BAHÇE 48(2): 79–85 (2019)
ISSN 1300–8943

BİTKİ ISLAHINDA GENOTİP VERİM DEĞERİNİN REGRESYON YÖNTEMLERİ İLE TAHMİNİ

Bengü KOYUNCU[1]*, Murat GÖK[2]

[1]Tarla Bitkileri Merkez Araştırma Enstitüsü Müdürlüğü, Tarla Bitkileri Bölümü, Ankara; ORCID: 0000-0003-0613-0802
[2] Doç. Dr., Yalova Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Yalova; ORCID: 0000-0003-2261-9288
Geliş Tarihi / Received: 10.06.2019 *Kabul Tarihi / Accepted: 16.07.2019*

ÖZ

Bilimsel araştırmaların amacı yapılan çalışmaların gözlem ve denemelerinden genel sonuçlara ulaşmaktır. Gelişen teknolojilerle beraber bu sonuçlar dijital olarak kayıt altına alınmakta ve bu kayıtlar büyük veri (big data) yığınlarını meydana getirmektedir. Bu yığınların işlenmesi yani anlamlı bilgiye dönüştürülmesi 1950'li yıllarda başlamış ve veri madenciliği kavramı ortaya çıkmıştır. Tahmin ya da karar verme süreçlerinde kullanılan veri madenciliği, günümüzde tarımsal faaliyetlerin tahmin çalışmalarında da kendine yer bulmaktadır. Bitki ıslah çalışmalarının temeli, istenilen fenotip ve genotip özelliklerinin verim ve çevre şartlarına göre karşılaştırılması esasına dayanmaktadır. Bu sonuçların değerlendirilmesinde çeşitli istatistik paket programları kullanılmaktadır. Kullanılan bu programlar bir ıslahçının verim ile yapılacak genotip seçimleri için gerekli analiz ve raporlama kabiliyetlerini tam olarak karşılamamakladır. Bu çalışmada, 12 lokasyondan, 24 genotipe ait 4 tekerrürlü toplam 1153 adet verim değerine göre genotipe ait verim tahmini yapılmıştır. Tahminlemede bitki ıslahında kullanılan doğrusal regresyonun yanında makine öğrenmesi metotlarından Sıralı Minimal Optimizasyon (SMO), En Yakın k–Komşu (k–EYK), Rastgele Orman (RO) metotları seçilmiştir. Seçilen metotların başarıları Ortalama Karesel Hatanın Karekökü ve Ortalama Mutlak Hata metriklerine göre karşılaştırılmıştır. RO, diğer üç yönteme göre daha yüksek performans göstermiş ve bitki ıslah programlarında kullanılan doğrusal regresyon yöntemi ile beraber kullanılması önerilmiştir.

Anahtar Kelimeler: Regresyon analizi, istatistik, ıslah, sıralı minimal optimizasyon regresyon, doğrusal regresyon, rastgele orman

REGRESSION METHODS ESTIMATION OF GENOTYPE YIELD VALUE IN PLANT BREEDING

ABSTRACT

The aim of scientific research is to reach general results from the observations and experiments of the studies. Together with the developing technologies, these results are recorded digitally, and these records form big data stacks. The process of processing these masses into meaningful information began in the 1950s and the concept of data mining emerged. Data mining, which is used in forecasting or decision making processes, now finds its place in forecasting agricultural activities. The basis of the plant breeding studies is based on the comparison of the desired phenotype and genotype properties according to the efficiency and environmental conditions. Various statistical package programs are used in the evaluation of these results. These programs do not fully meet the analysis and reporting capabilities required for a breeder's genotype selection. In this study, the yield of the genotype was estimated from 12 locations with a total of 1153 yields of 24 replicates of 24 genotypes. In addition to the linear regression used in plant breeding, sequential Minimal Optimization (SMO), Nearest k–Neighbor (k–EYK), Random Forest (RO) methods were selected from the methods of machine learning. The success of selected methods was compared according to the mean square root of the mean square error and the mean absolute error metrics. RO has shown higher performance than the other three methods and it has been proposed to use with the linear regression method used in plant breeding programs.

Keywords: Regression analysis, statistics, breeding, sequential minimal optimization regression, linear regression, random forest

*Sorumlu yazar / *Corresponding author*: bengu.esmer@tarimorman.gov.tr

GİRİŞ

Bitkilerin genetik yapılarındaki ve doğal yayılış alanlarındaki çeşitlilikten yararlanılarak kalıtım yoluyla aranan niteliklere sahip yeni bitkiler elde edilmesine 'bitki ıslahı' denir [26].

Bitki ıslahı çalışmalarında amaç, eldeki çeşit ve hatların yer aldığı melezleme programlarından daha verimli ve kaliteli, yeni çeşitlerin oluşturulmasıdır. Bu amaç doğrultusunda oluşturulan genotiplerin başka çevre koşullarında da aynı verimi ya da benzer verimi vermesi yani stabil olması beklenmektedir [23]. Bu bakımdan farklı lokalitelerden alınan genotipe ait verim değerleri çeşitli istatistik metotları vasıtasıyla değerlendirilmekte ve genotip için verim tahminleri yapılmaktadır. Kıran [10]'ın, 1995–1998 yılları arasında Ege bölgesindeki arpa çeşit adaylarının değerlendirilmesinde çeşitli istatistik parametreleri kullanmış ve çalışmasının sonunda verim tahmini için varyasyon formülünü belirtmiştir. Özberk ve ark. [17], Haran ovasında ekmeklik buğday çalışmasında çeşitli verim öğeleri ile tane verimi arasında istatistiki önemde bulunan regresyon eşitliklerini belirtmiştir. Özberk [16]'in, 1982/1983–1998/1999 yılları arasında Diyarbakır 81 ve Harran–95'in regresyon analizi bulgularını araştırmasının sonuna eklemiştir. Sabancı [21]'nın çalışmasında, bitki ıslah çalışmalarında stabilitenin belirlenmesinde regresyon katsayılarının kullanılması birçok araştırıcı tarafından benimsendiğini belirtmiştir. Bu ve benzeri, ülkemiz bitki ıslah çalışmalarında genotipe ait verim tahminlerinde sıklıkla regresyon analizi ile yapıldığı görülmüştür.

Islah programlarında yer alan genotiplerin farklı çevrelerde verimlerinin denenmesi ve bu bilgilerden anlamlı sınıflandırmalar ya da tahminler yapılması günümüz teknolojileri arasında yer alan makine öğrenmesi yöntemleriyle de mümkün olmaktadır. Makine öğrenmesi, matematiksel ve istatistiksel yöntemler kullanılarak, elde bulunan verilerden sonuçlar çıkaran, bu sonuçlarla bilinmeyene dair tahminlerde bulunan değerler dizisidir [4].

Makine öğrenmesinde, gözetimli öğrenme başlığı altında regresyon yöntemleri bulunur. Regresyon analizlerinde her gözlem için öğrenilen değerlere bakarak reel bir tahminleme yapılması söz konusudur. Yukarıda da örneklerle belirtilen ülkemiz ıslah çalışmalarında doğrusal regresyon yöntemi dışında verime yönelik bir tahminleme çalışmasına rastlanmamıştır. Bu bakımdan çalışmamızda makine öğrenmesi yöntemleri içinde yer alan ve doğrusal regresyonun dışında başka regresyon yöntemleri kullanılarak, genotipe ait verim tahminleri yapılmıştır ve modelleme başarı oranı en iyi sonucu veren metot, literatüre sunulmuştur.

Bu araştırmanın amacı, bitki ıslahında yer alan genotip verimine yönelik tahminleme çalışmalarına, doğrusal regresyon yöntemin dışında, regresyona dayalı başka makine öğrenmesi metotları ile denenmesi ve sonuçların bitki ıslahçılarıyla paylaşılmasıdır.

MATERYAL VE METOT

Materyal

Çalışmanın ana materyalini, 2018 yılına ait arpa ıslah programında kullanılan, 12 farklı lokasyondan 4 tekerrürlü olmak üzere 24 genotipe ait 1152 adet verim değerleri ile her lokasyonun ortalama verim bilgilerinden oluşmaktadır.

Veri madenciliğinde verilerin değerlendirmeye almadan önce veriler üzerinde ön işleme aşamalarından öznitelik çıkarımı yapılarak metinsel veriler sayısallaştırılmaktadır [9]. Bu nedenle her bir lokasyon 1–12, genotip 1–24 arasında sayısal kodlar verilerek veri seti hazırlanmıştır.

Veri setimize ait ilk dört örnek Çizelge 1'de verilmiştir. Çizelge 1'de bulunan, Lokasyon farklı çevre koşullarını temsil etmekte, Genotip kullanılan çeşit adaylarını göstermekte, Verim başlığı çeşit adaylarının aynı satırda yer alan çevredeki verimini ve VerimOrt aynı satırdaki çevreye ait verim ortalamasıdır.

Çizelge 1. Metinsel verilerin, sayısal verilere dönüştürme işlemi ile edilen sonuçlar

Table 1. Results of converting textual data into numerical data

Lokasyon *Location*	Gonotip *Genotype*	Verim *Yield*	VerimOrt *YieldAvg*
1	1	426.3	337.5
1	2	400.0	337.5
1	3	516.7	337.5
1	4	492.8	337.5

Metot

Bitki ıslah programlarında amaç, mevcut popülasyonlardan, çeşitli varyasyonlarla istenilen özellikte yeni bireyler elde etmektir. Yapılan bu çalışmalarda ıslahçı birçok veriye sahip olmakta ve yapmış olduğu gözlemsel denemeleri ayrıca istatistiksel olarak da kanıtlamak istemektedir. Yani elde bulunan yoğun verilerden, anlamlı sonuçlar elde etmek istemektedir. Bu sonuçlar için tanıtıcı istatistikteki bir takım analizlerden yararlanmaktadır. Yoğun verilerden anlamlı sonuçlar çıkartmak için makine öğrenmesi yöntemleri de kullanılabilir. Makine öğrenmesi, 1959 yılında bilgisayar biliminin yapay zekâda sayısal öğrenme ve model tanıma çalışmalarından geliştirilmiş olan bir sistemdir [18]. Bu sistem işlev olarak elde bulunan yığın bilgiler vasıtasıyla öğrenebilen ve veriler üzerinde tahmin yapabilen algoritmalar bütünüdür. Sistem içindeki algoritmaların bir kısmı tahmin ve kestirim (regresyon) bir kısmı da sınıflandırma yapabilme yeteneğine sahiptir [13]. Makine öğrenmesi gözetimli ve gözetimsiz öğrenme olmak üzere ikiye ayrılmaktadır [4].

Gözetimsiz Öğrenme; öğrenme süreci etiketsiz gözlemlerden oluşmaktadır. Kullanılan algoritmaların kendi kendilerine keşifler yapması, belli olmayan örüntülerin keşfedilmesi beklenir. Kümeleme ve boyut azaltımı alt sınıflarına ayrılarak incelenmektedir.

Gözetimli Öğrenme; sınıflandırma ve regresyon olarak iki alt başlıkta incelenen bu öğrenme metodu, öğrenme süreci için, etiketlenmiş gözlemleri kullanır, yani her girdiye ait çıkışın ne olduğu önceden sisteme verilir ve bu yolla öğrenme gerçekleştirilir. Aitkenhead [2] yabancı otların mahsullerden ayrılmasında, Shahin, Tollner ve Gitaitis [24]'in hastalıklı soğanların belirlenmesinde, Varjovi ve Talu [27]'nun, kayısı toplam ürün tahmininde, Acar ve Özerdem [1]'in, alan imgelerinin ekinin ürün gelişimine göre sınıflandırılmasında ve benzeri sınıflandırma, tahmin etme gibi sonuçları içeren tarımsal faaliyetlerde de gözetimli öğrenme metotlarından faydalanıldığı görülmüştür.

Bitki ıslahı çalışmalarında genotipe ait verimin, farklı çevre koşullarından etkilenip etkilenmediğini görmek için stabilite kavramı içerisinde doğrusal regresyon analizinden faydalanılmaktadır. Bu nedenle, çalışmamızda makine öğrenmesi yöntemlerinden regresyon analizlerini içeren algoritmalar seçilmiş ve bu yöntemler detaylandırılmıştır.

Regresyon yöntemlerini içeren makine öğrenmesi yöntemleri

Regresyon, bir yanıt değişkeni ile bu yanıt değişkenine bağlı olabilecek açıklayıcı değişkenlerin bir fonksiyonu olarak ifade edilebilmektedir [14]. İki ya da daha fazla değişken arasındaki doğrusal ilişkinin fonksiyon olarak gösterimi olarak ta ifade edilmektedir. Regresyonda tayin edilmesi gereken önemli konulardan birisi iki değişken arasındaki (x ve y) ilişkinin ne şekilde olduğudur [15]. Bu ilişki doğrusal, quatratik veya kübik olabilir [3].

Bitki ıslahında, en yüksek verim değerine sahip genotipin, durağan (stabil) olanını bulmak için kullanılan basit regresyon analizi, yanıt değişkeni ile tek bir açıklayıcı değişken arasındaki doğrusal ilişkiyi açıklar[20]. Eğer tek bir yanıt değişkeni ve birden fazla açıklayıcı değişken arasındaki doğrusal veya eğrisel bir ilişki tanımlanmak istenirse, ilişki çoklu doğrusal regresyon analizi ile incelenir [15, 28].

Basit doğrusal regresyonda, Y yanıt değişkeni, X_1 açıklayıcı değişkeni, β_0 ve β_1 bu değişkenin bilinmeyen parametrelerini ve $i\varepsilon$ şansa bağlı hata terimlerini ifade etmek üzere basit doğrusal regresyon modeli;

$Y = \beta_0 + \beta_1 X_1 i + i\varepsilon \; i = 1, 2, \ldots , n$

$$b_1 = \frac{\sum(xi - \bar{x})(yi - \bar{y})}{\sum(xi - \bar{x})}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

olarak yazılabilmektedir [11].

Gözetimli öğrenme metotları içinde regresyon analizine dayalı algoritmaya sahip bir diğer yöntem SMO'dur. SMO, gözetimli bir makine öğrenme algoritmasına sahip olan Destek Vektör Manikalarının (DVM) kuadratik programlama sorunlarını, fazladan matris depolaması olmadan ve her bir alt problem için tekrarlı sayısal bir tekrara başvurmadan hızlı bir şekilde çözen basit bir algoritma olarak tanımlanmaktadır [19]. Bu yöntem bütün kayıp değerlerini yenisi ile değiştirir ve nominal olan öznitelikleri iç ve dış döngü olmak üzere iki parçaya ayırır. Dış

döngüde en iyi seçim yapılırken, iç döngüde seçilen değerlere göre iki adet Langrage çarpanı hesaplanır ve tüm örnekler istenilen düzeye gelinceye kadar bu döngü çalıştırılır [7]. SMO sınıflandırma problemlerinin dışında regresyon işlemlerinde de kullanılabilmektedir [22].

Regresyon yöntemlerinde yukarıda anlatılan tanımlarda da belirtildiği gibi giriş ve çıkış parametreleri yer almaktadır. Çalışmamız için seçmiş olduğumuz SMO yönteminde, üç adet giriş (lokasyon, genotip, lokasyona ait verim ortalaması) kullanılarak bir adet çıkış parametresi (verim) değerinin tahmin edilmesi amaçlanmıştır. Doğrusal regresyon analizinde ise tek bir giriş (lokasyonların verim ortalaması) kullanılarak genotipe ait verim tahminlemesi yapılmıştır. Çizelge 2'de çalışmamıza ait verilen tahmin modeli ve formülleri gösterilmektedir.

RO, hem sınıflandırma hem de regresyon analizinde kullanılmaktadır. Aşırı öğrenme problemi olmayan bu metot, ilk olarak birden fazla sınıflandırma ağacı oluşturulur ve her ağaca sınıflandırma yapması için giriş vektörü verilir. Her ağacın sonucu arasında en yüksek oyu alan sınıf sonucu seçilir. Her ağaç, eğitim setinde bulunan örneklerden rasgele seçilerek yenisiyle değiştirilmesiyle oluşturulur. Her ağaç eğitim seti oluşturulurken verilerin üçte biri ağaç dışına ayrılır, geri kalanı ile sınıflandırma hatası hesap edilir [12]. Çalışmamızın verileri RO yöntemi ile değerlendirilirken, doğrusal regresyon yönteminde olduğu gibi tek bir giriş (lokasyonların verim ortalaması) kullanılarak genotipe ait verim tahminlemesi yapılmıştır.

k–EYK, parametrik olmayan bir sınıflandırma ve regresyon yöntemidir. Bir örneğin k adet seçilen eğitim uzayındaki sınıfları belli olan noktalara olan uzaklıklarına göre sınıflandırma işlemi yapmaktadır [6]. Uzaklık hesapları Minkowski, Öklid, Manhattan olmak üzere üç grupta incelenmektedir. Bu çalışmada Öklid uzaklık ölçüsü kullanılmış ve k=10 adet en yakın komşuya göre analiz yapılmıştır, lokasyonlara ait verim ortalamaları giriş etiketi olarak sistemin öğrenimi gerçekleştirilmiş ve genotipe ait verim tahmini yapılması istenmiştir.

Çizelge 2. Doğrusal ve SMO regresyon analizi formül sonuçları

Table 2. Linear and sequential minimal optimization regression analysis formula results

Kullanılan metot *Method used*	Formül *Formula*
Doğrusal regresyon *Linear regression*	Verim = 0.9999 × OrtVerim + 0.0596
SMO	Verim = –0.0196 × lokasyon – 0.0783 × genotip + 0.3628 × OrtVerim + 0.2577

BULGULAR VE TARTIŞMA

Başarım Metrikleri

Çalışmamızın seçilen model üzerindeki başarımları, Ortalama Karesel Hatanın Karekökü (Root Mean Squared Error–OKHK) ve Ortalama Mutlak Hata (Mean Absolute Error–OMH) metrikleri ile değerlendirilmiştir.

OKHK, bir model tarafından tahmin ettiği değerler ile gerçek değerleri arasındaki mesafeyi ölçen kuadratik bir metriktir ve bu tip farklılıkları ölçmek için sıkça kullanılır [25].

$$\mathrm{OKHK} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{E}_{\mathrm{tahmin}} - \mathrm{E}_{\mathrm{gerçek}})^2}$$

OMH, sıfıra yakınlığının modelin tahmin yeteneği için iyi olduğu düşünülen bu hata metriğinin ölçüm değerleri ile model tahminleri olan sürekli iki değişken arasındaki farkın ölçütüdür [8].

$$\mathrm{OMH} = \frac{1}{\mathrm{n}}\sum_{\mathrm{j}=1}^{\mathrm{n}}\left|\mathrm{e_j}\right|$$

Deneysel Sonuçlar

Çalışmamızda, 12 lokasyondan, 24 genotipe ait 4 tekerrürlü toplam 1153 adet verim değerine göre genotipe ait verim tahmini, makine öğrenmesi yöntemleri içinde regresyona dayalı metotlar kullanılarak yapılmıştır. Başarım metrikleri ise OKHK, OMH olarak seçilmiştir. Buna göre, doğrusal regresyon analizinde 8.60 OKHK, 10.82 OMH, SMO analizinde 10.67 OKHK, 8.34 OMH, k–EYK analizinde 7.73 OKHK, 6.00 OMH, RO analizinde ise 7.61 OKHK, 5.89 OMH değerleri bulunmuştur. Bu değerler toplu olarak Çizelge 3'te verilmiştir.

Çalışmamızda sadece SMO yönteminde giriş etiketi olarak, veri setimizde bulunan 3 öznitelik (genotip, lokasyon, lokasyona ait ortalama verim) kullanılırken, diğer yöntemlerimizde sadece verim kullanılmıştır.

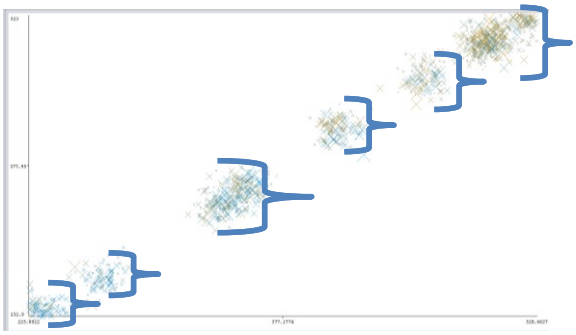
Sonuçların tahminlemesinde, SMO ve Doğrusal Regresyon için formül (Çizelge 1), RO için 23 alt dala ayrılmış bir ağaç yapısı (Çizelge 4) ve k–EYK yönteminde ise birbirine yakın 6 farklı küme sonucu (Şekil 1) söz konusu olmuştur.

Aslında kümeleme algoritmasına sahip olan k–EYK, regresyona dayalı hesaplarda da kullanıldığı için araştırmamıza eklenmiştir. k–EYK sonuçları; verim ortalama değeri 225–232 arasında ise, tahmini değer 226.7, 270–270.3 arasında ise 260.23, 337.5–358.2 arasında ise 339.86–358.26, 410.9 ise 403.21, 461.2 ise 463.18 ve son olarak 496–529 arasında ise 463.18 ile 498.56 arasında olmuştur. Ancak Şekil 1'de de görüldüğü üzere veriler 6 küme üzerinde ayrılmış ve tam anlamıyla ıslahçı için kullanışlı bir tahminleme sonucuna varılamamıştır.

Çizelge 3. Uygulanan metotlar ve hata metrik değerleri

Table 3. Applied methods and error metric values

Kullanılan metot *Method used*	Ortalama karesel hatanın karekökü *Root mean squared error*	Ortalama mutlak hata *Mean absolute error*
Doğrusal regresyon *Linear regression*	8.60	10.82
SMO	10.67	8.34
k–EYK	7.73	6.00
RO	7.61	5.89



Şekil 1. k–EYK kümeleme yapısı

Figure 1. k–EYK clustering structure

RO analizine ilişkin ayrıntılı bilgi Çizelge 4'te verilmiş olup, tahmin için giriş değerleri 384.55'ten küçükse çıkış değerlerinin 227.75 ile 375.9 arasında olduğu, büyükse 422.95 ile 531.26 arasında değerler aldığı gözlenmiştir. Ayrıca genotipe ait verim tahmini modeller üzerindeki hata payı bu analizsin sonucundan elde edilmiştir. Bu sonuçlar doğrultusunda genellikle bitki ıslah programlarında verim tahmin çalışmalarında kullanılan doğrusal regresyon yöntemi ile daha yüksek doğruluk sağlayan RO yöntemi beraber kullanılması önerilmektedir.

Elde edilen hata oranlarının daha çok düşürülebilmesi için bitki ıslah verim tahmin çalışmalarına başka öznitelikler (bitkiye ait olgunlaşma süresi, bitki boyu, başaklanma süresi vb.) eklenerek tekrar yapılması ve sonuçlarının tekrar karşılaştırılması gerektiği ortaya çıkmıştır.

Çizelge 4. RO ağaç yapısı

Table 4. RO tree structure

```
OrtVerim < 384.55
|   OrtVerim < 303.9
|   |   OrtVerim < 251.6: 227.75
|   |   OrtVerim >= 251.6: 275.99
|   OrtVerim >= 303.9
|   |   OrtVerim < 352.45
|   |   |   OrtVerim < 342.1: 356.96
|   |   |   OrtVerim >= 342.1: 336.07
|   |   OrtVerim >= 352.45: 375.9
OrtVerim >= 384.55
|   OrtVerim < 476.85
|   |   OrtVerim < 436.05: 422.95
|   |   OrtVerim >= 436.05: 458.47
|   OrtVerim >= 476.85
|   |   OrtVerim < 512.75
|   |   |   OrtVerim < 501.4
|   |   |   |   OrtVerim < 494.4: 489.64
|   |   |   |   OrtVerim >= 494.4
|   |   |   |   |   OrtVerim < 498.3: 499.67
|   |   |   |   |   OrtVerim >= 498.3: 495.78
|   |   |   OrtVerim >= 501.4: 510.89
|   |   OrtVerim >= 512.75: 531.26
```

SONUÇ

Günümüzde ticaret, eğitim, mühendislik, borsa, bankacılık, tıp ve telekomünikasyon gibi birçok alanda başarılı uygulamaları olan makine öğrenmesinin tarım alanında da uygulanması ve yeni çıkarımlar ortaya çıkarması kaçınılmazdır. Bitki ıslahı bilimi, toplumun gıda ihtiyacını karşılamak için önemli bir disiplindir. Bu çalışmada 2018 yılına ait arpa ıslah programında kullanılan, 24

genotipte 12 farklı lokasyondan 4 tekerrürlü olmak üzere toplanan genotipe ait verim bilgileri, makine öğrenmesi yöntemleri içinde yer alan regresyona dayalı algoritmalar kullanılarak genotip verim tahmini yapılmıştır. Hesaplanan hata metrik sonuçlarına göre, RO'nun doğrusal regresyondan daha yüksek performans gösterdiği görülmüştür. Yapılan literatür çalışmalarında, bitki ıslahında genotip veriminin tahminine yönelik doğrusal regresyon dışında başka makine öğrenmesi yöntemine rastlanmamıştır. Bu nedenle bitki ıslah programlarında kullanılan doğrusal regresyon yöntemi ile RO analizinin birlikte kullanılması önerilmektedir. İlerideki çalışmalarda, bitkiye ait farklı öznitelikler ekleme ve başka makine öğrenmesi algoritmaları yoluyla tahmin performansını artırmayı hedeflemekteyiz. Ayrıca çevrimiçi milli bir ıslah istatistik web uygulaması geliştirmeyi ve bu uygulama ile ıslah çalışmalarında kullanılan analizleri tek bir çatıda toplamayı amaçlamaktayız.

KAYNAKLAR

1. Acar, E., Özerdem, M., 2016. Kızıltepe tarımsal Kızıltepe tarımsal alan imgelerinin ekinin ürün gelişimine göre sınıflandırılması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi 5(1).* (Retrieved from; http://dergipark.org.tr/tbbmd/issue/22244/238793).
2. Aitkenhead, M.J., Dalgetty, I.A., Mullins, C.E., McDonald, A.J.S. and Strachan, N.J.S., 2003. Weed and crop discrimination using image analysis and artificial intelligence methods. *Computers and Electronics in Agriculture 39(3):157-171.*
3. Alparslan, M.A., İnal, A., 2005. Deneme tekniği ders kitabı. *Ankara Üniversitesi Ziraat Fakültesi Yayınları: 496, 435s.*
4. Baykal, C., Baykal, A., 2011. Veri madenciliğinde sınıflandırma algoritmalarının bir örnek üzerinde karşılaştırılması (http://ab.org.tr/ab11/bildiri/67.pdf; Erişim Tarihi: 18.01.2019).
5. Bishop, Christopher, M., 2006. Pattern recognition and machine learning *Springer*.
6. Cedeño, W., Agrafiotis, D.K., 2003. Using particle swarms for the development of QSAR models based on K-nearest neighbor and kernel regression. *Journal of Computer Aided Molecular Design 17(2-4):255-263.*
7. Çakır, A., Çalış, H., Küçüksille, E., 2009. Data mining approach for supply unbalance detection in ınduction motor. *Expert Systems with Applications 36(9):11808-13.*
8. Eren, B., Eyüpoğlu, V., 2011. Yapay sinir ağları ile Ni(II) iyonu geri kazanım veriminin modellenmesi. *6. International Advanced Technologies Symposium (IATS'11), 16-18 Mayıs 2011. Elazığ.*
9. Guyon, I., Gunn, S., Nikravesh, M., Zadeh, L. A., (Eds.) 2008. Feature extraction: foundations and applications. *Springer Press Verlag, Berlin Heilderberg. pp:775.*
10. Kıran, A., 1999. Ege Bölgesinde geliştirilen iki sıralı arpa (*Hordeum vulgare* L.) çeşit adaylarının bazı istatistik parametrelerce değerlendirilmesi. *Anadolu Journal of AARI 9(1):141-148.*
11. Kutner, M.H., Nachtsheim, C.J., Neter, J., Li, W., 2005. Applied linear statistical models. *McGraw-Hill Irwin Companies Inc. New York.*
12. Liaw, A., Wiener, M., 2002. Classification and regression by random forest. *R News 2(3):18-22.*
13. Michie, D., Spiegelhalter, D.J., Taylor, C.C., 1994. Machine learning, neural and statistical classification. *Ellis Horwood Ltd.*
14. Montgomery, D.C., Peck, E.A., Vining, G.G., 2012. Introduction to linear regression analysis. *John Wiley & Sons, Vol:821.*
15. Okur, S., 2009. Parametrik ve parametrik olmayan doğrusal regresyon analiz yöntemlerinin karşılaştırmalı olarak incelenmesi (Yüksek Lisans Tezi). *Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Adana.*
16. Özberk, İ., 2004. Güneydoğu Anadolu koşullarında bazı makarnalık buğday çeşitlerinin verim stabilitesi. *Atatürk Üniversitesi Ziraat Fakültesi Dergisi 35(1-2):11-19.*
17. Özberk, İ., Özberk, F., Öktem, A., 2002. Harran ovası koşullarında ekmeklik buğday bölge verim denemelerinde bazı istatistik analizler. *Çukurova Üniversitesi Ziraat Fakültesi Dergisi 17(3):111-118.*
18. Özekes, S., 2003. Veri madenciliği modelleri ve uygulama alanları. *İstanbul Ticaret Üniversitesi Dergisi 3:65-82.*

19. Platt, J., 1998. Fast training of support vector machines using sequential minimal optimization. *Microsoft Research Technical Report MSR-TR 98:14.*
20. Ronningen, V.V., 1986. Selection index theory with practical applications. *Chapter 10.*
21. Sabancı, C., 1997. Stabilite analizlerinde kullanılan yöntemler ve stabilite parametreleri. *Anadolu Dergisi 7(1):75-90.*
22. Senel, B., Caglar, M.F., Senel, F.A., 2016. Behavioral modeling of 2.4 GHz RF power amplifier using data mining. *2. International Conference on Engineering and Natural Sciences (ICENS 2016), At Sarajevo 10.*
23. Seyfi, T., Çeri, S., Kaya, Y., Akçura, M., Ayrancı, R., Özer, E., 2004. Bazı ekmeklik buğday (*T. aestivum* L.) genotiplerinin Orta Anadolu Bölgesi kuru koşullarında dane verimi stabilitesi. *Bitkisel Araştırma Dergisi 2:21-26.*
24. Shahin, M.A., Tollner, E.W., Gitaitis, R.D., 2002. Classification of sweet onions based on internal defects using image processing and neural network techniques. *Transactions of the ASAE 45(5):1613-1618.*
25. Sheta, A., Rine, D., Ayesh, A., 2008. Development of software effort and schedule estimation models using soft computing techniques. *IEEE Congress on Evolutionary Computation, pp:1283-1289.*
26. Vardar, K.Ç., Aras, S., Cansaran D.D., 2010. Bitki ıslahında moleküler belirteçlerin kullanımı ve gen aktarımı. *Türk Hijyen ve Deneysel Biyoloji Dergisi 67(1):33-43.*
27. Varjovi, M.H., Talu, M.F., 2016. Kayısı için otomatik rekolte tahmin sistemi. *International Conference on Artificial Intelligence and Data Processing, IDAP, 2016.*
28. Weisberg, S., 2005. Applied linear regression. *John Wiley & Sons, Inc. Hoboken, New Jersey.*