz doğrulama (Cross validation)

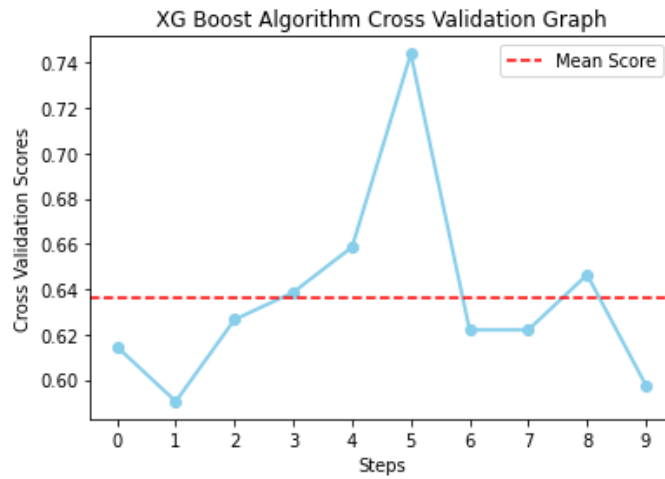
Gerçekleştirilen çalışmada veriler, 10 parça çapraz doğrulama yöntemi ile eğitilerek teste tabi tutulmuştur. Alt küme sayısı 10 olarak seçilmiştir. Random Forest algoritması için ortalama skor 0.6444, Bagging algoritması için ortalama skor 0.6554 ve XG Boost algoritması için ortalama skor 0.6360 olmuştur. Random Forest, Bagging ve XG Boost algoritmaları için çapraz doğrulama grafikleri şekil 8, 9 ve 10'da verilmiştir.



Şekil 8. Random Forest algoritması çapraz doğrulama grafiği (Random Forest algorithm cross-validation graph)



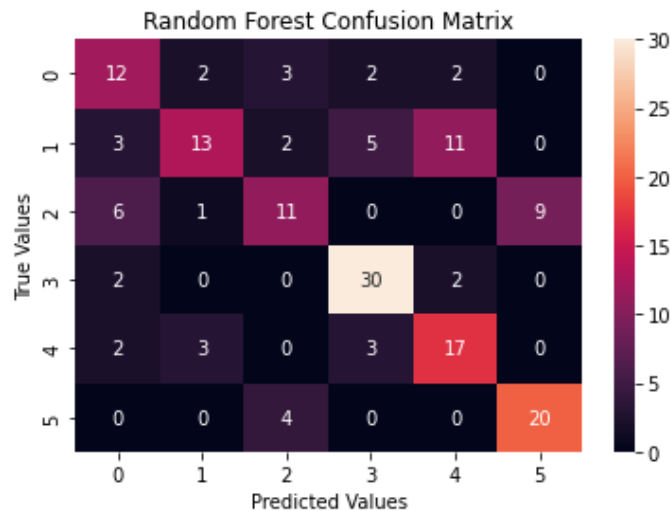
Şekil 9. Bagging algoritması çapraz doğrulama grafiği (Bagging algorithm cross-validation chart)



Şekil 10. XG Boost algoritması çapraz doğrulama grafiği (XG Boost algorithm cross-validation chart)

3.1.2. Karmaşıklık matrisi karşılaştırması (Complexity matrix comparison)

Bu çalışmada sınıflandırmanın doğruluğunu hesaplamak için kullanılan bir diğer yöntem karmaşıklık (confusion) matrisidir. Random Forest modeli, label encoding değeri 0 olan ürünü %48 doğru tahmin ederken label encoding değeri 1 olan ürünü %68, label encoding değeri 2 olan ürünü %55, label encoding değeri 3 olan ürünü %81, label encoding değeri 4 olan ürünü %53 ve label encoding değeri 5 olan ürünü de %68 doğru tahmin etmiştir. Şekil 11'de Random Forest algoritmasına ait karmaşıklık matrisleri verilmiştir.



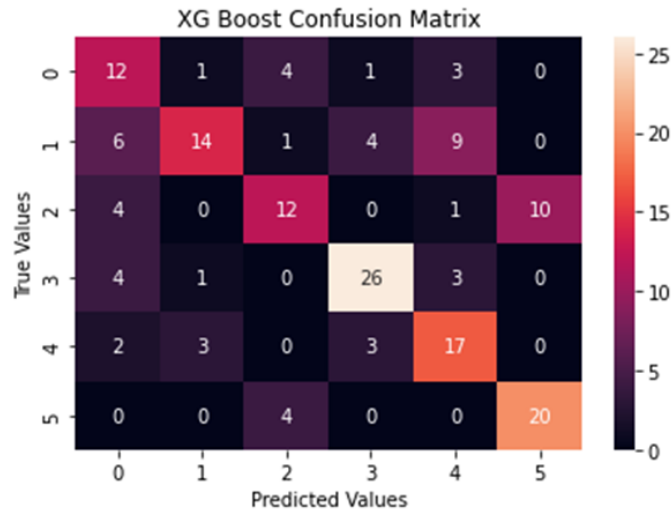
Şekil 11. Random Forest algoritması karmaşıklık matrisi ısı haritası grafiği (Random Forest algorithm complexity matrix heat map chart)

Bagging modeli, label encoding değeri 0 olan ürünü %55 doğru tahmin ederken label encoding değeri 1 olan ürünü %78, label encoding değeri 2 olan ürünü %53, label encoding değeri 3 olan ürünü %77, label encoding değeri 4 olan ürünü %52 ve label encoding değeri 5 olan ürünü de %68 doğru tahmin etmiştir. Şekil 12'de Bagging algoritmasına ait karmaşıklık matrisleri verilmiştir.



Şekil 12. Bagging algoritması karmaşıklık matrisi ısı haritası grafiği (Bagging algorithm complexity matrix heat map chart)

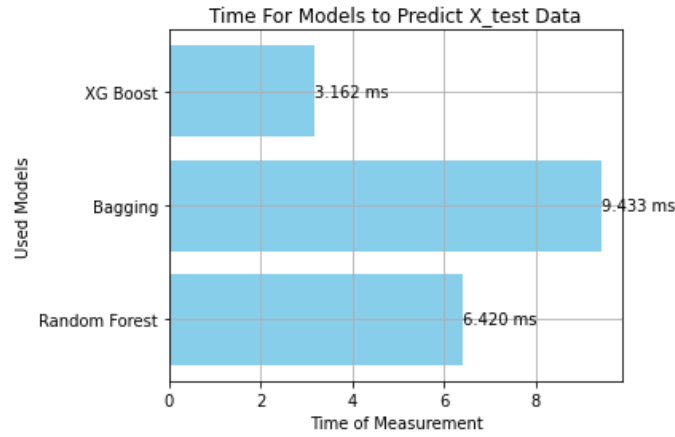
XG Boost modeli, label encoding değeri 0 olan ürünü %42 doğru tahmin ederken label encoding değeri 1 olan ürünü %73, label encoding değeri 2 olan ürünü %57, label encoding değeri 3 olan ürünü %76, label encoding değeri 4 olan ürünü %51 ve label encoding değeri 5 olan ürünü de %66 doğru tahmin etmiştir. Şekil 13'te XG Boost algoritmasına ait karmaşıklık matrisleri verilmiştir.



Şekil 13. XG Boost algoritması karmaşıklık matrisi ısı haritası grafiği (XG Boost algorithm complexity matrix heat map chart)

3.1.3. Model performanslarının zamansal karşılaştırılması (Temporal comparison of model performances)

Kullanılan modellerin, 165 adet veriden oluşan test veri setini ne kadar sürede tahmin ettikleri Python'da bulunan time kütüphanesi ile belirlenmiştir. Time kütüphanesinde bulunan time metodu ile modellerin tahmin süreleri saniye cinsinden hesaplanmış ve milisaniyeye çevrilmiştir. Bu hesaplamalar sonucunda XG Boost algoritması 3.162ms, Random Forest algoritması 6.420ms ve Bagging algoritması 9.433ms sürelerinde tahminde bulunmuşlardır. En hızlı model XG Boost olmasına karşın veriler üzerinde genelleme yeteneği en düşük olan modeldir. Hesaplama sonuçları Şekil 14'te verilmiştir.



Şekil 14. Modellerin tahminde bulunma süreleri (Prediction times of models)

3.1.4. Modelin gerçek veriler ile test sonucu (Test result of the model with real data)

Yapılan denemeler sonucunda tasarlanan programda toprak verileri girildikten sonra doğruluk olarak en uygun bulunan Bagging algoritmasına göre sonuç vermektedir. Örnek olarak girilen gerçek toprak verilerinin ilçe, köy, tarım şekli, potasyum, fosfor, kireç, organik madde, toplam tuz ve pH parametrelerine göre öneri sonucu Tablo 2’de verilmektedir.

Tablo 2. Giriş parametrelerine göre sonuç değeri (Result value according to input parameters)

Giriş parametreleri	Girilen Değerler
İlçe	Senirkent
Köy	Büyükkabaca
Tarım şekli	Sulu
Potasyum	210.8925
Fosfor	2.1870
Kireç	16.8866
Organik madde	1.7328
Toplam tuz	0.0085
pH	7.63
Satürasyon	63”
Çıkış parametresi	Ürün
Tahmin Sonucu	Elma

Tabloda verilen değerler gerçek toprak verisidir ve mevcut halde elma yetiştirilmektedir. Tahminleme sonucunda toprağın elmaya yakın olduğu Bagging algoritması ile önerilmiştir.

4. Sonuç ve Öneriler (Conclusion and Recommendations)

Yapay zekâ yöntemleri ile Isparta şehrinde toprak değerlerine göre en çok yetiştirilen 6 ürüne ait verilerin kullanılması ile bir makine öğrenimi modeli geliştirilmiş ve tasarlanan arayüzden girilen toprak değerlerine göre o toprağın yakın olduğu bitki önerme sistemi gerçekleştirilmiştir. Çalışmada giriş parametreleri olarak “ilçe, köy, tarım şekli, potasyum, fosfor, kireç, organik madde, toplam tuz ve pH” ve çıkış parametresi “ürün” olarak belirlenmiştir. Ürün tahmini için Isparta İl Tarım ve Orman Müdürlüğü ve Isparta İl Özel İdare’si tarafından sağlanan 7500+ gerçek toprak verisi uygun şekilde Excel tablosu haline getirilmiş ve incelemeler yapılmıştır. Kategorik ifadeler label encoder yöntemi ile sayısal hale getirilmiştir. Üç farklı algoritma ile tahminlenen test veri setinde accuracy, F1_score, mean absolute error, mean squared error, r^2 ve confusion matrix olarak en iyi sonucu Bagging algoritması vermiştir. Zamansal olarak da Random Forest ve Bagging

algoritmalarına göre XG Boost algoritması daha hızlı tahminleme yapmıştır.

Çalışmanın sonuçları değerlendirildiğinde giriş parametrelerinin içinde çiftçinin yetiştirmeyi tercih ettiği ürünler de bulunmaktadır. Mevcut yetişen ürünlerden yola çıkılarak yetiştirilmesi verimli olan ürünlerin tespiti sağlanmıştır. Model eğitilirken çiftçinin yetiştirmeyi tercih ettiği ürünler kullanılmış olduğu için gerçek veri sayısının değişimine bağlı olarak uygun bulunacak üründe de değişiklik olabilecektir. Bu da veri sayısının artmasına bağlı olarak daha doğru tahminleme sağlayacaktır.

Gerçek çiftçiler tarafından yapılan analizler sonucu oluşturulan veri setinin mevcut dağılımına bakılarak da görüleceği üzere korelasyon negatif çıkmaktadır. Buna bağlı olarak Bagging algoritması ile elde edilen %63'lük değer yapay zekâ tahminlemesi için düşük bir değer gibi görünse de aslında elde edilen sonucun, toprağın yatkinlik değerini verdiği unutulmamalıdır. Başka bir deyişle eğer yapay zekadan alınan değer %99-98 gibi değerler olsaydı, bu toprakta sistemin bize önerdiği üründen başka bir ürün yetiştirilemez demek olacaktır. Örneğin kullanılan algoritma ile verilen değerlere göre %98 elma yetişir önermesi ile, mevcut tarım alanında elmadan başka ürün yetiştirmek mümkün değildir sonucu elde edilecekti. Öte yandan tahminlenen sonuçla sistem; "büyük olasılıkla elma yetiştirebilirsin- daha düşük olasılıkla kiraz da yetiştirebilirsin" önermesini vermektedir. Ayrıca elde edilen sonuç ne kadar yüksek olursa bu durum toprağın anaç yapısına uygun olmayan bir yanılığa neden olacaktır. Bu sonuç doğal olarak kullanıcıyı yanıltmış olacak ve çiftçiyi tekdüze bir seçime zorlayacaktır. Böylece toprağın aslında aynı anda birden fazla ürüne yatkin olduğunu ama üretimi tercih edilen ürüne göre elde edilecek verimin değişiklik gösterdiği sonucuna da varılmaktadır. Bunlara ek olarak toprağın özelliklerinin yanında iklim değerlerinin önemi de büyüktür. Kullanılan toprak verileri ilçe, köy gibi parametreler ile dolaylı yoldan iklim parametreleri barındırsa da bölgenin gerçek iklim özellikleri ile bu tahminler daha doğru sonuçlar ortaya çıkarabilecektir.

Teşekkür (Acknowledgment)

Bu çalışmada kullanılan toprak analiz verileri Isparta İl Tarım ve Orman Müdürlüğü ve Isparta İl Özel İdare'sinden sağlanmıştır. Destekleri için bu kurumlara teşekkür ederiz.

Çıkar Çatışması Beyanı (Conflict of Interest Statement)

Yazarlar tarafından herhangi bir çıkar çatışması bildirilmemiştir.

Kaynaklar (References)

- [1] H. C. Bayrakçı, R. S. Çiçekdemir and M. Özkahraman, "Tarım Arazilerinde Harcanan Su Miktarını Yapay Zekâ Teknikleri Kullanarak Belirlenmesi," *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, vol. 9, no. 6, pp. 237-250, Ara. 2021, doi:10.29130/dubited.1015690
- [2] Türk Dil Kurumu, sozluk.gov.tr [Online] Available: <https://sozluk.gov.tr>. [Erişim tarihi: 15 Aralık 2023].
- [3] Z. Doğan, S. Arslan and A. Berkman, "Türkiye'de Tarım Sektörünün İktisadi Gelişimi ve Sorunları: Tarihsel Bir Bakış", *Niğde Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 8, no. 1, pp. 29-41, Ocak 2015, ISSN: 2148-5801, e-ISSN: 1308-4216.
- [4] M. Direk, *Tarım Tarihi ve Deontoloji*, Eğitim Akademi Yayınevi, 2012.
- [5] İ. Terzi, M. M. Özgüven, Z. Altaş and T. Uygun, "Tarımda Yapay Zekâ Kullanımı," *International Erciyes Agriculture, Animal Food Sciences Conference, 24-27 April 2019, Erciyes University, Kayseri, Turkey*, 2019, pp. 245-255.
- [6] H. Ekdahl, "Agricultural Technology and the Total System," *Journal of Agricultural Engineering Research*, vol. 76, no. 3, pp. 249-250, 2000. doi:10.1006/jaer.2000.0575
- [7] H. H. Tıraş, "Sürdürülebilir Kalkınma ve Çevre: Teorik Bir İnceleme," *Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, vol. 2, no. 2, pp. 57-73, Haz. 2012.
- [8] K. Temizel and Y. Koç, "Coğrafi Bilgi Sisteminin Hassas Tarımda Uygulanmasının Yararları: Gübreleme Örneği," *Anadolu Tarım Bilimleri Dergisi*, vol. 30, no. 2, pp. 130-135, Jul. 2015, doi:10.7161/anajas.2015.30.2.130-135
- [9] İ. Boz, A. Şahin, M. Paksoy, F. H. Giray and M. Direk, "Çevre Amaçlı Tarımsal Arazilerin Korunması Programının (ÇATAK) Yayılması ve Benimsenmesi," *TÜBİTAK Projesi Kesin Sonuç Raporu*, Proje No: 110O747, 2013. Available: <https://search.trdizin.gov.tr/tr/yayin/detay/613804>. [Erişim tarihi: 15 Aralık 2023].
- [10] E. Kılavuz and İ. Erdem, "Dünyada Tarım 4.0 Uygulamaları ve Türk Tarımının Dönüşümü," *Social Sciences*, vol. 14, no. 4, pp. 133-157, Oct. 2019. doi:10.12739/NWSA.2019.14

- [11] M. F. Baran, A. Çelik ve K. Bellitürk, *Türkiye’de Sürdürülebilir Tarım Uygulamaları: Zorluklar ve Potansiyeller*. Ankara: İksad Publishing House, 2021.
- [12] B. Güzel ve E. Okatan, *Tarım ve Yapay Zekâ. Yapay Zekânın Değiştirdiği Dinamikler*. İstanbul: Eğitim Yayınevi, 2022, pp. 199-224.
- [13] B. Çakmak, M. Yıldırım ve T. Aküzüm, “Türkiye’de Tarımsal Sulama Yönetimi, Sorunlar ve Çözüm Önerileri,” *TMMOB 2. Su Politikaları Kongresi, Ankara, Türkiye, 20-22 Mart 2008*, pp. 215-224.
- [14] H. Yılmaz, V. Demircan and M. Gül, “Üreticilerin Kimyasal Gübre Kullanımında Bilgi Kaynaklarının Belirlenmesi ve Tarımsal Yayım Açısından Değerlendirilmesi,” *Süleyman Demirel Üniversitesi Ziraat Fakültesi Dergisi*, vol. 4, no.1, pp. 31-44, 2009, ISSN: 1304-9984.
- [15] S. Ren, X. Cao, Y. Wei and J. Sun, “Global Refinement of Random Forest,” *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015*, pp. 723-730, doi: 10.1109/CVPR.2015.7298672
- [16] E. Akdağlı, “Makine Öğrenmesinde Random Forest Algoritması,” *medium.com*, 4 Mart 2021. [Çevrimiçi]. Mevcut: <https://ece-akdagli.medium.com/makine-%C3%B6%C4%9Frenmesinde-random-forest-algoritmas%C4%B1-a79b044bbb31>. [Erişim: 14 Aralık 2023].
- [17] M. Hao, S. Hejiang, L. Junjie and W. Shen, “Developing Window Behavior Models for Residential Buildings Using Xgboost Algorithm,” *Energy and Buildings*, vol. 205(2019), pp. 1-8, 2019. doi:10.1016/j.enbuild.2019.109564
- [18] A. C. Kelle and H. Yüce, “MQTT Trafığında DoS Saldırılarının Makine Öğrenmesi ile Sınıflandırılması ve Modelin SHAP ile Yorumlanması,” *Journal of Materials and Mechatronics*, vol. 3, no. 1, pp. 50–62, 2022, doi:10.55546/jmm.995091
- [19] G. Ngo, R. Beard and R. Chandra, “Evolutionary Bagging for Ensemble Learning,” *Neurocomputing*, vol. 510, pp. 1-14, 2022. doi:10.1016/j.neucom.2022.08.055
- [20] Ş. Ay, “Ensemble Learning-Bagging ve Boosting,” *medium.com*, 16 Aralık 2019. [Çevrimiçi]. Mevcut: <https://medium.com/deep-learning-turkiye/ensemble-learning-bagging-ve-boosting-50643428b22b>. [Erişim: 15 Aralık 2023].

* This paper was presented at the 5th International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAAME 2023) and the abstract was published as an e-book.

This is an open access article under the CC-BY license

