

# GAZİ

## JOURNAL OF ENGINEERING SCIENCES

### Hyperparameter Optimization in Polynomial Kernel SVM Algorithms and its Application

Hamit Armağan<sup>a</sup>

Submitted: 27.11.2023 Revised: 21.12.2023 Accepted: 22.12.2023 doi:10.30855/gmbd.0705S21

#### ABSTRACT

**Keywords:** machine learning, svm, bayesian optimization

<sup>a\*</sup> Süleyman Demirel Üniversitesi,  
Rektörlük, Enformatik Bölümü, 32200  
Isparta/TÜRKİYE  
Orcid: 0000-0002-8948-1546  
e mail: hamitarmagan@sdu.edu.tr

\*Corresponding author:  
hamitarmagan@sdu.edu.tr

In this study, regression analyses and parameter optimizations of support vector machines (SVMs) with different polynomial kernels were performed. Using linear-svm, quadratic-svm and cubic-svm regression algorithms, the performance of each model is examined through an optimization process on the Box-Constraint, Kernel Scale, Epsilon and Standardize parameters. Setting these parameters correctly is critical for the error and approximation metrics of the model. The optimization process was performed with Matlab Regression Learner using the Bayesian Optimization algorithm. Hyperparameter optimized polynomial-based regression models can predict nutrient deficiency in plants with high accuracy. In particular, accurate determination of calcium content in leaves is important to increase fertilization success during the fruit development period in apple trees. In the application part of our study, data sets obtained by digitizing leaf surfaces and data sets obtained from chemical laboratory analysis were modeled. This study using SVM algorithms provides a more efficient model than chemical methods in terms of cost and time.

### Polinom Çekirdekli SVM Algoritmalarında Hiper-Parametre Optimizasyonu ve Uygulaması

#### ÖZ

Bu çalışmada, destek vektör makinelerinin (SVM) farklı polinom çekirdeklerini içeren regresyon analizleri ve parametre optimizasyonları ele alınmıştır. Svm-linear, svm-quadratic ve svm-cubic regresyon algoritmaları kullanılarak her bir modelin performansı, box-constraint, kernel scale, epsilon ve standardize parametreleri üzerinde gerçekleştirilen optimizasyon süreciyle incelenmiştir. Bu parametrelerin doğru bir şekilde ayarlanması, modelin hata ve yaklaşım metrikleri için kritik önem taşımaktadır. Optimizasyon süreci, Bayesian Optimization algoritması kullanılarak matlab regression learner ile gerçekleştirilmiştir. Hiperparametre optimizasyonu yapılmış polinom tabanlı regresyon modelleri, bitkilerde besin elementlerinin eksikliğini yüksek doğrulukla tahmin edebilmektedirler. Özellikle yapraklardaki kalsiyum miktarının doğru bir şekilde belirlenmesi, elma ağaçlarında meyve gelişim döneminde gübreleme başarısını artırmak açısından önem taşımaktadır. Çalışmamızın uygulama bölümünde yaprak yüzeylerinin sayısallaştırılması ile elde edilen veri setleri ile kimyasal laboratuvar analizlerinden elde edilen veri seti modellenmiştir. SVM algoritmaları kullanılarak yapılan bu çalışma, maliyet ve zaman açısından kimyasal yöntemlere göre daha verimli bir model sunmaktadır.

**Anahtar Kelimeler:**  
makine öğrenmesi,  
svm,  
bayesian optimization

## 1. Giriş (Introduction)

Yapay zeka algoritmalarında veriler arasındaki ilişkileri modellemek için polinom çekirdekli regresyon algoritmaları çok kullanılan yöntemlerdir. Birçok durumda, tek girdi değişkeni kullanan lineer regresyon, günlük hayattaki olaylarla ilişkili modellerde yeterli tahmin gücü sunamayabilir [1]–[3]. Bu durum araştırmacıları, regresyon modeli için girdi olarak aynı anda iki ya da daha fazla değişkenin kullanıldığı lineer olmayan regresyon tekniklerine yönelmiştir [4]–[6]. Bu modellerin yazılım algoritmalarında tek başına ya da optimize edilmeden kullanılmaları regresyon uygulamalarında istenen başarının elde edilememesine neden olmaktadır. Bazı veri kümeleri için regresyon uygulamalarında polinom derecesinin artırılması ya da algoritmada kullanılan parametrelerin optimize edilmesi gerekmektedir. Ancak, polinom derecesindeki artışlar modelin karmaşıklaşmasına ya da aşırı uyum (overfitting) riskinin ortaya çıkmasına neden olmaktadır [5], [7]–[9]. Bu durum yazılım algoritmalarının kullanılabilirliğini azaltmakta ve işlem süresini uzatmaktadır. Parametre optimizasyonu ile parametrik matematiksel modelleri kullanan algoritmaların daha başarılı ve hızlı olması sağlanmaktadır [10]–[12].

Bu çalışmada kullanılan görüntü işleme tabanlı tahmin modelleri svm-linear, svm-quadratic ve svm-cubic regresyon algoritmalarıdır. Bu modeller, özellikle regresyon analizleri için kullanılan destek vektör makinelere farklı polinom çekirdeklerini temsil etmektedir. Her bir modelin performansını optimize etmek amacıyla, box constraint, kernel scale, epsilon ve standardize gibi kritik parametreler üzerinde optimizasyon süreci gerçekleştirilmiştir. Bu parametrelerin doğru bir şekilde ayarlanması, modelin tahmin yeteneklerini ve genel performansını artırmak için kritik öneme sahiptir [1], [2], [6], [13].

Optimizasyon süreci, "Bayesian Optimization" algoritması kullanılarak gerçekleştirilmiştir. Bu algoritma, modelin hiperparametre uzayını tarayarak en iyi kombinasyonu bulmayı amaçlar. Bayesian Optimization, modelin performansını sürekli olarak izler, daha iyi sonuçlar elde edilebilecek yeni hiperparametre kombinasyonlarını deneyerek optimizasyon sürecini günceller [2]. Bu sayede, her bir regresyon modeli için en uygun parametre seti belirlenmiş ve bu modellerin elma ağaçlarında besin elementlerinin değerini tahmin etme yetenekleri maksimum seviyeye çıkartılmıştır [14], [15].

Ticari bahçelerde meyve gelişim döneminde toprak ve ağaçlardan alınan veriler bahçe yönetim etkinliğini artırır. Veriye dayalı olmayan ya da hatalı veriye dayanan uygulamalar sonucunda kullanılan kimyasallar toprağın ve çevrenin gereksiz kirletilmesine, ekonomik olarak maliyetlerin artmasına ve iş gücü kaybına neden olmaktadır [12], [16]. Bilgisayarla görü tabanlı modeller kullanılarak bitkilerin beslenme durumunun doğru tespiti oldukça önemlidir. Bu yöntemlerle bitkilerde besin maddelerinin eksikliği ya da fazlalığı yüksek doğrulukta tahmin edilebilmektedir.

Yapraklardaki kalsiyum değeri, ağaç beslenmesini değerlendirmek için etkili parametrelerden biridir. Elma ağaçlarında meyve gelişim döneminde yapraklardaki kalsiyum miktarının tespiti gübreleme başarısını arttırmak açısından önemlidir. Fakat bu analizlerin yapılması maliyetli ve zaman alıcıdır [17]–[19].

Bu çalışmanın amacı, laboratuvar ortamında kimyasal yöntemlerle ölçülen yaprak kalsiyum içeriğini, görüntü analizi ve polinom çekirdekli svm algoritmalarını kullanarak tahmin etmektir. Bu modeller, numune alma işlemini ve maliyetini azaltabilir ve özellikle numunelerin işlenmesinin zor olduğu durumlarda hassasiyeti artırabilir. Yapraklardaki kalsiyum miktarının doğru bir şekilde belirlenmesi, meyve üretimi ve bahçe yönetimi için önemlidir. Standart laboratuvar analizleri güvenilir olsa da zaman alıcı ve maliyetlidir. Bu amaçla, yaprakların sayısal analizi ve regresyon modelleri ile ağaçların besin element analizlerinin yapılması, çevresel sürdürülebilirliği ve bahçe verimliliğine katkı sağlayacaktır.

## 2. Materyal ve Metot (Material and Method)

### 2.1. Materyal (Material)

Çalışmada kullanılan görüntüler "Göller Yöresi" bölgesinde yer alan ticari bir elma bahçesinden alınmıştır. Yaprak örnekleme, besin elementi içerikleri kararlı hale geldiği dönem olan tam çiçeklenme tarihinden 111 gün sonra konu uzmanları tarafından yapılmıştır. Yaprak örnekleme için güneş gören dalların orta kısmında yer alan ~80 yaprak sapsarı ile birlikte koparılmıştır. Her bir ağaç farklı bir ürün yükünü temsil edecek şekilde

toplamda 20 ağaç belirlenerek etiketlenmiş ve yaprak analiz örnekleri/verileri aynı ağaçlardan toplanmıştır. Çalışmada bitki besin elementi analizleri yapılmış olup, bu analizler Burdur Mehmet Akif Ersoy Üniversitesi Bilimsel ve Teknoloji Uygulama ve Araştırma Merkezinde laboratuvarlarında gerçekleştirilmiştir.

## 2.2. Veri toplama (Data collection)

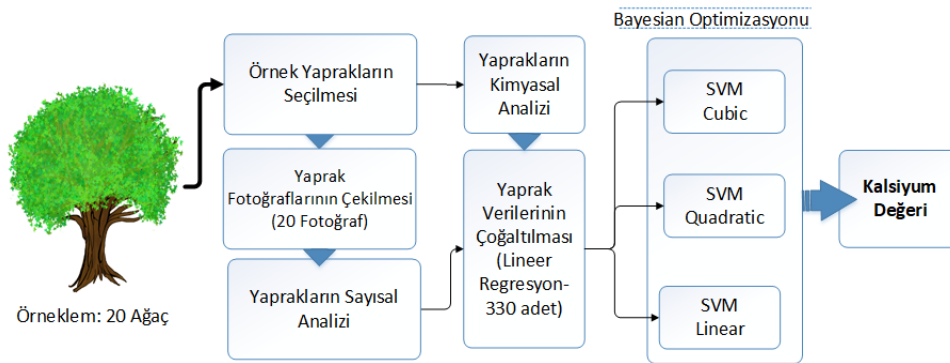
Regresyon algoritmalarında kullanılmak üzere çekilen resimler 4032x3024 boyutunda olup, 24 bit derinliğinde, 72 dpi çözünürlükte ve sRGB renk uzayı ile renkleri temsil edilmektedir. Görüntü işleme çalışmalarında kolaylık sağlaması için resimler ön işleme tabi tutulmuştur. Bu ön işlemden ağaçta yer alan yapraklar diğer nesnelere ve gölgelerinden ayırma işlemi için görüntü bölütleme işlemi uygulanmıştır. Görüntü bölütleme, görüntüdeki piksellerin özelliklerine bağlı olarak bir görüntüyü birden çok parçaya veya bölgeye ayırmak için kullanılan bir yöntemdir. Görüntü bölütleme ile ön planı arka plandan ayırma işlemi, renk ya da şekil benzerliklerini kümeleme en çok kullanılan uygulamalarıdır. Bu çalışmada görüntü bölütleme işlemi için matlab image segmenter uygulamasında yer alan grafik kesim graph cut yöntemi kullanılmıştır (Şekil 1). Grafik kesim, bir görüntüyü ön plan ve arka plan öğelerine ayırmak için kullanılan bölümlenme tekniğidir. Grafik kesim yöntemi hızlı bölümlenme için grafik teorisini kullanmaktadır. Tanımlanan ön plan ve arka plan bilgilerine göre pikseller arası bağlantı ağırlığı belirlenir ve algoritma pikseller arasındaki bağlantı ağırlıklarına göre görüntüdeki nesnelere bölünmesini sağlar.

## 2.3. Veri çoğaltma (Synthetic data generation)

Veri çoğaltma işlemleri iki aşamalı olarak uygulanmıştır. İlk olarak çekilen fotoğraflar arka plan gürültüleri temizlendikten sonra kırmızı(r), yeşil(g) ve mavi(b) renk kanallarının değeri sıfırdan büyük olan pikseller eşit eleman sayılarından oluşacak şekilde alt kümelere ayrılmıştır. Alt küme sayısı 330 olarak seçilmiştir.

$$y = f(r, g, b) \quad (1)$$

Bu işlem ile  $f(r, g, b)$  fonksiyonunun tanım kümesindeki eleman sayısı lineer regresyon yöntemi kullanılarak 20'den, 330'a çıkarılmıştır. Her ağaçtan 4 yaprak alınmıştır. Çekilen her fotoğrafta aynı ağaçtan alınan dört yaprak kullanılmıştır. Veri toplama, çoğaltma ve matematiksel modellerin kullanımı Şekil 1'de detaylı gösterilmiştir.



Şekil 1. Polinom tabanlı SVM algoritmalarında veri toplama, çoğaltma ve tahminleme süreçleri (Data collection, synthetic data generation and prediction processes in polynomial kernel SVM algorithms)

## 2.4. Matematiksel model ve parametreler (Mathematical model and parameters)

SVM modeli, regresyon görevinde kullanılan bir makine öğrenimi algoritmasıdır. Bu model, veri setindeki örnekler arasında en iyi uyum sağlayan bir hiperdüzlemi bulmaya çalışarak hedef değişkenin tahminini gerçekleştirir. Destek vektör makinelerinin (SVM) regresyonu için kullanılan hiperparametreler, modelin hata ve yaklaşım değerlerini etkileyen önemli araçlardır [19], [20].

C(Cost), hata noktalarının ne kadar ihmal edilebileceğini kontrol eden bir parametredir. Büyük C(cost) değerleri, modelleme hatasına daha fazla ceza verir ve modelin eğitim veri setine daha fazla uymasını sağlar.

Ancak, aşırı uyuma yol açabilir. Küçük C değerleri genelleme yeteneğini artırabilir ancak eğitim veri setindeki hatalara hassas değildir [21]-[23].

Epsilon ( $\epsilon$ ), regresyon hatasının kabul edilebilir bir sınırdan ne kadar sapma göstermesine izin verileceğini kontrol eder. Epsilon değeri büyüdükçe, model regresyon hatasına daha fazla tolerans göstermektedir [21]-[23].

Kernel, SVM'de veri noktalarını bir özellik uzayına dönüştürmek için kullanılan bir fonksiyondur. Regresyon kernel fonksiyonlarının genel fonksiyonu denklem 1'de verilmiştir.

$$K(x, y) = (\gamma \cdot x \cdot y + r)^d \quad (2)$$

Denklem 2'de  $\gamma$  kernel katsayısı, d polinom derecesi ve r sapma değeridir. Hiperparametrelerin en uygun değerlerini bulmak için çapraz doğrulama yöntemi ve farklı hiperparametre kombinasyonlarını denemek içinde Bayesian Optimizasyon algoritması kullanılmıştır.

Veri setinde toplam 330 gözlem bulunmakta olup, tahmin yapmak için 3 adet tahminci değişken (R,G,B) kullanılmıştır. Hedef değişken "Ca" olarak belirlenmiştir. Modelin performansı, 5 katlı çapraz doğrulama yöntemi kullanılarak değerlendirilmiştir.

Deneyimdeki veri seti üzerinde gerçekleştirilen 5 katlı çapraz doğrulama, modelin genelleme yeteneğini değerlendirmek ve performansını daha güvenilir bir şekilde ölçmek için kullanılmıştır. Her bir iterasyonda veri seti rastgele beşe bölünmüş, model dört katmanda eğitilmiş ve beşinci katmanda test edilmiştir. Bu işlem beş kez tekrarlanarak genel performans değerlendirmesi sağlanmıştır [18], [19], [23].

Bu otomatikleştirilmiş regresyon analizi deneyimi, veri bilimi ve makine öğrenimi alanındaki uygulamalara hızlı ve etkili bir çözüm sunmaktadır. SVM modeli, tahmin değişkeni "Ca" üzerindeki ilişkileri anlamak ve model performansını değerlendirmek için kullanılmıştır. Elde edilen sonuçlar, modelin belirli bir veri setinde başarılı bir şekilde genelleme yapabilme yeteneğini yansıtmaktadır.

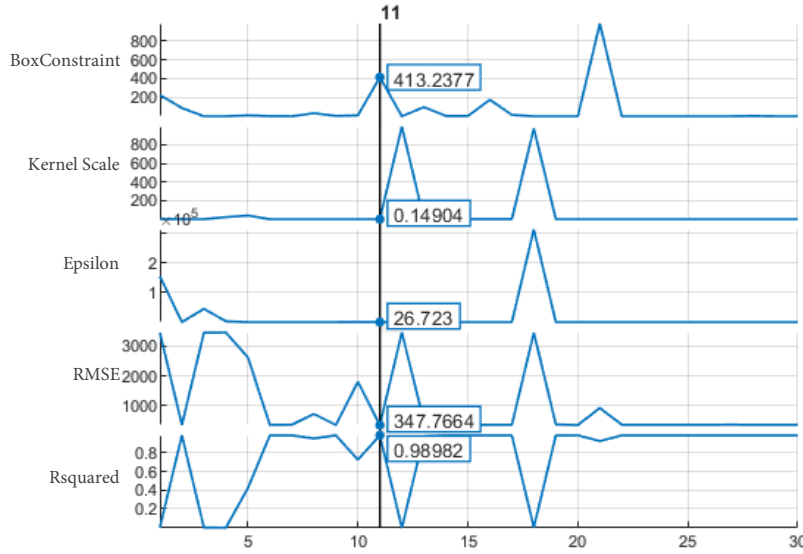
Hiperparametreler; box constraint (C), modelin karmaşıklığını kontrol eden bir hiperparametredir. [0.001, 1000] aralığında gerçel sayılar içerir ve logaritmik bir dönüşümle işlenmiştir. Kernel scale, SVM'nin çekirdek fonksiyonlarının ölçeğini belirleyen bir hiperparametredir ve [0.001, 1000] aralığında gerçel sayılar içerir ve logaritmik bir dönüşümle işlenmiştir. Epsilon, SVM regresyonunun hatası olan epsilon değerini belirler. [3.2158, 321575.7631] aralığındaki gerçel sayılar içerir ve logaritmik bir dönüşümle işlenmiştir. Standardize(standartlaştırma), veri setini standartlaştırma durumunu belirleyen bir hiperparametredir "true" veya "false" değerlerini alabilir.

Bu hiperparametreler, modelin karmaşıklığını, hatasını ve standartlaştırma durumunu kontrol eden önemli araçlardır. Bayesian Optimizasyon, bu hiperparametre setini optimize ederek modelin daha iyi bir performans göstermesini sağlar.

Bayesian Optimization sürecindeki işlem basamakaları; hiperparametre uzayı, İlk olarak, modelin optimize edilecek hiperparametrelerinin bir uzayı belirlenir. Bu uzay, box constraint, kernel scale, epsilon, standardize gibi hiperparametreleri içerir. Hiperparametrelere başlangıç değeri ataması yapılır. Amaç fonksiyonu belirlenir. Bu fonksiyon, modelin performansını ölçen bir metrik üzerinden tanımlanır. Belirlenen cost(C) dahilinde hiperparametre uzayında iteratif olarak değerlendirme yapılır. Her adımda, modelin performansı ölçülür ve bu değere bağlı olarak bir sonraki hiperparametre seti belirlenir. Belirlenen cost(C) içinde en iyi performansı sağlayan hiperparametre seti bulunur. Bu set, Bayesian Optimization sürecinin sonucudur. İteratif bir süreç olarak, her bir döngüde hiperparametre seçimi, model eğitimi, performans değerlendirmesi ve güncelleme adımları tekrarlanarak en iyi hiperparametre setinin bulunması hedeflenir.

### 3. Sonuçlar ve Tartışma (Results and Discussion)

#### 3.1. SVM-Linear algoritması (SVM-Linear algorithm)



Şekil 2. Bayesian Optimization yöntemi kullanılarak SVM-Linear algoritmasında elde edilen optimize edilmiş parametrelerin hata ve yaklaşım performansları

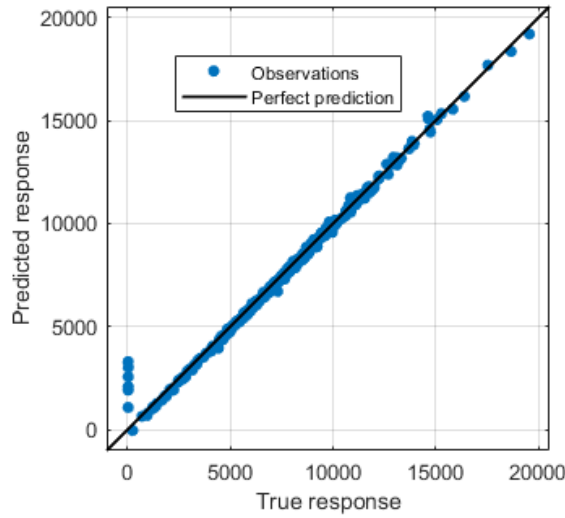
(Error and approximation performances of optimized parameters obtained in the SVM-Linear algorithm using the Bayesian Optimization method)

Destek vektör makinesi (SVM) modelinin hiperparametre optimizasyonu için kullanılan Bayesian Optimization yöntemi, belirlenen hiperparametre setinin doğrulama performansını ayrıntılı olarak şekil 2'de verilmiştir. Aynı zamanda, Bayesian Optimization yöntemiyle elde edilen hiperparametre kombinasyonları arasında en yüksek başarıya sahip olan ilk 10 sonuç tablo 1'de paylaşılmıştır. Bu doğrultuda belirlenen hiperparametre seti ile geliştirilen Destek Vektör Makinesi (SVM) modeli, yüksek R kare değeriyle veriyi etkili bir şekilde açıklamakta ve düşük Root Mean Square Error (RMSE) değeriyle tahminlerin gerçek değerlere yakın olduğunu göstermektedir. Bayesian Optimization stratejisi kullanılarak belirlenen hiperparametre seti ve oluşturulan SVM modelinin; doğrulama verisi üzerinde başarılı bir performans sergilediğini göstermektedir.

Tablo 1. Bayesian Optimization yöntemi kullanılarak SVM-Linear algoritmasında elde edilen optimize edilmiş parametreler ve doğrulama performanslarının kombinasyonları ( Sıralama RMSE değerine göre yapılmış olup ilk 10 değeri göstermektedir. Toplam kombinasyon sayısı 30 dur.)

(Combinations of optimized parameters and verification performances obtained in the SVM-Linear algorithm using the Bayesian Optimization method (The ranking is made according to the RMSE value and shows the first 10 values. The total number of combinations is 30.))

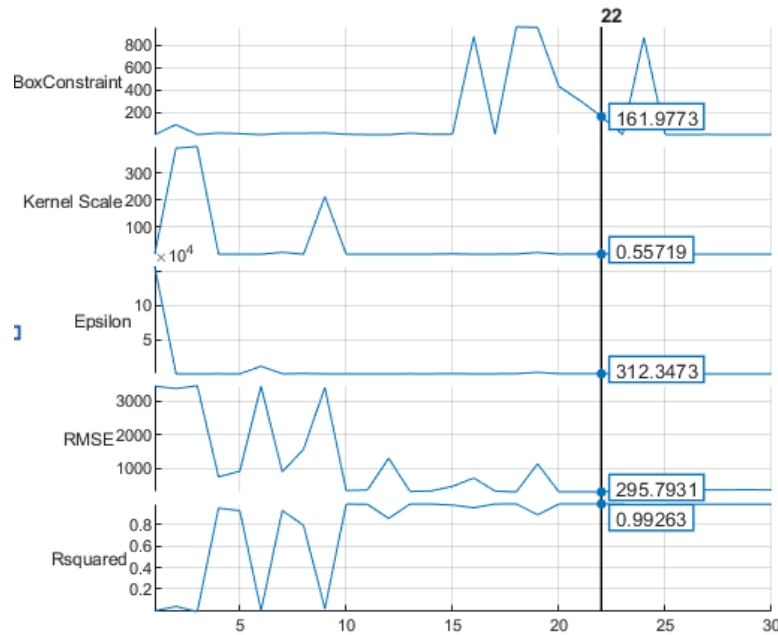
BoxConstraint	Kernel Scale	Epsilon	Validation RMSE	Validation Rsquared
413.2376902	0.149042976	26.72295986	347.7663911	0.989816633
0.026945007	0.028940967	28.96798805	348.3532578	0.989782235
3.35792329	0.129978567	20.90234852	350.402532	0.989661664
3.682985661	0.068124603	19.84131668	350.9247292	0.989630827
87.31014359	0.270066044	17.74144123	351.7494962	0.989582029
1.427989577	0.167288076	11.59887758	353.8129505	0.989459441
0.001101378	0.004131366	8.959736054	354.7195096	0.989405357
3.140283237	0.147229454	8.530972267	354.8003963	0.989400525
173.8278945	0.285662457	7.559932223	355.4594489	0.98936111
0.034528455	0.015737249	5.650026688	355.8737029	0.989336299



Şekil 3. SVM- Linear algoritmasında gerçek değer ve tahmin yanıtının karşılaştırma performansları  
(Comparison performances of actual value and predicted responses in the SVM- Linear algorithm)

Şekil 3'te diyagonal çizgi ideal durumu temsil eder; yani, gerçek ve tahmin edilen değerler arasındaki uyumu gösterir. Görsel olarak bu çizgiye yakın bir dağılım, modelin gerçek değerlere yakın ve tutarlı tahminlerde bulunduğunu gösterir. Bu durum, belirlenen hiperparametre seti ve Bayesian Optimization'u kullanılarak oluşturulan SVM modelinin veriyi etkili bir şekilde açıkladığını ve doğrulama verisi üzerinde yüksek performans sergilediğini göstermektedir.

### 3.2. SVM-Quadratic algoritması (SVM-Quadratic algorithm)



Şekil 4. Bayesian Optimization yöntemi kullanılarak SVM- Quadratic algoritmasında elde edilen optimize edilmiş parametrelerin hata ve yaklaşım performansları  
(Error and approximation performances of optimized parameters obtained in the SVM-Quadratic algorithm using the Bayesian Optimization method)

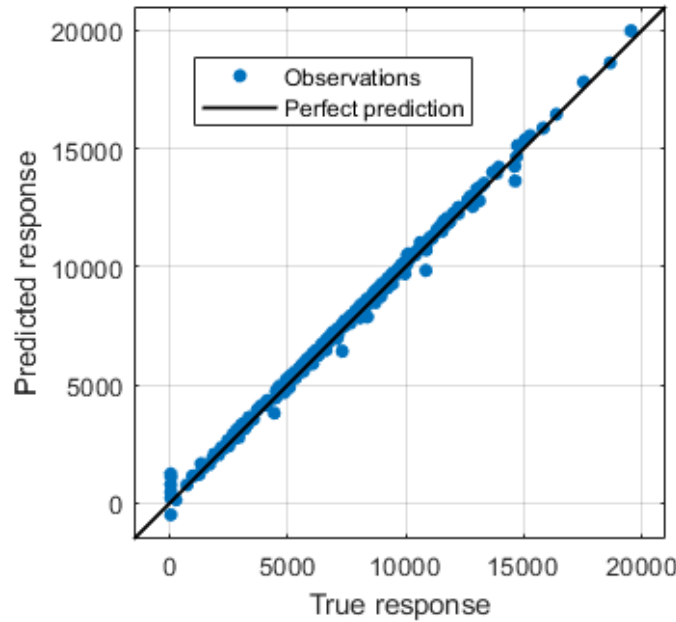
Destek vektör makinesi (SVM) modelinin hiperparametre optimizasyonu için Bayesian Optimization yöntemi kullanılarak belirlenen hiperparametre setinin doğrulama performansı şekil 4'de verilmiştir. Ayrıca Bayesian Optimization yöntemi kullanılarak elde edilen hiperparametre kombinasyonlarından en yüksek başarıya sahip ilk 10 sonuç tablo 2'de paylaşılmıştır.

Tablo 2. Bayesian Optimization yöntemi kullanılarak SVM-Quadratic algoritmasında elde edilen optimize edilmiş parametre ve doğrulama performans kombinasyonları (sıralama RMSE değerine göre yapılmış olup ilk 10 değeri göstermektedir. Toplam kombinasyon sayısı 30 dur.)

(Combinations of optimized parameters and verification performances obtained in the SVM- Quadratic algorithm using the Bayesian Optimization method (The ranking is made according to the RMSE value and shows the first 10 values. The total number of combinations is 30.))

BoxConstraint	Kernel Scale	Epsilon	Validation RMSE	Validation Rsquared
161.9773328	0.557191839	312.3473457	295.7931324	0.992632972
967.9262171	0.875948995	346.9786153	297.271893	0.992559127
433.4541301	0.311059872	240.5526151	299.6919974	0.992437481
304.829562	0.771002448	207.0737476	302.4558597	0.99229735
13.13843562	0.325438467	150.3384645	312.7825197	0.991762392
3.380708157	0.270484319	114.9687094	322.2166498	0.991257973
2.994487851	0.242315929	98.55338541	326.5159236	0.991023131
0.429380191	0.050528408	65.25387426	335.7532089	0.990508026
3.397631434	0.085163718	51.95381339	339.7573508	0.990280277
0.265001018	0.001016695	25.42854692	348.7772761	0.989757345

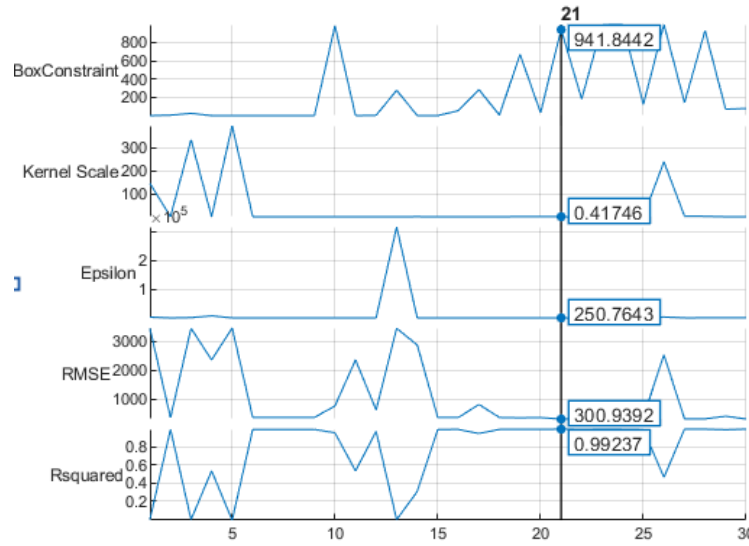
Belirlenen hiperparametre seti ile elde edilen SVM modeli, yüksek R kare değeri ile veriyi yüksek oranda açıklamakta ve düşük RMSE değeri ile tahminlerin gerçek değerlere yakın olduğunu göstermektedir. Bayesian Optimization stratejisi kullanılarak elde edilen hiperparametre seti ve SVM modelinin, doğrulama verisi üzerinde yüksek performans sergilediğini göstermektedir.



Şekil 5. SVM-Quadratic algoritmasında gerçek değer ve tahmin yanıtlarının karşılaştırma performansları (Comparison performances of actual value and predicted responses in the SVM- Quadratic algorithm)

Şekil 5’de, SVM-Quadratic regresyon modelinin gerçek ve tahmini değerlerini karşılaştırmak için kullanılmıştır. Bu grafik, modelin gerçek verilere yaklaşımını değerlendirmek adına önemli bir araçtır. İyi bir modelde, gerçek ve tahmin değerleri tam olarak örtüşür ve diyagonal çizgi oluşturmaktadır. Grafik üzerindeki her noktanın diyagonal çizgiden uzaklığı, o tahminin hatasını ifade eder. Grafikte de görüleceği gibi gerçek değerler ve tahmin değerleri yüksek oranda örtüşmekte ve diyagonal çizgi oluşturmaktadır.

### 3.3. SVM-Cubic algoritması (SVM-Cubic algorithm)



Şekil 6. Bayesian Optimization yöntemi kullanılarak SVM-Cubic algoritmasında elde edilen optimize edilmiş parametreler ve doğrulama performansları  
(Error and approximation performances of optimized parameters obtained in the SVM-Cubic algorithm using the Bayesian Optimization method)

Bu çalışmada, destek vektör makinesi (SVM) modeli için Bayesian Optimization yöntemi kullanılarak belirlenen hiperparametre setinin doğrulama performansı incelenmiştir. Bu sonuçlara göre, belirlenen hiperparametre seti ile elde edilen SVM modelinin yüksek performans sergilediğini göstermektedir. Düşük RMSE değeri, modelin tahminlerinin gerçek değerlere yakın olduğunu, yüksek R kare değeri ise modelin veriyi yüksek oranda açıkladığını göstermektedir.

Bayesian Optimization'ın kullanılması, belirlenen hiperparametre setini optimize etmede etkili olmuştur. Bu sonuçlar, SVM modelinin belirli bir problem üzerinde başarılı bir şekilde uyarlandığını ve hiperparametre optimizasyonunun bu başarıyı artırdığını göstermektedir. Özellikle, BoxConstraint ve Kernel Scale gibi hiperparametrelerin model performansındaki etkisi önemlidir. Standardizasyonun false olması, veri setinin standartlaştırılmadığını göstermektedir. Ayrıca Bayesian Optimization yöntemi kullanılarak elde edilen hiperparametre kombinasyonlarından en yüksek başarıya sahip ilk 10 sonuç tablo 3'de ve doğrulama performansı da şekil 7'de verilmiştir.

Tablo 3. Bayesian Optimization yöntemi kullanılarak SVM-Cubic algoritmasında elde edilen optimize edilmiş hiperparametreler ve doğrulama performans kombinasyonları ( sıralama RMSE değerine göre yapılmış olup ilk 10 değeri göstermektedir. Toplam kombinasyon sayısı 30 dur.)

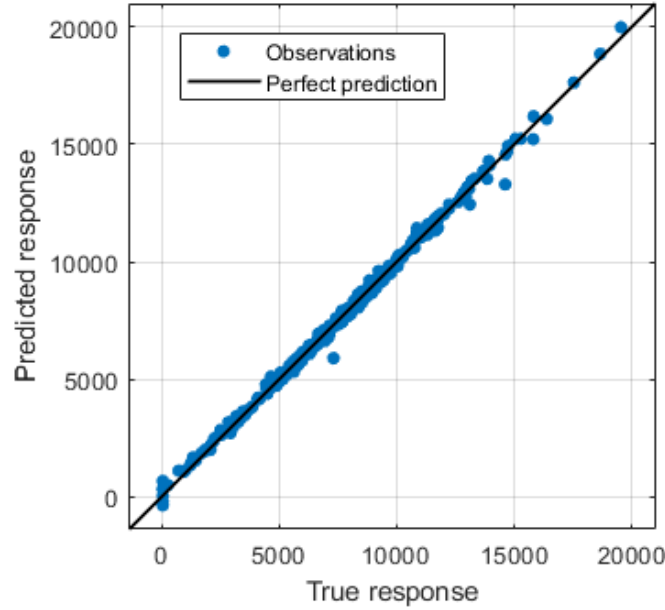
(Combinations of optimized parameters and verification performances obtained in the SVM-Cubic algorithm using the Bayesian Optimization method  
(The ranking is made according to the RMSE value and shows the first 10 values. The total number of combinations is 30.))

BoxConstraint	Kernel Scale	Epsilon	Validation RMSE	Validation Rsquared
941.8442131	0.417461642	250.7642617	300.9392457	0.992374403
78.10845401	0.289259491	326.2277033	302.4268688	0.992298826
142.2002401	2.60377804	255.1919797	305.3886486	0.992147247
929.7379756	1.873427275	385.3543677	308.1705874	0.992003526
995.9629942	8.300488299	321.5053366	312.7371392	0.991764782
997.0468113	1.672638739	131.1677841	319.5328459	0.991402995
181.6804707	0.234791432	92.64252791	329.4120231	0.99086318
670.7817518	0.776443467	35.48066837	345.8667607	0.98992758
5.092075509	0.100515873	18.30578891	351.8103139	0.989578426
31.77291005	0.506563232	17.66780319	351.822748	0.989577689



Şekil 7, SVM-Cubic regresyon modelinin hata ve yaklaşım değerini gösteren bir grafikdir. Grafikte, modelin gerçek değerlere olan yaklaşımını ve tahmin yeteneği görselleştirilmiştir. İdeal bir modelde, gerçek ve tahmin edilen değerler birbirine çok yaklaşıp ve bu durum diyagonal bir çizgi oluşturur. Grafikte her noktanın diyagonal çizgiden uzaklığı, o noktanın tahmin hatasını temsil eder.

Gözlemlendiği üzere, SVM-Cubic regresyon modeli tarafından gerçekleştirilen tahminler ile gerçek veriler arasında yüksek bir uyum söz konusudur. Grafikteki noktaların genel eğilimi, diyagonal bir çizgi oluşturacak şekilde gerçek ve tahmin değerlerinin yakın olduğunu göstermektedir. Bu durum, modelin verileri etkili bir şekilde öğrendiğini ve genellemeye uygun bir performans sergilediğini işaret etmektedir.



Şekil 7. SVM-Cubic algoritmasında gerçek değer ve tahmin yanıtının karşılaştırma performansları (Comparison performances of actual value and predicted responses in the SVM-Cubic algorithm)

### 3.4. Karşılaştırma ve değerlendirme (Comparison and evaluation)

RMSE değerleri, SVM Quadratic modelinin RMSE değeri (295.7931324), diğer iki modele göre daha düşüktür. Düşük RMSE, modelin daha iyi tahminler yaptığını göstermektedir. Rsquared (R kare) değerleri; SVM Quadratic modelinin R squared değeri (0.992632972), diğer modellere göre daha yüksektir. Bu durum, SVM Quadratic modelinin veriyi daha iyi açıkladığını göstermektedir. Hiperparametre değerleri; SVM Cubic modeli, diğerlerine göre daha yüksek BoxConstraint ve Kernel Scale değerlerine sahiptir. Bu durum, SVM Cubic modelinin daha karmaşık bir yapıya sahip olduğunu göstermektedir. Sonuç olarak; SVM Quadratic modeli, düşük RMSE ve yüksek R squared değerleri ile modelin daha iyi bir performans sergilediğini göstermektedir.

### Çıkar Çatışması Beyanı (Conflict of Interest Statement)

Yazarlar tarafından herhangi bir çıkar çatışması bildirilmemiştir.

### Kaynaklar (References)

- [1] H. Fattahi and N. Babanouri, "Applying Optimized Support Vector Regression Models for Prediction of Tunnel Boring Machine Performance," *Geotech. Geol. Eng.*, vol. 35, no. 5, pp. 2205–2217, Oct. 2017. doi:10.1007/S10706-017-0238-4/TABLES/8.
- [2] V. Strijov and G. W. Weber, "Nonlinear regression model generation using hyperparameter optimization," *Comput. Math. with Appl.*, vol. 60, no. 4, pp. 981–988, Aug. 2010. doi:10.1016/J.CAMWA.2010.03.021.
- [3] P. Yu, M. Y. Low, and W. Zhou, "Design of experiments and regression modelling in food flavour and sensory analysis: a

- review,” *Trends Food Sci. Technol.*, vol. 71, pp. 202–215, Jan. 2018. doi:10.1016/j.tifs.2017.11.013.
- [4] A. Kamilaris and F. X. Prenafeta-Boldú, “Deep learning in agriculture: A survey,” *Comput. Electron. Agric.*, vol. 147, pp. 70–90, Apr. 2018. doi: 10.1016/J.COMPAG.2018.02.016.
- [5] P. Freund, R. J., Wilson, W. J., Sa, *Regression analysis: Statistical Modeling of a response variable (2nd ed)*. California, USA: Elsevier, 2006.
- [6] N. Panigrahi and B. S. Das, “Evaluation of regression algorithms for estimating leaf area index and canopy water content from water stressed rice canopy reflectance,” *Inf. Process. Agric.*, vol. 8, no. 2, pp. 284–298, 2021. doi:10.1016/j.inpa.2020.06.002.
- [7] I. Keramatlou, M. Sharifani, H. Sabouri, M. Alizadeh, and B. Kamkar, “A simple linear model for leaf area estimation in Persian walnut (*Juglansregia L.*),” *Sci. Hortic. (Amsterdam)*, vol. 184, pp. 36–39, Mar. 2015. doi:10.1016/j.scienta.2014.12.017.
- [8] Basak, J. K., Qasim, W., Okyere, F. G., Khan, F., Lee, Y. J., Park, J., and Kim, H. T., “Regression Analysis to Estimate Morphology Parameters of Pepper Plant in a Controlled Greenhouse System,” *J. Biosyst. Eng.*, vol. 44, no. 2, pp. 57–68, Jun. 2019. doi:10.1007/S42853-019-00014-0/FIGURES/11.
- [9] H. Armağan, “Color Based Segmentation with k-Means Clustering Algorithm and Numerical Analysis of the Effect of Color Spaces on Image Quantities,” *El-Cezeri*, vol. 9, no. 4, pp. 1506–1517, Dec. 2022. doi:10.31202/ECJSE.1141148.
- [10] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, “Machine Learning in Agriculture: A Review,” *Sensors*, Vol. 18, no. 8, p. 2674, Aug. 2018. doi: 10.3390/S18082674.
- [11] T. U. Rehman, M. S. Mahmud, Y. K. Chang, J. Jin, and J. Shin, “Current and future applications of statistical machine learning algorithms for agricultural machine vision systems,” *Comput. Electron. Agric.*, vol. 156, pp. 585–605, Jan. 2019. doi:10.1016/J.COMPAG.2018.12.006.
- [12] M. Pathan, N. Patel, H. Yagnik, and M. Shah, “Artificial cognition for applications in smart agriculture: A comprehensive review,” *Artif. Intell. Agric.*, vol. 4, pp. 81–95, Jan. 2020. doi:10.1016/J.AIIA.2020.06.001.
- [13] “K-means clustering based image segmentation - MATLAB imsegkmeans.” [Online]. Available: <https://www.mathworks.com/help/images/ref/imsegkmeans.html> [Accessed Sep. 22, 2022].
- [14] M. Shahhosseini, G. Hu, and H. Pham, “Optimizing ensemble weights and hyperparameters of machine learning models for regression problems,” *Mach. Learn. with Appl.*, vol. 7, p. 100251, 2022. doi:10.1016/j.mlwa.2022.100251.
- [15] P. M. Granitto, H. D. Navone, P. F. Verdes, and H. A. Ceccatto, “Weed seeds identification by machine vision,” *Comput. Electron. Agric.*, vol. 33, no. 2, pp. 91–103, 2002. doi:10.1016/S0168-1699(02)00004-2.
- [16] J. G. A. Barbedo, “Detection of nutrition deficiencies in plants using proximal images and machine learning: A review,” *Comput. Electron. Agric.*, vol. 162, pp. 482–492, Jul. 2019. doi:10.1016/J.COMPAG.2019.04.035.
- [17] X. Ye, S. Abe, and S. Zhang, “Estimation and mapping of nitrogen content in apple trees at leaf and canopy levels using hyperspectral imaging,” *Precis. Agric.*, vol. 21, no. 1, pp. 198–225, Feb. 2020. doi:10.1007/S11119-019-09661-X/FIGURES/15.
- [18] Chen, S., Hu, T., Luo, L., He, Q., Zhang, S., Li, M., Cui, X. and Li, H., “Rapid estimation of leaf nitrogen content in apple-trees based on canopy hyperspectral reflectance using multivariate methods,” *Infrared Phys. Technol.*, vol. 111, p. 103542, Dec. 2020. doi:10.1016/J.INFRARED.2020.103542.
- [19] Morellos, A., Pantazi, X. E., Moshou, D., Alexandridis, T., Whetton, R., Tziotziou, G., Weibensohn, J., Bill, R. and Mouazen, , “Machine learning based prediction of soil total nitrogen, organic carbon and moisture content by using VIS-NIR spectroscopy,” *Biosyst. Eng.*, vol. 152, pp. 104–116, Dec. 2016. doi:10.1016/J.BIOSYSTEMSENG.2016.04.018.
- [20] G. Liu, S. Mao, and J. H. Kim, “A mature-tomato detection algorithm using machine learning and color analysis,” *Sensors (Switzerland)*, vol. 19, no. 9, pp. 1–19, 2019. doi:10.3390/s19092023.
- [21] C. Lisu, S. Yuanyuan, and W. Ke, “Rapid diagnosis of nitrogen nutrition status in rice based on static scanning and extraction of leaf and sheath characteristics,” *Int. J. Agric. Biol. Eng.*, vol. 10, no. 3, pp. 158–164, May 2017. doi:10.25165/IJABE.V10I3.1860.
- [22] M. Yang, D. Xu, S. Chen, H. Li, and Z. Shi, “Evaluation of Machine Learning Approaches to Predict Soil Organic Matter and pH Using vis-NIR Spectra,” *Sensors*, vol. 19, no. 2, p. 263, Jan. 2019. doi:10.3390/S19020263.
- [23] M. Altalak, M. A. Uddin, A. Alajmi, and A. Rizg, “Smart Agriculture Applications Using Deep Learning Technologies: A Survey,” *Appl. Sci.*, vol. 12, no. 12, Jun. 2022. doi:10.3390/app12125919.

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

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

