

# BAFONLINE: Uzaktan Spektral Veri Analizleri için Bir Web Uygulaması

## BAFONLINE: A Web Application for Remote Spectral Data Analysis

Fatih KAHRIMAN

Çanakkale Onsekiz Mart Üniversitesi,  
Ziraat Fakültesi, Çanakkale, Türkiye  
fkahriman@hotmail.com  
ORCID: 0000-0001-6944-0512

Musa TOPÇAKIL

BAF Elektronik Yazılım Tarım S ve Tic AŞ  
Çanakkale Tek. Geliştirme Bölgesi, Türkiye  
musa@baf-eyt.com.tr  
ORCID: 0000-0001-5805-9572

Aycan SÜTAL

BAF Elektronik Yazılım Tarım S ve Tic AŞ  
Çanakkale Tek. Geliştirme Bölgesi, Türkiye  
aycan@baf-eyt.com.tr  
ORCID: 0000-0001-5039-4595

Ömer GEZER

İstanbul Teknik Üniversitesi,  
Makine Müh. Bölümü, İstanbul, Türkiye  
gezero@itu.edu.tr  
ORCID: 0000-0002-0492-3844

### Öz

Spektral veri analizleri için geliştirilmiş çeşitli programlar ve web tabanlı uygulamalar mevcuttur. Son yıllarda web tabanlı uygulamalara olan ilgi bir hayli artmıştır. R program dili ve shiny paketi kullanılarak geliştirilmiş çok farklı alanlarda kullanılabilen web uygulamaları mevcuttur. Bu çalışma spektral veri analizlerini internet üzerinden yapabilen bir web uygulamasının geliştirilmesi amacıyla yürütülmüştür. Geliştirilen uygulama (BAFONLINE) RStudio arayüzü kullanılarak geliştirilmiştir. BAFONLINE, 1200-2400 nm arasında alınan spektral verinin destek vektör makineleri regresyonu (DVMR) tabanlı modelleme yaklaşımı ile analiz edilmesi sonrasında hedef değişken ile ilgili sonucu analiz ekranında gösterebilen kullanıcı dostu bir uygulamadır. Uygulamanın geliştirme aşamasında DVMR ve Rassal Orman (RO) yöntemlerine göre modeller oluşturulmuştur. Bu modeller, 250 adet öğütülmüş mısır numunesinden masaiüstü NIRS (Near InfraRed Spectroscopy)

cihazı ile 1200-2400 nm arasında toplanan spektral veriler bağımsız değişken olarak, bağımlı değişken olarak ise nem içeriği atanarak R platformunda geliştirilmiştir. Geliştirilen modellerin değerlendirme istatistiklerine göre DVMR modelinin ( $RMSE_{Cal}=0,484$ ,  $R^2_{Cal}=90,9$ ,  $RPD_{Cal}=3,32$ ,  $RMSE_{Val}=0,317$ ,  $R^2_{Val}=95,6$ ,  $RPD_{Val}=4,79$ ,  $RMSE_{Cal}=0,634$ ,  $R^2_{Cal}=52,2$ ,  $RPD_{Cal}=1,45$ ) RO modelinden daha daha yüksek güvenilirliğe sahip olduğu anlaşılmıştır. Geliştirilen DVMR modeli uygulamanın arayüzüne uygun kodlar kullanılarak entegre edilmiştir. Geliştirilen arayüze <https://baf.shinyapps.io/BAFONLINE/> linkinden erişilebilmekte ve veri girişinin ardından spektral analizler kolaylıkla gerçekleştirilebilmektedir. İleriki araştırmada bu uygulamaya farklı spektral cihazlar ile farklı tarımsal ürünlere yönelik yeni tahmin modellerinin eklenmesi düşünülmektedir.

**Anahtar Sözcükler:** R programlama, Spektral analizler, Makine öğrenmesi

### Abstract

There are many software and web-based applications for spectral data analysis. Recently, the interest in web based applications has

Gönderme ve kabul tarihi: 02.10.2020 - 17.11.2020

Makale türü: Araştırma

increased. The shiny software package which has been developed with the R language has been used in different web applications in many fields. This study has been conducted for development of a web based application that can analyze spectral data online. The application (BAFONLINE) has been developed using the RStudio interface. BAFONLINE is a user friendly application where one can analyze the spectral data within 1200-2400 nm interval utilizing support vector machine regression (SVMR) based modeling approach and can display the result of the target variable. SVMR and Random Forest (RF) methods were used and compared in the development step of application. These models, the spectral data has been acquired using a desktop NIRS (Near InfraRed Spectroscopy) device from 250 ground corn grains, in the 1200-2400 nm spectral range as dependent variable, and the spectral models have been developed under the R platform, selecting the moisture content as the dependant variable. According to the evaluation statistics of the developed models, it was understood that SVMR model ( $RMSE_{Cal}=0.484$ ,  $R^2_{Cal}=90.9$ ,  $RPD_{Cal}=3.32$ ,  $RMSE_{Val}=0.458$ ,  $R^2_{Val}=87.4$ ,  $RPD_{Val}=2.77$ ) has higher robustness than RF model ( $RMSE_{Cal}=0.317$ ,  $R^2_{Cal}=95.6$ ,  $RPD_{Cal}=4.79$ ,  $RMSE_{Val}=0.634$ ,  $R^2_{Val}=52.2$ ,  $RPD_{Val}=1.45$ ). This model have been integrated into the interface using appropriate codes. The interface can be accessed from the <https://bafr.shinyapps.io/BAFONLINE/> address and the spectral analyses can be easily performed. In the future studies, support for additional spectroscopy devices and agricultural products is planned to be added.

**Keywords:** R Programming, Spectral analyses, Machine learning

## 1. Giriş

Spektral veri, çeşitli cihazlardan alınan ve farklı şekillerde kaydedilen büyük veri setleridir. Özellikle spektral tarama amaçlı geliştirilen cihazlardan elde edilen veriler oldukça fazla bilgi içermektedir. Bu veri kullanılan spektral cihazlara göre ışık ile madde etkileşiminden doğan sayısallaştırılmış ölçümlerdir. Bu ölçümlerin anlamlı hale getirilebilmesi için tahmin modellerine ihtiyaç duyulmaktadır. Masaüstü spektral cihazların büyük kısmında bu modeller

cihaza entegre olarak bulunmakta veya kullanıcı ihtiyaçları doğrultusunda lokal ve genel kalibrasyonlar olarak geliştirilme suretiyle daha sonra da yüklenebilmektedir [1].

Mevcut spektral cihazlarda veya bu cihazlar ile toplanan verilerin anlamlı hale getirilebilmesi için çeşitli modelleme yöntemlerinden yararlanılmaktadır. Spektral analizlerde kullanılan modelleme yöntemleri öğrenme tabanlı olmayan modeller ve öğrenme tabanlı modeller olarak iki grup altında sınıflandırılmaktadır [2]. Öğrenme yaklaşımı içermeyen yöntemler klasik modelleme teknikleridir ve Kısmi En Küçük Kareler Regresyonu (KEKR), Çoklu Doğrusal Regresyon (ÇDR), Temel Bileşen Regresyonu (TBR) bunlardan birkaçıdır. Öğrenme yaklaşımli modeller ise yapay sinir ağları (YSA), destek vektör makineleri (DVM), rassal orman (RO) gibi makine öğrenmesine dayalı metotlardır [3, 4]. Bu modeller içerisinde DVM ve RO teknikleri yenilikçi yöntemlerdendir ve hem kantitatif hem de kalitatif sınıflama modellerin geliştirilmesi amacıyla kullanılmaktadır [4, 5]. DVM yönteminin kalitatif ayırım gerektiren problemlerin çözümünde kullanılan alt metoduna Destek Vektör Makineleri Sınıflaması (DVMC), kantitatif verilerle ilgili problemlerin çözümünde kullanılan alt metoduna ise Destek Vektör Makineleri Regresyonu (DVMR) adı verilmektedir. Kalitatif ister kantitatif değişkenlere yönelik çözümlerde kullanılan ve doğrusal olmayan modelleme imkânı veren bu DVM yöntemlerinin önemli bir yönü çekirdek fonksiyonudur. Yaygın olarak bu yöntemde farklı çekirdek fonksiyonu kullanılmakla birlikte radyal tabanlı fonksiyon (RTF) doğrusal olmayan problemlerin çözümünde en fazla tercih edilen yöntemler arasındadır [6]. RTF'yi diğer çekirdek fonksiyonlarından ayıran en önemli özellik diğer polinomial çekirdek fonksiyonlarından daha düşük sayıda hiper parametresinin bulunması ve daha az sayısal problem içermesidir [7]. Bu nedenle RTF regresyon problemlerinin DVM ile çözümünde yaygın olarak tercih edilen yöntemlerdendir.

Yakın geçmişe kadar ticari olarak kullanılan veya prototip olarak geliştirilen spektral ölçüm cihazlarının büyük kısmında klasik yöntemler olarak kabul edilen ÇDR ve KEKR gibi yöntemlere dayalı yerleşik kalibrasyon modelleri

yaygın olarak bulunmaktadır. Ancak bu tahmin modellerinde kullanılan klasik yöntemler ışık ile madde arasındaki doğrusal olmayan ilişkileri yeterince açıklayamamaktadır. Işık ve madde arasındaki ilişkiler her zaman doğrusal olmadığından, bu ilişkiye yönelik doğrusal olmayan ve daha kuvvetli yöntemlere ihtiyaç duyulmaktadır. Bu nedenle DVM gibi kuvvetli makine öğrenmesi yöntemlerine olan rağbet bilimsel alanda olduğu kadar uygulamada da gün geçtikçe artış göstermektedir. Spektral analiz cihazlarında kullanılmak üzere tahmin modellerin oluşturulabilmesi için günümüzde farklı yazılım çözümleri üretilmiştir. Bu yaklaşımlarla geliştirilen yazılımları temel olarak üç grupta toplamak mümkündür. Birincisi yerleşik kurulum gerektiren (stand-alone) yazılımlardır ve bu yazılımlar ölçüm yapan spektral cihazlar üzerine kurulu ve çoğunlukla cihaza özel olarak geliştirilmiş programlardır. İkinci grup yazılımlar ise taşınabilir (portable) özellik taşıyan programlardır ancak bu çözümler kısa süre içerisinde yerini üçüncü gruba yani uzaktan erişim sağlanabilen yazılımlara bırakmıştır. Üçüncü grup olan ve yenilikçi yaklaşımlar içeren bu çözüm yolu yalnızca spektral veri analizlerinde değil, kimyasal analiz cihazlarından elde edilen verilerin analizi gibi farklı amaçlarla da kullanılmaya başlanmıştır.

## 2. Geçmiş Çalışmalar

Spektral cihazlarla tarımsal ürün analizlerinin yapılmasına yönelik araştırmaların geçmişi oldukça eskilere dayanmakla birlikte son yıllarda bu alanda yürütülen çalışmalar önemli bir ivme kazanmıştır. Spektral analizler ile tahıl, kahve, çay, et, süt, yumurta, yağ gibi çok çeşitli tarımsal ürünlerin kalitesini ölçmek amacıyla kullanılmıştır [8]. Tarımsal ürünlerde protein, karbonhidrat, yağ, nem gibi içerik analizleri için spektral analiz cihazlarından yararlanılmaktadır [1]. Nem içeriği tarımsal ürünler içerisinde en yaygın kullanılan ölçümlerden birisidir. Nitekim yakın kızıl ötesi spektroskopisi ile tarımsal ürün analizlerini konu eden ilk araştırmalar da nem içeriği üzerine olmuştur [9]. Nem içeriği gerek ürün fiyatlandırılmasında, hasat zamanlarının belirlenmesinde ve bu ürünlerin depolanması gibi kritik konularda ilk ölçümlenen özelliklerdendir. Nem içeriğinin hızlı ve doğru olarak ölçülebilmesi

için pratik araçlara ihtiyaç duyulmuştur. Bu amaçla yakın kızıl ötesi spektroskopisinden farklı tarımsal ürünlerin nem içeriğini tespit etmeye yönelik tahmin modelleri geliştirilmiştir. Bu çalışmalarda nem içeriğini tahminlemeye yönelik modellerin geliştirilmesi ya da model geliştirmede kullanılan kemometrik tekniklerin kıyaslanmasını konu edinmektedir [10, 11].

Spektral cihazlarda analiz edilmek istenen bileşene yönelik sonuç alınabilmesi için bu bileşene yönelik olarak oluşturulmuş kalibrasyon adı verilen tahmin modellerine ihtiyaç duyulmaktadır. Günümüzde kullanılan spektral cihazlarda bu tahmin modelleri cihazlara yerleşik olarak bulunmaktadır. Tahmin modelinin güncellenmesi uzaktan veya lokal bağlantı ile gerçekleştirilebilir de, bu yol pratik bir çözümden uzaktır. Dahili tahmin modellerine alternatif olarak uzaktan erişim ile spektral veri analizi yapabilen yazılımların veya platformların geliştirilmesi güncel konu başlıklarındandır.

Uzaktan erişimli yazılımlar, diğer bir ifade ile çevrimiçi uygulamalar kod yazımı gerektirmeyen menü veya butonlarla kontrol edilebilen kullanıcı dostu çözümlerdir. Bu uygulamaların geliştirildiği çok farklı platformlar bulunmakla birlikte, R programlama diline dayalı platformlar bu çevrimiçi uygulamaların geliştirildiği başlıca seçeneklerden birisidir. R programlama dili, Ihaka ve Gentelmann [12] tarafından geliştirilen ve günümüzde R Core Team [13] olarak bilinen bir platformdur. RStudio arayüzü ile eşgüdümlü çalışabilen ve çevrimiçi arayüz geliştirmeye yönelik olarak oluşturulan shiny paketi ve çerçevesi [14] sayesinde bu platform popüler hale gelmiştir. Tabii ki R programında veri analizine yönelik uygulamaların geliştirilebilmesi için öncelikle bu verilere uygun paketlerin geliştirilmesine ihtiyaç duyulmuştur. Gerek spektral verilerin görüntülenmesi gerekse spektral model oluşturma amacıyla farklı paketler geliştirilmiştir. Bu paketlerden bazıları pls [15], mdatools [16], caret [17], ChemometricwithR [18], e1071 [19] ve prospectr [20] olarak örneklendirilebilir. Bu paketlerin fonksiyonları spektral veri analizi ve modelleme amaçlı olarak kullanılabilmesine karşın tamamının R arayüzüne kod yazımı gerektirmesi kullanıcıları zorlamaktadır. Diğer taraftan bu paketlerin hiçbirisi model oluşturmadan analiz imkânı

sunmamaktadır. Bu anlamda söz konusu paketlerin çevrimiçi veri analizine yönelik çözüm sunduklarını söylemek zordur. Bu anlamda en pratik örneklerden birisi [21] tarafından geliştirilmiş olan Speclayzer uygulamasıdır. Bu uygulama sayesinde taşınabilir spektrometre cihazları ile toplanan veri spektral indisler denilen çeşitli hesaplamalara çevrimiçi olarak dönüştürebilmektedir. Masaüstü cihazlarda bulunan dahili modeller genel olarak analiz edilen örneklere yönelik biyokimyasal içerik analizlerine yöneliktir.

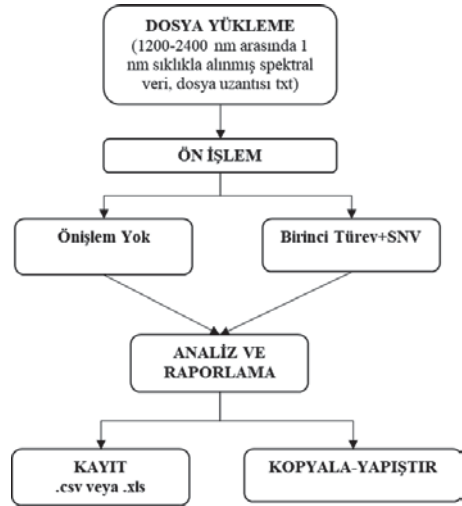
Bilimsel çalışmalarda masaüstü spektral cihazlarından alınan verileri internet üzerinden işlemeye yönelik bir uygulamaya rastlanmamıştır. Bu noktadan hareketle, bu çalışma masaüstü spektral cihazlardan alınan verilerin web üzerinden analiz edilmesini sağlayabilen bir web uygulaması geliştirmek amacıyla yürütülmüştür. Geliştirilen uygulamada örnek tarımsal ürün olarak mısır unu ve analiz edilecek değişken olarak nem içeriği kullanılmıştır.

### 3. Yöntem

#### 3.1 Uygulamanın Yapısı ve Geliştirilmesi

Geliştirilen arayüzün kullanım sürecine ait akış diyagramı Şekil 1'de gösterilmiştir. Bu uygulamada analiz işleminin gerçekleştirilebilmesi için uygulamada kullanılan tahmin modelindeki değişken yapısına uygun verinin girilmesi gerekmektedir. Bu nedenle ilk adım olarak öğütülmüş mısır numunesinden 1200-2400 nm arasında her 1 nm'de alınmış 1201 adet dalga boyunun reflektans verisine ihtiyaç duyulmaktadır. Analiz edilecek veri seti uygulamaya yüklendikten sonra spektral ön işlem seçeneği belirlenir. Spektral ön işlem uygulamaları cihazlardan alınan verilerde istenmeyen bazı değişimleri (gürültü vb.) ortadan kaldırmak amacıyla kullanılan teknikleri kapsamaktadır. Uygulamada birinci türev (FD) ve standart normal değişim (SNV) ön işlem seçenekleri birlikte kullanılmıştır. Uygulama arayüzünde belirlenen spektral ön işleme göre (Transformasyon Yok, FD+SNV) veri uygun modele uygulanır. Ön işlem seçiminin ardından spektral veriye ait grafik görüntülenir. Şayet spektral veride bir problem yok ise analiz

aşamasına geçilir. Analiz sonuçları menüsünden hesaplama sonuçları görüntülenir ve kullanıcı istediğine bağlı olarak csv veya excel olarak sonuçlar kayıt altına alınabileceği gibi, ekranda görülen sonuçlar kopyala-yapıştır yöntemiyle boş bir dosyaya da aktarılabilir.



Şekil-1: Geliştirilen uygulamanın akış diyagramı

Arayüz geliştirme işlemi RStudio platformunda shiny paketi kullanılarak [6] gerçekleştirilmiştir. BAFONLINE temel olarak üç menüden oluşmaktadır. Birinci menü (HAKKINDA) BAFONLINE ile genel bilgileri içeren sayfadır. İkinci menü (DENEME SÜRÜMÜ) kullanıcıların deneme sürümü olarak arayüzü test etme olanağı sağlayan menüdür ve temelde beş fonksiyon taşımaktadır. Bu fonksiyonlar veri yükleme, veri transformasyonu, grafiksel gösterim, sonuçların görüntülenmesi ve elde edilen sonuçların dosya olarak indirilmesidir. Arayüz başlangıç dosyası (app) birleşik dosya olarak shiny, Roxygen2 ve devtools paketleri kullanılarak oluşturulmuştur. Oluşturulan app dosyası standart olarak kullanıcı arayüzü (UI) ve fonksiyonların internet üzerinden çalışmasını sağlayan ayrı bir bölümden (server) oluşmaktadır. Arayüzün yukarıda belirtilen menü görünümünü sağlamak amacıyla shiny paketinin standart kodlarından yararlanılmış ve kullanıcı arayüzü oluşturulmuştur. Arayüze fonksiyonların kazandırılması amacıyla ise Çizelge 1'de belirtilen paketlerin fonksiyonlarından yararlanılmıştır.

**Çizelge-1: Arayüz fonksiyonları ve bu fonksiyonları işlevsel hale getirmek amacıyla faydalanan paketler**

Paket Fonksiyonu	Kullanılan paket	Kaynak
Veri Yükleme	shiny	Chang vd. [14]
Grafiksel gösterim	mdatools	Kucheryavskiy [16]
Veri transformasyonu	prospectr	Stevens and Ramirez-Lopez [20]
Analiz	e1070	Meyer vd. [19]
Raporlama	DT	Xie vd. [22]

### 3.2. Uygulamaya İlişkin Tahmin Modellerinin Geliştirilmesi

R tabanlı web uygulamalarında analiz ve raporlama fonksiyonlarının aktif hale gelebilmesi için uygulamanın model dosyasına ihtiyacı vardır. Spektral cihazlarda kalibrasyon veya kütüphane olarak daha dosyalar spektral veri yardımıyla tahmin edilmek istenen değişkene ilişkin matematiksel modeli içermektedir. Çalışmamızda bu tahmin modelleri öğütülmüş mısır numunelerinde nem içeriğinin belirlenmesi amacıyla R programında geliştirilmiştir. Uygulama arayüzünde deneme sürümü için geliştirilen model daha önce Kahrıman ve ark. [23] tarafından yürütülen bir araştırma projesinde, masaüstü spektral bir cihaz (Spectrastar 2400, Unity Scientific, USA) ile mısır ununda nem içeriğinin tespitini mümkün kılan bir tahmin modelinin verileri ile oluşturulmuştur. Masaüstü NIR cihazı 1200-2400 nm arasında her 1 nm'de ölçüm alabilmektedir ve modellerde tahminleyici değişken olarak toplamda 1201 dalga boyu kullanılmıştır. Bu cihaz ile toplanmış spektral veriler uygun programlar kullanılarak metin dosyasına dönüştürülmüş ve model geliştirme amacıyla kayıt altına alınmıştır. Aynı çalışmada [23] referans analizlerle elde edilen spektrumların alındığı örneklerle ait referans yöntem ile elde edilmiş nem içeriği değerleri kullanılmıştır. Model geliştirme amacıyla spektral veriler ve nem içeriğine ait sonuçlar RStudio arayüzüne aktarılmıştır. Model geliştirme öncesi spektral verilere birinci türev ardından *SNV* dönüşümü uygulanmıştır. Bu dönüşüm ile spektral verideki düzensizlikler ve istenemeyen değişimler uzaklaştırılmaya çalışılmış olup birinci türev (FD) ve *SNV* uygulamasına ilişkin matematiksel

formüller Denklem 1 ve Denklem 2'de sunulmuştur. Bu formüllerde  $X_i$ : örnek spektrum değerini,  $N$ : normalizasyon katsayısını,  $k$ : gap genişliğini,  $c_h$ : polinom ve türev derecesine göre hesaplanan katsayısı,  $\bar{X}_i$ : örnek spektrumunun ortalamasını,  $\sigma_{xi}$ : örnek spektrumuna ait standart sapmayı göstermektedir. Birinci türev hesaplamasında gap genişliği 15, polinom ve türev derecesi ise 1 alınmıştır.

$$FD_{i,f} = \frac{1}{N} \sum_{h=-k}^k c_h x_{i+h} \quad (1)$$

$$SNV = \frac{x_i - \bar{X}_i}{\sigma_{xi}} \quad (2)$$

Dönüşüm işleminin ardından destek vektör makineleri regresyonu (DVMR) yöntemi kullanılarak oluşturulan tahmin modellerinde nem içeriği bağımlı değişken, cihazlardan alınan veriler ise ayrı ayrı bağımsız değişken olarak atanmak suretiyle tahmin modeli geliştirilmiştir. Destek vektör makineleri yönteminde eps-regresyon metodu ve kernel tipi olarak radyal tabanlı fonksiyon kullanılmıştır. Hiper parametre optimizasyonu için gamma, cost ve epsilon değerleri grid tarama yöntemi ile belirlenmiş ve tahmin modelinde bu parametreler kullanılmıştır. Grid taramada cost değeri (1, 5, 10, 50, 100, 1000, 10000), gamma değeri (0,00001, 0,01, 0,1, 1, 2, 5, 10) ve epsilon değeri (0,1, 0,01, 0,001) aralık taramasında kullanılmıştır. Bu tarama sonuçlarına göre model oluşturulur iken gamma değeri 0,00001, cost değeri 10000 ve epsilon değeri 0,01 olarak atanmıştır. Denklem 3'te sunulan radyal tabanlı fonksiyonda  $\gamma$ : en yüksek gamma değerini gösteren parametredir,  $u$  ve  $v$ : çekirdek fonksiyonundaki destek vektörleridir. Denklem 4'te sunulan destek vektör makineleri hesaplamaları için kullanılan genel formülde  $x$ : girdi vektörü,  $\omega$ : hiper düzeleme olan uzaklık,  $b$ : bias değeridir.

$$RBF = \exp\{-\gamma|u - v|^2\} \quad (3)$$

$$DVM = \omega \times x + b \quad (4)$$

İkinci modelleme yöntemi olarak kullanılan RO yöntemine göre geliştirilen modellerde dönüşüm uygulanmış spektral veri seti bağımsız değişken, nem içeriği ise bağımlı değişken olarak alınmıştır. RO yönteminde maksimum ağaç sayısı ve her kademedeki değişken sayıları grid tarama yöntemi ile belirlenmiştir. Bu taramaya sonucunda

belirlenen optimizasyon parametreleri (ağaç sayısı=250, düğüm büyüklüğü=24, maksimum düğüm sayısı=26) geliştirilen modele tanımlanmıştır. RO yöntemi ile regresyon çözümüyle oluşturulan ağaç yapısındaki her katmandaki hata kareler ortalamasına (HKO) göre gerçekleştirilmiştir. HKO ortalaması aşağıdaki Denklem 5'e göre belirlenmiştir. Bu formülde N veri sayısını, fi i. nokta için modele ait tahmin sonucunu, yi ise i. noktanın gerçek değerini göstermektedir.

$$HKO = \frac{1}{N} \sum_{i=1}^N (f_i - y_i)^2 \quad (5)$$

Tahmin modellerinin güvenilirlik ve geçerlilik düzeylerini sınamak amacıyla Denklem 6, 7 ve 8'de sunulan hesaplama formüllerinden yararlanılmıştır. Bu formüllerden elde edilen istatistikler kalibrasyon (Cal) ve dış doğrulama setleri (Val) için ayrı ayrı hesaplanmıştır.

$$RMSE = \sqrt{\frac{\sum (Y_{pred} - Y_{ref})^2}{n}} \quad (6)$$

$$R^2 = \left( \frac{\sum_{i \in S_k} (\hat{Y}_{pred} - \bar{Y}_{pred}) \cdot (Y_{ref} - \bar{Y}_{ref})}{(n-1) \cdot STD_{pred} \cdot STD_{ref}} \right)^2 \quad (7)$$

$$RPD = \frac{STD_{ref}}{SE_{pred}} \quad (8)$$

Formüllerde; *RMSE*: hata kareler ortalamasının karekökünü; *Y<sub>pred</sub>*: tahminlenen değer; *Y<sub>ref</sub>*: referans analizle elde edilen değer; *n*: örnek sayısı; *R<sup>2</sup>*: regresyon katsayısını; *RPD*: standart hatanın standart sapmaya oranı; *STD<sub>ref</sub>*: referans analizlere ait veri setindeki standart sapma, *STD<sub>pred</sub>*: tahmin setine ait veri setindeki standart sapma, *SE<sub>pred</sub>*: tahminlenen değerlere ait standart hatayı göstermektedir. Bu istatistiklere göre başarılı bulunan tahmin modelleri RDS uzantılı olarak bir dosya kayıt edilmiş ve bu dosya uygulama dosyası ile (app) aynı klasöre kayıt edilmiştir. Tüm işlemler tamamlanıp uygulamanın ön testleri tamamlanmasının ardından uygulama yayınlama ile ilgili standart prosedürler uygulanarak dosyalar shiny server üzerine yüklenmiştir. Arayüzü shiny server üzerinden kullanmak için <https://baf.shinyapps.io/BAFONLINE/> bağlantısı üzerinden erişime açılmıştır.

BAFONLINE uygulamasından elde edilen sonuçların kıyaslanması ve dış doğrulama işlemi

için kalibrasyon setinde bulunmayan 182 farklı öğütülmüş mısır örneğinden masaüstü NIR cihazı ile spektrum verisi alınmış ve bu örneklerin nem içerikleri kalibrasyon geliştirme amacıyla kullanılan referans yöntemine göre belirlenmiştir. Bu amaçla spektrum verisi örneklerden yaklaşık 2 g örnek ağırlığı kaydedilmiş petrilere konulmuş ve 105 °C'de 3 saat süre ile kurutulmuştur. Kurutma sonrası ağırlık hassas terazide belirlenmiş ve Denklem 8 yardımıyla örneklerin nem içerikleri tespit edilmiştir.

$$Nem = 100 - \left( \frac{Kur.Son.Ağ. - Petri.Ağır.}{Örnek.Ağırlığı} \times 100 \right) \quad (8)$$

Geliştirilen uygulamanın fonksiyonel olarak çalışıp çalışmadığını test etmek amacıyla dış doğrulama örneklerine ait spektral veri BAFONLINE arayüzüne internet üzerinden yüklenmiştir. Uygulama üzerinden hesaplama gerçekleştirilmiş ve referans analiz ile elde edilen sonuçlarla uygulamadan elde edilen veriler kıyaslanmıştır. Elde edilen verileri karşılaştırmak amacıyla BAFONLINE ve referans analizlerle elde edilen tanımlayıcı istatistiklerden yararlanılmıştır. Referans analizler ile BAFONLINE uygulamasından elde edilen ortalamalar arasındaki farklar Kruskal-Wallis testi ile kıyaslanmıştır. İki veri grubu arasındaki ilişkiler korelasyon ve regresyon analizleri ile irdelenmiş, sonuçlar da grafiksel olarak gösterilmiştir. Ayrıca dış doğrulama seti için *RMSE*, *SEE* ve *RPD* hesaplamaları tekrar edilmiştir.

## 4. Uygulamanın Gerçekleştirilmesi

### 4.1 Tahmin Modellerinin Değerlendirilmesi

Deneme sürümünde kullanılmak üzere nem içeriği analizi için geliştirilen modellerden hangisinin kullanılacağını belirlemek amacıyla hesaplanan istatistikler Çizelge 2'de gösterilmiştir.

NIR spektroskopi kalibrasyon modellerinin güvenilirliğini değerlendirmek amacıyla *RMSE* istatistiklerinin düşük, *RPD* değerinin ise yüksek olması istenmektedir. Ancak *R<sup>2</sup>* ve *RPD* değerlerine göre NIR kalibrasyon modellerinin güvenilirlikleri hakkında farklı değerlendirmeler mevcuttur. Williams ve Sobering [24] kalibrasyon

modelinin güvenilir kabul edilmesi için RPD değerinin alt sınırını 3 olarak rapor etmiştir. Diğer bir çalışmada ise başarılı kalibrasyon modellerinde  $R^2$  değerinin 90'ın üzerinde RPD değerinin ise 3,0-4,0 arasında olması gerektiğini vurgulamıştır [25]. Bu değerlendirmelere göre, Çizelge 2'de nem içeriği için oluşturulan modeller içerisinde DVMR modelinin RMSE değerlerinin düşük,  $R^2$  değerinin ise yüksek olduğu görülmektedir. RO yöntemine göre geliştirilen tahmin modeli kalibrasyon setinde başarılı bulunmasına karşın, dış doğrulama işleminde güvenilirlik düzeyi bakımından DVMR modelinin gerisinde kalmıştır. Onaç ve ark. [11] mısır ununda nem içeriğinin NIR ile tespitine yönelik oluşturduğu EKKR yöntemiyle geliştirdiği tahmin modelinde SECV değerini 0,468,  $R^2$  değerini 91,2 ve RPD<sub>Val</sub> değerini 3,40 olarak belirlemiştir. DVMR yönteminden elde edilen sonuçlar bu çalışma sonuçlarına benzerlik göstermektedir.

**Çizelge-2: Kalibrasyon ve dış doğrulama setleri için hesaplanan değerlendirme parametreleri**

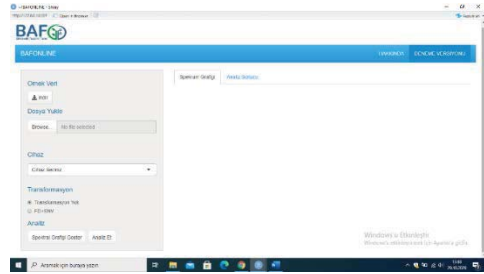
Kalibrasyon (n=250)	RMSE <sub>Cal</sub>	$R^2$ <sub>Cal</sub>	RPD <sub>Cal</sub>
DVMR	0,484	90,9	3,32
RO	0,317	95,6	4,79
Dış Doğrulama (n=182)	RMSE <sub>Val</sub>	$R^2$ <sub>Val</sub>	RPD <sub>Val</sub>
DVMR	0,458	87,4	2,77
RO	0,634	0,522	1,45

Uygulamada kullanılmak üzere farklı yöntemlerle oluşturulan modeller içerisinde DVMR yönteminin daha başarılı sonuç verebileceği anlaşılmıştır. Bu sonuçlara dayalı olarak geliştirilen uygulamaya DVMR yöntemi ile oluşturulan model tanımlanmıştır.

## 4.2 Uygulamanın Testi

BAFONLINE uygulamasının testi için dış doğrulama setinde kullanılan veri seti (Ek Dosyalar:TEST.txt) kullanılmıştır. Bu veriye dergi web sayfasındaki ek dosyalardan veya uygulama arayüzünden “Örnek Dosya” seçeneği altındaki indirme butonu kullanılarak erişilebilmektedir. Analiz için bu veri setinin girişi uygulama arayüzünden dosya yükleme

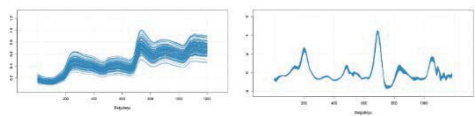
bölümünden gerçekleştirilmiştir. Bunun için Şekil 2’de görüldüğü gibi öncelikle uygulamaya ait sayfaya (<https://bafir.shinyapps.io/BAFONLINE/>) internet üzerinden erişim sağlanmıştır. Yüklenen test dosyası model geliştirmede kullanılan spektral cihazdan alınmış ve metin dosyasına dönüştürülmüş halde ve sütunlarda dalga boyları satırlarda başlıksız olarak örnek kodları bulunacak şekilde yüklenmiştir. Nem içeriği için geliştirilen model FD+SNV dönüşümüne tabi tutulmuş verilerle oluşturulduğundan, “Transformasyon Yok” seçeneği seçilerek uygulamanın “FD+SNV” işlemini gerçekleştirilmiştir. Bu seçim kullanılan modellerin tahmin başarısına önemli ölçüde etki ettiğinden mutlaka yüklenen veriye uygun seçim yapılmış olmalıdır.



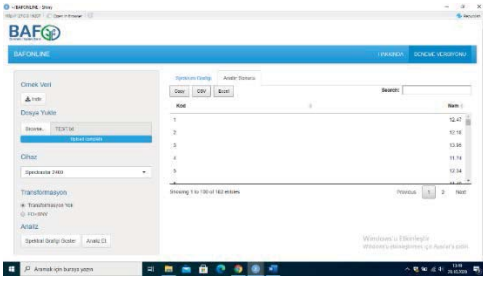
**Şekil-2: Geliştirilen uygulamanın kullanıcı arayüzü**

Spektral verinin yüklenmesi ve transformasyon seçeneğinin ardından kullanıcı veriye ilişkin grafiği izlemek amacıyla “Spektral Grafiği Göster” butonunu kullanılmıştır. Çalışmada kullanılan spektral verinin ön işlemsiz ve FD+SNV transformasyonu uygulanmış haline ait grafiksel çıktılar Şekil 3’te görülmektedir.

Spektral grafikte herhangi bir izlenmemiş ve Şekil 4’te görüldüğü gibi kullanıcı “Analiz Et” butonunu kullanarak arayize örnek spektral veriye ait sonuçları “Analiz Sonucu” ekranında görüntülenmiştir. Bu sonuçları toplu veya satır seçimi yaparak .csv ve .xls uzantılı olarak kaydedilerek bir sorun olup olmadığı kontrol edilmiştir. İlave olarak kopyalama butonu ile farklı bir dosyaya da aktarabilmektedir.



**Şekil-3: Arayüz üzerinden spektral verinin transformasyonsuz (sol) FD+SNV transformasyonu (sağ) sonrasına ait grafikleri**



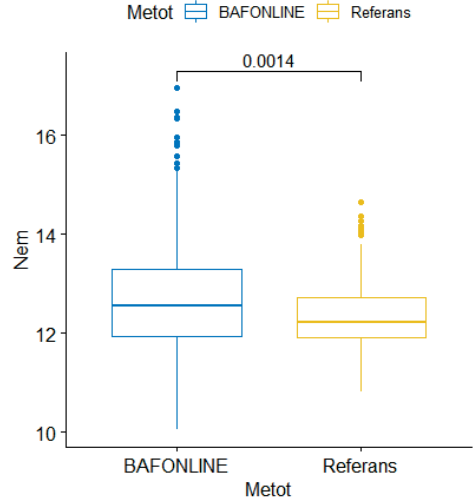
Şekil-4: Geliştirilen uygulamanın sonuç sayfası

### 4.3 Test Sonuçlarının Değerlendirilmesi

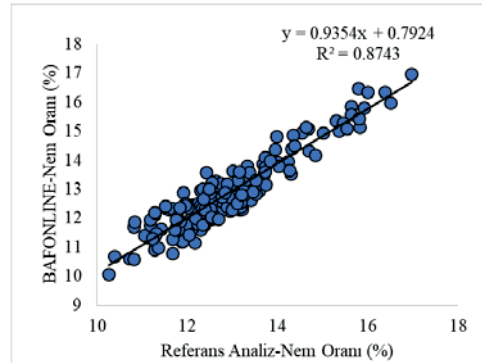
Çalışmada oluşturulan uygulama ile kalibrasyon setinde hiç kullanılmayan 182 adet mısır numunesine ait nem içeriği model sonuçlarının tanımlayıcı istatistikleri Çizelge 3’te sunulmuştur. Geliştirilen uygulama ile referans analiz ortalaması yakın bulunmuştur. BAFONLINE uygulamasından elde edilen sonuçlardaki değişim ile (Min=%10,26, Mak= 16,98) referans analiz sonuçları da benzerlik göstermiştir. Buna bağlı olarak Çizelge 3’te görüldüğü üzere BAFONLINE uygulamasından elde edilen sonuçlara ait standart sapma değeri referans analiz sonuçlarına oldukça yakın bulunmuştur.

Çizelge-3: BAFONLINE ve referans analizlerle elde edilen nem değerlerine ilişkin tanımlayıcı istatistikler

Veri Grubu	n	Ort.	Min	Maks	Std.Sap
BAFONLINE	182	12,84	10,26	16,98	1,26
Referans Analiz	182	12,80	10,05	16,95	1,27



Şekil-5: Dış doğrulama örnekleri ile referans analiz ve BAFONLINE sonuçları



Şekil-6: Dış doğrulama örnekleri ile referans analiz ve BAFONLINE uygulamasının sonuçları arasındaki regresyon grafiği

Araştırmada doğrulama seti olarak kullanılan örnek setinde referans analizler ve uygulama üzerinden DVMR modeli ile tespit edilen nem değerleri arasındaki farklar ve ilişkiler Şekil 5’te gösterilmiştir. Şekil 5’te görüleceği üzere dış doğrulama setinde referans analizler ile BAFONLINE uygulamasından elde edilen ortalamalar birbirine yakın olmasına rağmen, bu iki grup ortalaması arasında istatistikî açıdan önemli farklar bulunmuştur. Bu sonuç BAFONLINE uygulamasından elde edilen sonuçlardaki değişimin yüksek olmasından ve



örnek sayısının fazla olmasından ileri gelmektedir. Regresyon analizi sonuçlarına göre ise (Şekil 6) referans analizi ile uygulama sonuçları arasında yüksek düzeyde bir benzerlik olduğu ( $R^2=0,8545$ ) görülmüştür. Onaç ve ark [11] tarafından bu çalışmada model geliştirme amaçlı kullanılan veri seti ile Kısmi En Küçük Kareler Regresyonu (KEKR) yöntemi ile geliştirilen ve tüm spektral verileri kullanan modelin dış doğrulama setinde  $R^2$  değeri 0.861 olarak tespit edilmiştir. Bu çalışmada geliştirilen ve uygulamada kullanılan DVMR modeli de kalibrasyon seti dışındaki örneklerle yapılan doğrulama testi sonucunda benzer  $R^2$  değeri göstermiştir.

## 5. Sonuç

Bu araştırma ile spektral verilerin uzaktan analizine imkân veren fonksiyonel bir web uygulaması geliştirilmiştir. Çalışmada kullanılan veri setine uygun ölçüm yapan spektral cihazlar (1200-2400 nm aralığında 1 nm sıklıkla) veya daha geniş tarama aralığına sahip olan cihazlardan ayrıştırılan veriler ile mısır örneklerinden alınan spektral ölçümlere dayalı olarak nem içeriği tespit edilebilir. Uygulama mısır ıslahı ve araştırma geliştirme faaliyetleri yürüten kamu kurumları (üniversiteler, araştırma enstitüleri vb.) veya bu ürünün ticaretini yapan firmalar tarafından kullanılabilir. Tabii ki kullanıcıların geliştirilen modellerde kullanılan spektral veri toplama cihazlarına sahip olması gerekmektedir. Spektral cihazlarda çalışma alanına göre istenilen tüm ürün veya değişkenlere ilişkin kalibrasyonların bulunmaması veya bu kalibrasyonların yüksek fiyatlı olması bu kalibrasyonların geçici süre ile kullanımına olan ilgiyi artırmaktadır. BAFONLINE uygulaması da bu anlamda geçici süre ile spektral veri analiz etmeyi mümkün kılan bir uygulama örneği olarak geliştirilmiştir. Diğer taraftan geliştirilen uygulama ülkemizde benzer araştırmalar yürütecek araştırmacılara örnek bir yol ortaya koyması bakımından da önem arz etmektedir. Spektral ölçüm yapan farklı cihaz türlerinden (Raman spektroskopisi FT-IR vb.) alınan verinin tahmin modeline uygulanarak sonuç almaya yönelik her türlü araştırma için bu çalışmada izlenen yol ve platformlar kullanılarak yeni uygulamaların geliştirilebileceği düşünülmektedir.

R programlama dilinin uyumluluk ve hesaplama fonksiyonlarındaki etkinliği sayesinde BAFONLINE uygulamasının ileri hesaplama yöntemlerini kullanması ve pratikte faydalı olabilecek bir uygulamadır. Nitekim spektral cihazların tarımsal ürün analizlerindeki kullanımı gittikçe yaygınlaşmakta ve kullanıcıların kalibrasyon çeşitliliğindeki ihtiyacı da buna bağlı olarak artmaktadır. İleride yapılacak çalışmalar ile mısır dışındaki farklı bitkisel ürünlere ilişkin modellerin bu uygulamaya eklenmesi sayesinde spektral cihazlardan toplanan veriler ile analiz edilen örneklerin biyokimyasal içerik sonuçlarının alınması mümkün olabilir. Ayrıca kullanılan spektral cihazların çeşitlendirilmesi de bu gibi uygulamaların etkinliğini artırabilir.

## Teşekkür

Bu çalışma TÜBİTAK 1512 programı kapsamında 2180355 no'lu proje ile desteklenmiştir. Ayrıca uygulamanın deneme sürümünde oluşturulan kalibrasyon modeline konu olan spektral ve referans analiz verileri TÜBİTAK 2150867 no'lu projeden elde edilmiştir. Desteklerinden ötürü TÜBİTAK'a teşekkür ederiz.

## Kaynakça

- [1] Kahrıman, F., Egesel, C. Ö., *Using near infrared (NIR) spectroscopy in the analysis of cereal products: the example of maize*, in Recent Researches in Science and Landscape Management, Newcastle: Cambridge Scholars Publishing, pp. 507-521, 2018.
- [2] Roggo, Y., ve ark., *A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies*, Journal of Pharmaceutical and Biomedical Analysis vol. 44, pp. 683-700, July 2007.
- [3] Pasquini, C. *Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications*, Journal of the Brazilian Chemical Society vol. 14, pp. 198-219, April 2003.
- [4] Mekonnen, B. K., Yang, W., Hsieh, T-H., Liaw, S-K., Yang, F-L., *Accurate prediction of glucose concentration and identification of major contributing features from hardly*

- distinguishable near-infrared spectroscopy*, Biomed Signal Process Control, vol. 59, 101923, 2020.
- [5] Özdemir, S., *Random Forest Yöntemi kullanılarak potansiyel dağılım modellemesi ve haritalaması: Yukarıgökdedere Yöresi örneği*, Turk J Agric For, 2018, 19(1): 51-56.
- [6] Uğuz, S., *Makine öğrenmesi - teorik yönleri ve python uygulamaları ile bir yapay zeka ekolü*, ISBN: 978-605-033-176-9, 298 s, 2019.
- [7] Zareef, M. ve ark., *An overview on the applications of typical non-linear algorithms coupled with nir spectroscopy in food analysis*, Food Eng Rev vol: 12, pp. 173–190, 2020.
- [8] Grossi, M. ve ark., *An opto-electronic system for in-situ determination of peroxide value and total phenol content in olive oil*, J Food Eng vol.146, pp. 1–7, 2015.
- [9] Massie, D. R. ve Norris, K.H. *Spectral reflectance and transmittance properties of grain in the visible and near infrared*, Transactions of ASAE vol. 8, pp. 598–600, 1965.
- [10] Nakawajana, N. ve Posom, J., *Comparison of analytical ability of pls and svm algorithm in estimation of moisture content, higher heating value, and lower heating value of cassava rhizome ground using FT-NIR spectroscopy*, IOP Conf Ser: Earth Environ. Sci. vol. 301, MN:012032, 2019.
- [11] Onaç, İ. Kahrıman, F. ve Baytekin, H. *Mısırdan nem içeriğinin yakın kızıl ötesi (NIR) spektroskopisi ile tespitinde farklı kemometrik yöntemlerin etkisinin belirlenmesi*, ÇOMÜ Ziraat Fakültesi Dergisi, vol. 4, pp. 117-124, 2017.
- [12] Ihaka, R. ve Gentleman, R. R: *A language for data analysis and graphics*, The Journal of Computational and Graphical Statistics vol. 5, pp. 299-314, 1996.
- [13] Team, R C. R: *A language and environment for statistical computing*, R Found. Stat. Comput. Vienna, 2019, [Çevrimiçi]. Erişim adresi: <https://www.R-project.org/>.
- [14] Chang, W. ve ark., *shiny: Web application framework for R*. 2020.
- [15] Mevik , B-H. ve Wehrens, R., *The pls package: principal component and partial least squares regression in R*, Journal of Statistical Software vol. 18, pp. 1-23, Jan. 2007.
- [16] Kucheryavskiy, S., *mdatools – R package for chemometrics*, Chemom Intell Lab Syst vol. 198, s. 103937, Mar. 2020.
- [17] Kuhn, M. ve ark., *caret: Classification and regression training*. 2020.
- [18] Wehrens, R., *Chemometrics with R: multivariate data analysis in the natural sciences and life sciences*. Berlin Heidelberg: Springer-Verlag, 2011.
- [19] Meyer, D. ve ark., *e1071: Misc functions of the department of statistics, probability theory group* (Formerly: e1071), TU Wien. 2019.
- [20] Stevens, A. ve Ramirez-Lopez, L., *prospectr: miscellaneous functions for processing and sample selection of spectroscopic data*. 2020.
- [21] Koc, A., Henriksson T. ve Chawade A., *Specalzyer—an interactive online tool to analyze spectral reflectance measurements*, PeerJ, vol. 6, e5031, June, 2018.
- [22] Xie Y. ve ark., *DT: A wrapper of the javascript library “DataTables”*. 2020.
- [23] Kahrıman F., ve ark., *Mısır tanesinde gıda ve endüstriyel kullanım için önem taşıyan özelliklerin analizinde kullanılacak NIRS (yakın kızıl ötesi spektroskopisi) kalibrasyon modellerinin geliştirilmesi* (proje no: 2150867), TÜBİTAK Projesi Sonuç Raporu, 2017.
- [24] Williams P.C. ve Sobering D.C., *How do we do it: a brief summary of the methods we use in developing near infrared calibrations*. In: Davies AMC, Williams PC, editors. *Near Infrared Spectroscopy: The Future Waves*. Chichester, UK: NIR Publications, pp. 185–188, 1996.
- [25] Galvez–Sola L. ve ark., *The potential of near infrared reflectance spectroscopy (NIRS) for the estimation of agroindustrial compost quality*. Sci Total Environ vol: 408, pp. 1414–1421, 2010.