



Okuma Becerilerini Yordayan Özelliklerin Belirlenmesi: Genetik Algoritma Kestirimi*

İzzettin AYDOĞAN**
Selahattin GELBAL***

Öz

Bu araştırmayla öğrencilerin okuma becerilerini yordayan özelliklerin belirlenmesi amaçlanmıştır. Araştırmanın çalışma grubunu, PISA 2015 uygulamasına katılan 42 farklı ülkeden 5232 on beş yaş grubu öğrenci oluşturmuştur. Araştırma verileri, PISA 2015 programı verileri üzerinden sağlanmış olup, genetik algoritmalar yöntemi kestirimine dayalı regresyon modeli esasıyla analiz edilmiştir. Genetik algoritmalar yöntemi ile okuma becerilerini en iyi derecede yordayan değişkenlerden oluşan regresyon modeli için değişken seçim işlemi yapmak istenmiştir. Elde edilen sonuçlara göre, cinsiyet, baba eğitim durumu, evde internet kullanımı, evde konuşulan dil, sahip olunan e-kitap okuyucu sayısı, okuma becerisini ölçen maddeleri yanıtlama hızı ve evdeki kitap çeşitliliği ve sayısı değişkenlerinin öğrencilerin okuma becerilerini istatistiksel olarak anlamlı düzeyde yordadığı saptanmıştır. Yordama düzeyi anlamlı bulunan değişkenlerdeki farklılaşmanın öğrencilerin okuma becerilerinde de anlamlı düzeyde farklılaşmaya yol açtığı anlaşılmıştır.

Anahtar Kelimeler: Okuma becerileri, yordayıcı özellikler, regresyon, genetik algoritmalar

Determination of the Features Predicting Reading Skills: Genetic Algorithms Estimation

Abstract

The current study aimed to determine the features predicting students' reading skills. The study group of the study was comprised of 5232 students aged 15 years from 42 countries participating in the PISA 2015 application. The data of the study were obtained over the PISA 2015 program and analyzed by using the regression model based on the genetic algorithms method estimation. It was intended to perform a feature selection process for the regression model, which consisted of the variables that best predicted reading skills with the method of genetic algorithms. The results obtained revealed that the variables of gender, father's education level, use of the internet at home, the language used at home, the number of e-book readers, the speed at which the items measuring the reading skill was responded and the number and variety of the books at home significantly predicted the students' reading skills. The variation in the variables whose prediction power was found to be significant was understood to lead to a significant variation in reading skills.

Keywords: Reading skills, predictor features, regression, genetic algorithms

Giriş

Modern dünyaya ayak uydurmak için gerekli olan öğrenmelere metinsel öğeleri okuyarak ulaşılabileceği uzun zamandan beri kabul edilen bir durumdur (Kintsch, 1998). İnsanlar, gerek okul ve okul dışı ortamlardaki eğitim süreçlerinde; gerekse iş hayatlarında ihtiyaç duydukları bilgi ve

* Bu çalışma, birinci yazarın ikinci yazar danışmanlığında hazırladığı doktora tezinden üretilmiştir.

** Doktora Öğrencisi, Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitim Bilimleri Bölümü, Ankara, ayd.izzet@gmail.com, ORCID: 0000-0002-5908-1285

*** Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara, gelbal@hacettepe.edu.tr, ORCID: 0000-0001-5181-7262

birikimleri okuma sayesinde elde etmektedirler (Yalçın, 2018). Okuma davranışı, ses ve harf arasındaki bağlantının kurulmasıyla başlayan, harflerin birleşmesiyle oluşturulan hecelerin ve bu hecelerin bir araya gelerek kelimeleri oluşturmasıyla devam eden ve kelimeler arasındaki bütünlüğün ifade ettiği cümle yapılarının elde edilmesiyle sonuçlanan bir süreçtir (Adams, 1990, s. 109-110). Grabe ve Stoller (2011, s. 3) okumayı, yazınsal bir ifadeden anlam oluşturma ve bu anlamları uygun şekilde yorumlama yeteneği olarak ifade etmektedir.

Okuma becerileri; bilgiye ulaşma ve bilgiyi geliştirmenin yanı sıra, bireylerin akademik, sosyal ve bireysel gelişimlerini yakından etkileyen bir unsur olarak karşımıza çıkmaktadır (Temizyürek, Çolakoğlu ve Coşkun, 2013; Yıldız ve Akyol, 2011). Coşkun (2002) okumayı; görme, algılama, yorumlama, değerlendirme, analiz etme gibi bir takım üst zihinsel süreçlerin bütünlüştüğü, bireylerin kültürlenmesi, çevresini tanıması ve anlaması, bilgi yüklenmesinde rol oynayan bir araç ve dil becerisi olarak ifade etmektedir. Okuma bireylere; kelimelerin anlaşılması ve tanınması, kelime dağarcığının genişlemesi, anlamlandırma, değerlendirme ve eleştirel düşünme yeteneğinin gelişmesi, farklı görüşler geliştirme, yorum ve sorgulama yeteneğinin gelişmesi gibi özellikler kazandırmaktadır (Sever, 1995, s. 18). Bunun yanında, okuma bireyler için önemli bir yaşam becerisi olarak tasvir edilmektedir. Bu becerileri gelişmeyen bireylerin toplumsal olaylara anlam vermesi, olaylar arasındaki kurguyu anlaması, sorgulama yetisine sahip olması, problem durumlarına karşı çözüm yolları üretmesi beklenemez (Kutlu, 2004). Benzer şekilde, okuma ve anlama becerisi gelişmeyen bireylerden akademik düzeyde başarı beklemek yersiz olacaktır (Özçelik, 1987, s. 102). Gelişmişliğin göstergesi ve dinamiği olarak kabul edilen bilginin işlenmesi, okuma alışkanlığına bağlı bir süreç olarak kabul görmektedir (Özdemir, 1993). Yakın geçmişe kadar toplumların gelişmişlik düzeyinde ölçüt okur-yazar oranı iken; günümüzde bunun yerini okuduğunu anlama olgusu almıştır. PIRLS (Progress in International Reading Literacy Study), PISA (Programme for International Students Assessment) gibi uluslararası standartlarda yapılan büyük ölçekli sınavlarda okuduğunu anlama ve anlama dayalı çıkarım yapma becerilerinin ölçülmesine yer verilmesi bu duruma kanıt olarak gösterilebilmektedir (Kurnaz ve Yıldız, 2015).

Okuma becerileri ile ilişkili özelliklerin belirlenmesi amacıyla yapılan bilimsel çalışmalar incelendiğinde, Adaba'nın (2016) çalışmasında okumaya ayrılan süre ve evde bulunan okuma kaynaklarının; Giambona ve Porcu'nun (2015) çalışmalarında evde bulunan kitap sayısı, cinsiyet, internet kullanımı ve ebeveynlerin eğitim düzeyinin; Harju-Luukkainen, Vettenrante, Ouakrim-Soivio ve Bernelius'un (2016) çalışmalarında evde konuşulan dilin; Kurnaz ve Yıldız'ın (2015) çalışmalarında cinsiyet, sınıf düzeyleri, okudukları kitap sayısı, ebeveynlerin okuma alışkanlığı ve internet kullanımının; Pitoyo'nun (2020) çalışmasında okumaya ayrılan süre, kütüphane ve internet erişiminin; Rajchert, Zultak ve Smulczyk'in (2014) çalışmalarında cinsiyetin okuma becerileri için yordayıcı unsur olduğu anlaşılmaktadır. Ayrıca Williams'a (2003, s. 159) göre okuma hızı okuduğunu anlama açısından çok önemli olup, hızın düşük olması okuduğunu anlama etkinliğini olumsuz yönde etkilemektedir. Bu sonuçlara bağlı olarak araştırmancının amacı, okuma becerileri ile ilişkileri literatürce desteklenen bu özelliklerin okuma becerilerini yordama düzeyinin karmaşık ve büyük veri yapısı üzerinden araştırılmak istenmesidir. Farklı ülkelerden öğrencilerin yer aldığı ve gözlem sayısının fazla olduğu PISA uygulamasından üretilen veri grubuyla; cinsiyet, sınıf düzeyi, ebeveynlerin eğitim durumları, internet kullanımı, evde konuşulan dil, okuma becerilerini ölçen maddeleri yanıtlama hızları, evdeki kitap sayısı ve türü, okulda okuma becerileri için ayrılan haftalık ders saati, dijital araç kullanımı gibi özelliklerin okuma becerileriyle ilişkisinin araştırılması araştırmancının odağını oluşturmaktadır. PISA, OECD (Organisation for Economic Co-operation and Development) tarafından üç yılda bir 15 yaş grubu öğrencilere uygulanan dünya çapında bir ölçme ve değerlendirme programıdır. Fen, matematik ve okuma becerilerinin ölçülmesinin yanında öğrenci, öğrenme ortamları ve ailelere yönelik demografik bilgiler de toplanmaktadır. PISA programı, farklı ülkelerin eğitim sistemlerinin etkinliği ve etkililiği üzerine gerek eğitimci ve araştırmacılara, gerekse politikacılara önemli bilgiler sunmaktadır (OECD, 2017, s. 25).

Yordanan değişkenin okuma becerileri ve yordayıcı değişkenlerin yukarıda belirtilen özelliklerin olduğu en iyi doğrusal regresyon modelinin genetik algoritma yöntemiyle optimize (en iyileme) edilmesi araştırmadaki hedef durumun açıklaması olacaktır. Diğer bir ifadeyle, araştırmada

okuma becerilerini en iyi derecede yordayan değişkenlerden oluşan regresyon modeli için değişken seçimi (feature selection) işleminin gerçekleşmesi planlanmaktadır. Regresyon analizlerinin; yordayıcı değişkenlerin seçilmesi, regresyon modelini tanımlayan fonksiyonun belirlenmesi ve modeldeki parametrelerin kestirilmesi olmak üzere üç temel kullanımı söz konusudur (Paterlini ve Minerva, 2010; Yang, Chuang, Jeng ve Tao, 2011). Genetik algoritmalar, regresyon modellerinde en iyi modelin oluşturulması ya da model için en iyi yordayıcı değişkenlerin seçilmesinde kullanılan bir tekniktir (Vasant, 2013; Žegklitz ve Pošík, 2015). Genetik algoritmalar, sezgisel arama yaklaşımına dayanmakta olup canlıların biyolojik gelişiminden esinlenerek, doğal seçim mekaniği ve doğal genetiğe bağlı arama yapan olasılık temelli bir yöntemdir. İçinde bulunduğu grubun özelliklerine bağlı olarak, güçlü olan bireylerin hayatta kalması esasına dayanır. Canlılardaki gen yapısına benzer şekilde kodlama esasıyla oluşan kromozom düzeni prensibiyle çalışır. Çeşitli değişim ve gelişimlerden etkilenen bireyler arasından dayanıklılık gösterenlerin birleşmesiyle bir sonraki nesil oluşturulur (Gen ve Cheng, 1996, s. 1-3). Başlangıçta tüm bireyler başlangıç popülasyonunu oluşturur ve çözüm için belirlenen uygunluk fonksiyonu değerine göre uygun bireyler seçilerek bir sonraki popülasyona aktarılır. Bu işlem tekrarlanarak birkaç nesil sonra optimal çözümü temsil eden bireylere ulaşılmış olunur (Goldberg, 1989, s. 10-11). Genetik Algoritmaları diğer temel yöntemlerden ayıran önemli özelliklerinden biri diğer yöntemlerin deterministik, genetik algoritmaların ise olasılık esasına dayanmasıdır; bu sayede en yüksek uygunluk değerine sahip bireyler seçilebilmektedir. Ayrıca genel anlamda ikili kodlama prensibiyle çözüm üreten bir yöntem olması da ayrıcalıklı yönünü oluşturmaktadır (Örkcü, 2009). Leardi, Boggia ve Terrile (1992); Minghua, Qingxian, Benda ve Feng (2017); Paterlini ve Minerva (2010); Tolvi (2004) ve Trejos, Villalobus-Arias ve Espinoza'nın (2016) farklı problem durumları üzerinden genetik algoritma yöntemini regresyon modellerinde optimal model (değişken) seçimine ilişkin çalışmalarından elde edilen sonuçlar, bu yöntemin başarısını ortaya koymaktadır.

Tüm bu durumlar değerlendirildiğinde okuma becerisinin hem bireyler hem de toplumlar açısından önemli bir olgu olduğu anlaşılmaktadır. Sorgulama, eleştirme, araştırma, yorum yapabilme, değerlendirme yetisine sahip olma, çıkarım yapabilme gibi üst düzey bilişsel becerilerin okuma ve okuduğunu anlama yeteneğiyle ilişkilendirildiği dikkate alındığında bu becerilerin öneminin oldukça fazla olduğu söylenebilmektedir. Günümüz gereksinimlerinden olan bilginin işlenmesi ve en verimli şekilde kullanılmasının ülkelerin gelişmişlik düzeyini önemli ölçüde etkilemekte ve okuyan, anlayan, çıkarım yapan, araştıran, eleştiren, sorgulayan, analiz ve sentez yetilerini geliştiren bireylerin arzulan birey profili oluşturdukları anlaşılmaktadır. Bu bağlamda okuma becerilerini yordayan özelliklerin büyük ve karmaşık bir veri grubu aracılığıyla araştırılıyor olmasının ulaşılabilecek sonuçlar açısından alanyazına katkı sağlayacağı düşünülmektedir. Dolayısıyla bu araştırmayla, PISA 2015 uygulamasına katılan öğrenci verilerinden yararlanılarak oluşturulan veri grubu için, öğrencilerin okuma becerilerini yordayan değişkenler ve bu değişkenlere ait özellikler nelerdir?, sorusuna cevap aranacaktır. Bunun yanında analiz yöntemi olarak kullanılan genetik algoritma tekniğinin araştırma probleminin çözülmesinde diğer araştırmalardan farklı olarak alanyazına sağlayacağı katkının önem arz edeceği söylenebilmektedir.

Yöntem

Araştırmanın Modeli

Araştırmada, okuma becerilerini yordayan özelliklerin neler olduğu belirlenmek istenmektedir. Bu bağlamda, iki değişken arasındaki önemli ölçüdeki ilişkiden hareketle, değişkenlerin birinden diğer değişkenin yordandığı olduğu araştırmalar korelasyonel (ilişkisel) araştırmalara örnek olmaktadır (Fraenkel ve Wallen, 2006, s. 337). Yordayıcı değişken iki ya da daha fazla olduğunda, araştırma çok faktörlü yordayıcı korelasyonel desen olarak belirtilmektedir (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2014, s. 186).

Araştırma grubu

Araştırmanın çalışma grubu, PISA 2015 uygulamasından elde edilen verilerden yararlanılarak, okuma becerileri ile ilgili çoktan seçmeli maddeleri aynı sırada cevaplayan öğrencilerin seçilmesi esasıyla oluşturulmuştur. Bu işlem için öğrencilerin en çok yanıtladığı kitapçık olması ve yanıtlama

sürelerinin kaydedilmesi gibi araştırmanın amacına uygun koşulları içerdiğinden, bilgisayar tabanlı test (computer based) uygulamasıyla yanıtlanmış olan 36. kitapçığın bir formu tercih edilmiştir. Bu formu 9396 öğrenci yanıtlamıştır. Araştırmanın yordanan ve yordayıcı değişkenlerinin bir araya getirildiği veri setinde kayıp verilerin liste bazında veri silme yöntemiyle silinmesi sonucunda 5232 öğrenci verisiyle araştırmanın çalışma grubu elde edilmiştir. Çalışma grubunu oluşturan öğrencilere ait ülke bilgileri Tablo 1’de sunulmuştur.

Tablo 1.
Ülkelere Göre Öğrenci Sayıları

Ülke	N	Ülke	N	Ülke	N
Almanya	68	Hollanda	104	Litvanya	123
Avusturalya	215	Hong Kong	109	Lüksemburg	63
Avusturya	100	İrlanda	141	Macaristan	88
Belçika	138	İspanya	99	Makao	102
Birleşik Devletler (ABD)	86	İspanya (Bölgeler)	485	Norveç	91
Birleşik Krallık	221	İsrail	81	Polonya	135
Çekya	100	İsviçre	122	Portekiz	114
Çin	191	İtalya	184	Rusya	79
Çin Taipei (Tayvan)	173	İzlanda	50	Singapur	127
Danimarka	117	Japonya	108	Slovakya	96
Estonya	112	Kanada	309	Slovenya	112
Finlandiya	123	Karadağ	48	Türkiye	81
Fransa	102	Kore	105	Yeni Zelanda	65
Hırvatistan	101	Letonya	97	Yunanistan	87

Araştırma, çalışma grubunu oluşturan 42 farklı ülkeden 2699’u kız (%51,6) ve 2533’ü erkek (%48,4) olmak üzere toplam 5232 öğrenci verisiyle gerçekleştirilmiştir.

Veri Toplama Araçları

Araştırmanın yordanan değişkeni olan okuma becerilerini PISA 2015 uygulamasındaki 36. kitapçıkta yer alan 19 çoktan seçmeli maddeye verilen cevaplara ait öğrenci ham puanları toplamı temsil etmektedir. Ham puanlar hesaplanırken doğrular 1 ve yanlış ve boş yanıtlar 0 puan olarak değerlendirilmiştir. Araştırmanın yordayıcı değişkenleri olan özellikler ise öğrencilerin PISA 2015 programındaki ilgili maddelere verdiği yanıtlardan elde edilen; cinsiyet, evde internet kullanımı, e-kitap okuyucu sayısı, evdeki kitap türü ve sayısı, evde konuşulan dil, okuma becerilerine yönelik okutulan haftalık ders saati, anne eğitim durumu, baba eğitim durumu, sınıf düzeyi ve maddeleri yanıtlamak üzere ayrılan toplam süre olarak sıralanmaktadır. Ayrılan toplam süre dakika cinsinden değerlendirilmiştir. Yordanan değişken olan okuma becerilerini oluşturan maddelerin PISA uygulaması kodları Tablo 2’de sunulmuştur.

Tablo 2.
Okuma Becerilerine İlişkin Maddelerin PISA Kodları

Madde	Kod	Madde	Kod
1	CR067Q01S	11	CR083Q04S
2	CR102Q07S	12	CR442Q07S
3	CR220Q01S	13	CR245Q01S
4	CR220Q02S	14	CR245Q02S
5	CR220Q04S	15	CR101Q01S

Okuma Becerilerini Yordayan Özelliklerin Belirlenmesi: Genetik Algoritma Kestirimi

6	CR220Q05S	16	CR101Q02S
7	CR220Q06S	17	CR101Q03S
8	CR083Q01S	18	CR101Q04S
9	CR083Q02S	19	CR101Q05S
10	CR083Q03S		

Yordayıcı değişkenlere ilişkin PISA uygulaması kodları Tablo 3'te sunulmuş olup yanıtlama sürelerine ait kodlar Tablo 2'de sunulan her madde kodunun ilk sekiz karakterine TT ikili karakterinin eklenmesiyle ifade edilmektedir.

Tablo 3.

Yordayıcı Değişkenlere İlişkin Maddelerin PISA Kodları

Değişken	Kod	Değişken	Kod
Cinsiyet	ST004D01T	Sınıf düzeyi	ST001D01T
İnternet kullanımı	ST011Q06TA		ST005Q01TA01-5
Konuşulan dil	ST022Q01TA	Anne eğitim düzeyi	ST006Q01TA01
E-kitap okuyucu sayısı	ST012Q08NA		ST006Q02TA01
	ST011Q07TA01		ST006Q03TA01
	ST011Q08TA01		ST007Q01TA01-5
Kitap türü	ST011Q09TA01	Baba eğitim düzeyi	ST008Q01TA01
	ST011Q10TA01		ST008Q02TA01
	ST011Q11TA01		ST008Q03TA01
	ST011Q12TA01	Kitap sayısı	ST013Q01TA
Haftalık ders saati	ST059Q01TA		

Verilerin Toplanması

Araştırma verileri, PISA 2015 programına katılan öğrencilere ait verilerden yararlanılarak oluşturulmuştur.

Verilerin Analizi

Verilerin regresyon modeli esasıyla genetik algoritmalar yöntemiyle analiz edilmesinden önce analize hazır hale getirilmesi işlemi gerçekleştirilmiştir. Bu bağlamda PISA veri setinden 36. kitapçığı kullanan öğrencilere ait regresyon modelinin değişkenlerinden oluşan alt veri seti SPSS programı kullanılarak ayıklanmıştır. Ayıklanan veri setindeki kayıp veriler liste bazlı veri silme (listwise deletion) yöntemiyle temizlenmiştir (Enders, 2010, s. 39). Bu işlemler sonrasında regresyon modelinin yordanan ve yordayıcı değişkenlerini içeren 5232 öğrenciye ait araştırmanın veri seti elde edilmiştir.

Regresyon modelinin analiz edilmesinde kullanılan genetik algoritmalar gibi makine öğrenmesine dayalı yöntemlerde modelde yer alan tüm değişkenlerin sayısal (numeric) değerde olması gerekir. Bu durum kategorik yapıdaki değişkenlerin sayısal kodlamalara dönüştürülmesi anlamına gelmektedir. Bu amaçla kullanılan popüler tekniklerden biri sıralı kodlama (ordinal encoding) tekniğidir. Sıralı kodlama tekniğinde kategorik değişkene ilişkin düzeyler 1-2-3 gibi tamsayılarla ifade edilerek dönüşüm gerçekleştirilir (Brownlee, 2020, s. 241-243). Bu nedenle regresyon modelinde yer alan kategorik değişkenler sıralı kodlama esasıyla sayısal değerlere dönüştürülmüştür.

Genetik algoritmalar yöntemi parametrik istatistik yöntemlerinin gerektirdiği normalite, doğrusallık, homojenlik gibi sayıltıları gerektirmediğinden (Altunkaynak, 2009; Şen ve Öztopal, 2001), analiz için her hangi sayıltı testi uygulanmamıştır. Veriler R programı ortamına aktarılmış ve diğer analiz işlemleri bu platformda gerçekleştirilmiştir. Problem durumunun çözülmesi için gerçekleştirilen analizler; caret (Kuhn vd., 2021), GA (Scrucca, 2021), mlbench (Leisch ve Dimitriadou,

2021); modelr (Wickham ve Rstudio, 2020); recipes (Kuhn, Wickham ve Rstudio, 2021) ve rsample (Silge, Chow, Kuhn, Wickham ve Rstudio, 2021) paketleri aracılığıyla sağlanmıştır.

Genetik Algoritmalar

Charles Darwin'in Evrim Teorisi ilkeleri esas alınarak doğayı taklit etme ve güçlü olanın hayatta kalması prensibine dayanan genetik algoritmalar, doğal seçim ve arama yöntemidir. Canlıların biyolojik gelişiminden esinlenerek, doğal seçim mekaniği ve doğal genetiğe bağlı arama yapan olasılık temelli bir yöntemdir. İçinde bulunduğu grubun özelliklerine bağlı olarak, güçlü olan bireylerin hayatta kalması esasına dayanır. Canlılardaki gen yapısına benzer şekilde kodlama esasıyla oluşan kromozom düzeni prensibiyle çalışır. Çeşitli değişim ve gelişimlerden etkilenen bireyler arasından dayanıklılık gösterenlerin birleşmesiyle bir sonraki nesil oluşturulur (Michaleweich, 1996, s. 13-15).

Genetik Algoritmalar; Darwin'in savunduğu biyolojik süreci yapay ortamlarda modelleyerek, fonksiyonlara en iyi çözümü üreten algoritmalar. Eşit sayıda gen içeren kromozomlar birey rolüyle popülasyonları oluşturur. Kromozomlar, problemlerin çözümüne yönelik verilerin depolandığı yapılardır. Çözümü sağlayacak parametreler yan yana dizilerek gen yapısını oluşturmak üzere kromozomlara kodlanırlar. Parametre sayısı kromozom uzunluğunu belirlerken aynı zamanda problemin çözümünü sağlayacak bir çözüm önerisi anlamına gelmektedir (Michaleweich, 1996, s. 14-15). Genetik Algoritmalar ile bir probleme yönelik çözüm süreci genetik algoritma operatörlerinin işletilmesiyle gerçekleşmektedir. Bu operatörler çözüm uzayının taranarak en iyi çözümün bulunmasını sağlayan adımlar içermektedir. Bu operatörler; kopyalama, çaprazlama ve mutasyon olarak sıralanmaktadır (Goldberg, 1989, s. 10).

Problemin çözüm sürecinde her bir birey (kromozom) için bir uygunluk değeri hesaplanır. Uygunluk değerinin hesaplanması çözüm için geliştirilen amaç (uygunluk) fonksiyonu üzerinden gerçekleşir. Problemin yapısına göre amaç fonksiyonunun maksimum ya da minimum uygunluk değeri çözüm için esas alınır (Haupt ve Haupt, 1998, s. 30). Hesaplanan uygunluk değerlerine göre bireyler güçlü ya da zayıf olarak sınıflandırılır. Evrim mantığı esasıyla, güçlü bireylerin çoğalmalarının sağlanması adına hayatta kalma olasılıkları artırılırken; zayıf bireylerin ölme olasılıkları artırılarak yok olmaları sağlanır. Operatörler aracılığıyla iterasyonlar izlenerek, güçlü olan bireylerin üremesiyle çoğalan yavru bireylerle yeni bir popülasyon oluşturulur. İstenilen iterasyon sayısı ya da durdurma koşulu sağlandığında üreme ve iterasyon durdurularak en iyi uygunluk değerine ulaşılmış olur. Ulaşılan bu uygunluk değeri problemin çözüm noktası olarak belirlenir (Michaleweich, 1996, s. 17).

Genetik Algoritma yöntemi ile bir problemin çözümünü sağlayan işlem adımları aşağıdaki gibi sıralanır (Karakoca, 2009, s. 12):

1. Adım: Parametre değişim aralıkları, uygunluk fonksiyonu, çaprazlama ve mutasyon olasılıkları, popülasyon büyüklüğü gibi başlangıç değerleri belirlenir.
2. Adım: Başlangıç popülasyonu oluşturulur.
3. Adım: Popülasyonun uygunluk değerleri hesaplanır.
4. Adım: Sonlandırma koşulları sağlandıysa 8. adım'a, sağlanmadıysa 5. adım'a gidilir.
5. Adım: Kopyalama operatörü uygulanır.
6. Adım: Çaprazlama operatörü uygulanır.
7. Adım: Mutasyon operatörü uygulanır.
8. Adım: En iyi çözüm kaydedilir.

Geçerlik ve Güvenirlik

Oluşturulan modelin analiz sonuçlarına yönelik uyum ve performans bilgilerine araştırmanın bulgular kısmında yer verilmiştir.

Araştırma ve Yayın Etiği

Bu çalışmada "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altında belirtilen eylemlerden hiçbiri

gerçekleştirilmemiştir. Çalışmada kamusal erişime açık durumda olan, kişisel bilgi ve etik değerlerin uluslararası standartlarda korunduğu PISA verileri kullanıldığından etik kurul izni gerekmemektedir.

Bulgular

Araştırmanın problem durumu olan, PISA 2015 uygulamasına katılan öğrenci verilerinden yararlanılarak oluşturulan veri grubu için, öğrencilerin okuma becerilerini yordayan değişkenler ve bu değişkenlere ait özellikler nelerdir?, ifadesine yanıt aramak doğrultusunda elde edilen betimsel özelliklere ve oluşturulan regresyon modelinin genetik algoritma yöntemi esasıyla analiz edilmesi sonucunda ulaşılan bulgulara aşağıda yer verilmiştir.

Genetik algoritmalar yöntemi esasıyla analiz edilen ve yordanan değişkenin okuma becerileri; yordayıcı değişkenlerin ise cinsiyet, anne ve baba eğitim durumu, evde internet kullanımı, evde konuşulan dil, sahip olunan e-kitap okuyucu sayısı, evdeki kitap türü ve sayısı, okuma becerilerini ölçen maddeleri yanıtlama hızları, okulda okuma becerileri için ayrılan haftalık ders saati ve sınıf düzeyi olan regresyon modelinin değişkenlerine ilişkin betimsel özellikler Tablo 4a ve Tablo 4b'de sunulmuştur.

Tablo 4a.

Değişkenlere İlişkin Betimsel Özellikler

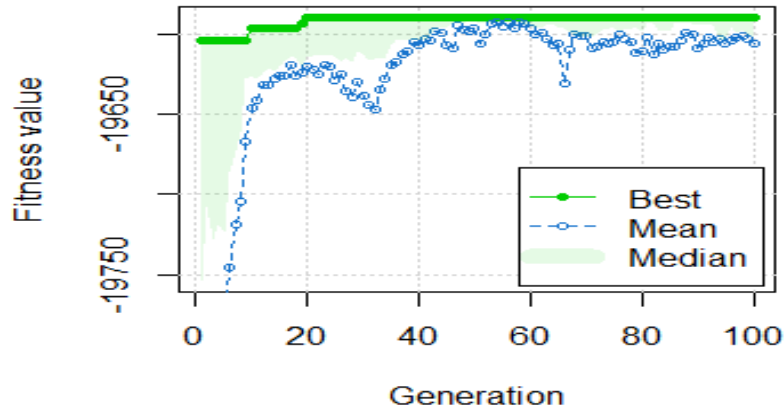
Değişken		N	%	Değişken		N	%
Cinsiyet	Kız	2699	51.6	internet kullanımı	Var	5072	96.9
	Erkek	2533	48.4		Yok	160	3.1
	Tamamlamadı	83	1.6	Konuşulan dil	Test dili	4683	89.5
İlkokul	229	4.4	Diğer		549	10.5	
Anne eğitim düzeyi	Ortaokul	739	14.1	Sınıf düzeyi	7	13	.2
	Lise	1313	25.1		8	164	3.1
	Yüksekokul	743	14.2		9	1718	32.9
	Lisans	981	18.8		10	2995	57.2
	Lisansüstü	1144	21.9		11	336	6.4
	Tamamlamadı	74	1.4		12	6	.1
	İlkokul	256	4.9		0-10	520	9.9
Baba eğitim düzeyi	Ortaokul	884	16.9	11-25	747	14.3	
	Lise	1240	23.7	Kitap sayısı	26-100	1581	30.2
	Yüksekokul	775	14.8		101-200	1029	19.7
	Lisans	867	16.6		201-500	888	17
Lisansüstü	1136	21.7	501-üstü		467	8.9	

Tablo 4b.

Değişkenlere İlişkin Betimsel Özellikler

Değişken	N	Min.	Maks.	Ortalama	S. sapma
Okuma becerileri	5232	0	19	13.14	3.8
Yanıtlama süresi	5232	2.65	38.38	16.98	4.52
E-kitap okuyucu sayısı	5232	0	3	.35	.67
Kitap türü	5232	0	6	4.11	1.61
Haftalık ders saati	5232	2	10	4.43	1.41

Genetik algoritmalar yöntemi gibi makine öğrenmelerine dayalı yöntemlerde kestirimlere yönelik yanlış hatalar üretilmesi problemi olarak açıklanan aşırı uyum sorunlarının önüne geçmek için veri seti eğitim (training) ve test (testing) seti olmak üzere iki alt gruba ayrılır (Kushchu, 2002; Žegklitz ve Pošík, 2015). Bu nedenle, verilerin %70'i eğitim seti ($n=3663$) ve %30'u test seti ($n=1569$) olmak üzere (Ahmed ve Elaraby, 2014; Pahmi, Saepudin, Maesarah, Solehudin ve Wulandari, 2018) ikiye ayrılarak analiz işlemleri gerçekleştirilmiştir. Regresyon modeli, eğitim seti verileri kullanılarak analiz edilmiş daha sonra oluşturulan modelin farklı veri grupları için performans düzeyinin belirlenmesinde test ve tüm veri seti kullanılmıştır. Genetik Algoritmalar yöntemiyle gerçekleşen analizde; başlangıç topluluğu büyüklüğü 50, maksimum nesil sayısı 100, mutasyon işlemi olasılığı .1, çaprazlama işlemi olasılığı .8 ve elitizm sayısı 2 olarak belirlenen ikili kodlama esasına dayalı genetik işlemler uygulanmıştır. Uygunluk (amaç) fonksiyonu olarak, en iyi regresyon modelinin belirlenmesini sağlayan değişkenlerin seçimi amacıyla model uyum kriteri olarak Bayes Bilgi Kriteri (BIC) kriterini en küçük yapan fonksiyon kullanılmıştır (Minghua vd., 2017; Trejos vd., 2016). Analizde kullanılan R programı GA paketi uygunluk (amaç) fonksiyonunu maksimum yapma özelliği gösterdiğinden, minimum BIC değerine ulaşmak adına negatif BIC değerinin maksimum değerinin elde edilmesi sağlanmıştır. Bu bağlamda en iyi model için BIC değeri -19,638.64 olarak saptanmış olup uygunluk (amaç) fonksiyonu değeri olan BIC kriterinin belirlenmesinde gözlenen iterasyonlar Şekil 1 ile sunulmuştur.



Şekil 1. Uygunluk (Amaç) Fonksiyonunun Belirlenmesinde Gözlenen İterasyonlar

Genetik algoritmalar yöntemine ait kestirim özellikleri ve en iyi regresyon modeli için seçilen değişkenlere ilişkin bilgileri içeren R programı analiz çıktısı Şekil 2'de sunulmuştur.

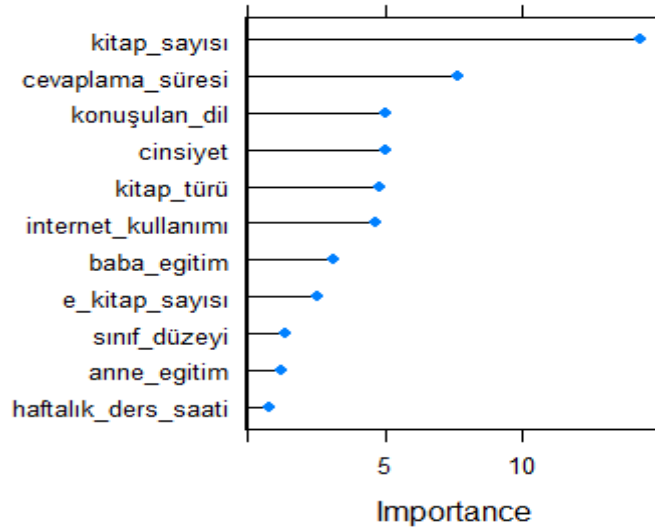
```
-- Genetic Algorithm -----
GA settings:
Type = binary
Population size = 50
Number of generations = 100
Elitism = 2
Crossover probability = 0.8
Mutation probability = 0.1

GA results:
Iterations = 100
Fitness function value = -19638.64
Solution =
  sınıf_düzeyi cinsiyet internet_kullanımı e_kitap_sayısı kitap_sayısı konuşulan_dil cevaplama_süresi
[1,] 0 1 1 1 1 1
  haftalık_ders_saati anne_egitim baba_egitim kitap_türü
[1,] 0 0 1 1
```

Şekil 2. Genetik Algoritmalar Analizi R Programı Ekran Alıntısı

Okuma Becerilerini Yordayan Özelliklerin Belirlenmesi: Genetik Algoritma Kestirimi

İkili kodlamaya dayalı genetik algoritmalar yöntemine yönelik analizlerde seçilen değişken 1; seçilmeyen değişken ise 0 kodlanarak çıktı olarak sunulur. Şekil 2 incelendiğinde, en iyi regresyon modeli için okuma becerilerini istatistiksel olarak anlamlı şekilde yordayan değişkenlerin; cinsiyet, internet kullanımı, e-kitap okuyucu sayısı, evdeki kitap sayısı, evde konuşulan dil, okuma becerilerini ölçen maddeleri yanıtlama hızları, baba eğitim düzeyi ve evde bulunan kitap türü sayısı olmak üzere sekiz özelliğin olduğu saptanmıştır. Sınıf düzeyi, okuma becerileri için ayrılan haftalık ders saati ve anne eğitim düzeyi değişkenleri ise model için okuma becerilerini istatistiksel olarak anlamlı düzeyde yordamamıştır. Yordayıcı değişkenlerin 0 ile 15 değerleri arasında ölçeklenen regresyon modeli için önem dereceleri Şekil 3'te sunulmuştur.



Şekil 3. Yordayıcı Değişkenlerin Önem Düzeyi

Şekil 3 incelendiğinde, öğrencilerin okuma becerilerini istatistiksel olarak anlamlı düzeyde yordayan en önemli değişkenin evde bulunan kitap sayısı olduğu anlaşılmaktadır. Bu değişkeni önem sırasına göre; okuma becerilerini ölçen maddeleri yanıtlama hızları, evde konuşulan dil, cinsiyet, evdeki kitap türü sayısı, evde internet kullanımı, baba eğitim düzeyi, sahip olunan e-kitap okuyucu sayısı takip etmektedir. Önem derecesi en düşük olan üç değişken; sınıf düzeyi, anne eğitim düzeyi ve haftalık ders saati değişkenlerinin ise öğrencilerin okuma becerilerini istatistiksel olarak anlamlı düzeyde yordamadığı saptanmıştır. Eğitim veri seti üzerinden oluşturulan regresyon modelinin performansı eğitim seti, test seti ve tüm veri seti için R^2 , RMSE ve MAE metrikleri ile araştırılmış (Özdemir, 2017; Paterlini ve Minerva, 2010) olup, bilgiler Tablo 5 ile sunulmuştur.

Tablo 5.

Regresyon Modelinin Veri Setleri İçin Performans Metrikleri

Veri seti	N	R^2	RMSE	MAE
Eğitim seti	3663	.1527	3.47	2.78
Test seti	1569	.1524	3.55	2.88
Tüm veri seti	5232	.1527	3.49	2.81

Eğitim seti verileriyle analiz edilen optimal regresyon modelinin performansının ifade edildiği Tablo 5 incelendiğinde, eğitim seti için ulaşılan performans metrikleri değerlerinin farklı veri seti özelliğinde olan test verisi ve tüm veri seti için yakın değerlerde olduğu gözlenmektedir. Bu bağlamda, oluşturulan modelin farklı veri yapılarında benzer derecede performans gösterdiği sonucu çıkarılabilmektedir. Ayrıca eğitim seti üzerinden elde edilen değerler dikkate alındığında, öğrencilerin

okuma becerilerini istatistiksel olarak anlamlı düzeyde yordayan değişkenlerin okuma becerilerindeki değişkenliğin %15,27'sini ($R^2=.1527$) açıkladığı anlaşılmaktadır.

Tartışma ve Sonuç

PISA 2015 uygulaması verilerinden yararlanılarak oluşturulan veri grubundaki öğrencilerin okuma becerilerini yordayan değişkenlerin belirlenmesinin genetik algoritmalar esasıyla en iyi (optimal) regresyon modelinin kestirilmesi amacıyla gerçekleştirilen analizler sonucunda; öğrencilerin okuma becerilerini sekiz değişkenin istatistiksel olarak anlamlı düzeyde yordadığı saptanmıştır.

Öğrencilerin okuma becerilerini istatistiksel olarak anlamlı düzeyde yordayan özelliklerden üçü; öğrencilerin sahip oldukları e-kitap okuyucu sayısı, evlerinde bulunan kitap sayısı ve türü olmuştur. Bu bağlamda öğrencilerin sahip oldukları e-kitap okuyucu sayısı, ulaşabildikleri kitap sayısı ve kitap türü çeşitliliği öğrencilerin okuma becerilerinde anlamlı düzeyde farklılaşmaya yol açmaktadır. Alanyazında benzer amaçla yapılan çalışmalar incelendiğinde, Adaba (2016); Dadandı, Dadandı ve Koca (2018); Giambona ve Porcu (2015); Kutlu, Yıldırım, Bilican ve Kumandaş (2011) ve Torres'in (2019) yapmış oldukları çalışmalarda araştırmanın sonuçlarına paralel olarak öğrencilerin ulaşabildikleri kitap sayısı ve türünün okuma becerilerini anlamlı düzeyde yordadığı sonucuna ulaşmışlardır. Mustadi ve Amri (2020) ve Pitoyo (2020) çalışmalarında ise evde ya da okulda kütüphane bulunmasının okuma becerileri için yordayıcı bir unsur olduğu gözlenmektedir. Ayrıca Bozkuş (2021); Dadandı ve diğerleri (2018) ve Gumus ve Atalmis (2011) çalışmalarında dijital araç kullanımının okuma becerilerini istatistiksel olarak anlamlı düzeyde yordadığını ifade etmektedirler.

Öğrencilerin okuma becerilerini istatistiksel olarak anlamlı düzeyde yordayan diğer iki değişken cinsiyet ve evde konuşulan dil olmuştur. Elde edilen sonuçlara göre; öğrencilerin cinsiyetleri ve evde konuştukları dilin eğitim aldıkları (test dili) dilden farklı olması öğrencilerin okuma becerilerinde farklılaşmaya yol açmaktadır. Lazarus (2020); Rajchert ve diğerleri (2014); Temizyürek ve diğerleri (2013); Torppa, Eklund, Sulkunen, Niemi ve Ahonen'in (2018) çalışmalarında cinsiyetin okuma becerileri için yordayıcı bir unsur olduğu gözlenmektedir. Cheung, Sit, Soh, leong ve Mak (2014) ve Harju-Luukkainen ve diğerlerinin (2016) çalışmalarında ise evde konuşulan dilin eğitim dilinden farklı olmasının öğrencilerin okuma becerileri açısından farklılık yarattığı sonucuna ulaşılmaktadır.

Araştırma sonuçlarına göre öğrencilerin okuma becerilerini istatistiksel olarak anlamlı düzeyde yordayan diğer değişken öğrencilerin internet kullanımı olmuştur. Öğrencilerin evlerinde internet kullanma imkânına sahip olup olmaması okuma becerilerinde farklılaşmaya yol açmaktadır. Giambona ve Porcu (2015); Kurnaz ve Yıldız (2015) ve Pitoyo'nun (2020) çalışmaları bu sonucu desteklemektedir.

Öğrencilerin okuma becerilerini yordayan bir diğer özellik baba eğitim düzeyi olup, anne eğitim düzeyinin öğrencilerin okuma becerilerini istatistiksel olarak anlamlı düzeyde yordamadığı saptanmıştır. Bu durumda öğrencilerin babalarının farklı eğitim kademelerinden mezun olmasının okuma becerilerinde farklılaşmaya yol açtığı anlaşılmaktadır. Giambona ve Porcu (2015); Kurnaz ve Yıldız (2015); Tebekana ve Cishe (2015); Torres (2019) ve Torres, Ordonez ve Calvo (2021) çalışmalarında ebeveynlerin eğitim düzeyinin öğrencilerin okuma becerilerini anlamlı şekilde yordadığı sonucuna ulaşmışlardır.

Yordama düzeyi istatistiksel olarak anlamlı bulunan son değişken öğrencilerin okuma becerilerini ölçen maddeleri yanıtlama hızları olmuştur. Öğrencilerin okuma becerilerini ölçen maddelere verdikleri yanıtlar için harcadıkları toplam sürenin okuma becerilerinde farklılaşmaya yol açtığı anlaşılmaktadır. Carretti, Toffalini, Saponaro, Viola ve Cornoldi (2019); Dyson ve Haselgrove (2001) ve Uysal ve Bilge'nin (2018) çalışmalarında okuma hızının okuma becerilerini istatistiksel olarak anlamlı düzeyde yordadığı sonucuna ulaşmışlardır.

Araştırmada öğrencilerin okuma becerilerini istatistiksel olarak anlamlı düzeyde yordayan değişkenlerin okuma becerilerindeki değişkenliğin %15,27'sini açıkladığı saptanmış olup, ulaşılan sonuçların alanyazındaki benzer araştırma sonuçlarıyla tutarlılık gösterdiği gözlenmiştir. Bu bağlamda cinsiyet, evde internet kullanımı, evde konuşulan dil, sahip oldukları e-kitap okuyucu sayısı, baba

eğitim düzeyi, okuma becerilerini ölçen maddeleri yanıtlama hızları, evdeki kitap türü ve sayısının farklılaşması öğrencilerin okuma becerilerinde de farklılaşmaya yol açtığı söylenebilmektedir.

Araştırma ve Yayın Etiği

Bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbirini gerçekleştirilmemiştir.

Yazarların Katkı Oranı

Yazarlar çalışmaya eşit oranda katkı sağlamışlardır.

Çıkar Çatışması

Çalışma çıkar çatışması teşkil edecek herhangi bir durum içermemektedir.

Kaynaklar

- Adaba, H. (2016). Assessing factors affecting the students reading speed and comprehension: Manasibu secondary school grade nineth in focus: Western Wallagga Zone. *International Journal of Language and Linguistics*, 4(5), 165-182.
- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge: MIT Press.
- Ahmed, A. B. ve Elaraby, I. S. (2014). Data mining: A prediction for student's performance using classification method. *World Journal of Computer Application and Technology*, 2(2), 43-47.
- Altunkaynak, A. (2009). Sediment load prediction by genetic algorithms. *Advances in Engineering Software*, 40(9), 928-934.
- Bozkuş, K. (2021). Digital devices and student achievement: The relationship in PISA 2018 data. *International Online Journal of Education and Teaching (IOJET)*, 8(3), 1560-1579.
- Brownlee, J. (2020). *Data preparation for machine learning: Data cleaning, feature selection, and data transforms in Python*. Machine Learning Mastery.
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2014). *Bilimsel araştırma yöntemleri* (16. bs.). Ankara: Pegem.
- Carretti, B., Toffalini, E., Saponaro, C., Viola, F. ve Cornoldi, C. (2020). Text reading speed in a language with a shallow orthography benefits less from comprehension as reading ability matures. *British Journal of Educational Psychology*, 90(Suppl 1), 91-104.
- Cheung, K. C., Sit, P. S., Soh, K. C., leong, M. K. ve Mak, S. K. (2014). Predicting academic resilience with reading engagement and demographic variables: Comparing Shanghai, Hong Kong, Korea, and Singapore from the PISA perspective. *The Asia-Pacific Education Researcher*, 23(4), 895-909.
- Coşkun, E. (2002). Lise hızlı okuma teknikleri öğretim programı ve uygulamalarının değerlendirilmesi. *Eğitim Araştırmaları*, 9(5), 41-51.
- Dadandı, P.U., Dadandı, İ. ve Koca, F. (2018). PISA Türkiye sonuçlarına göre sosyoekonomik faktörler ile okuma becerileri arasındaki ilişkiler. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi*, 7(2), 1239-1252.
- Dyson, M. C. ve Haselgrove, M. (2001). The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, 54(4), 585-612.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.
- Fraenkel, J. R. ve Wallen, N. E. (2006). *How to design and evaluate research in education* (6. bs.). New York: McGraw-Hill.
- Gen, M. ve Cheng, R. (1997). *Genetic algorithms and engineering design*. New York: Wiley.
- Giambona, F. ve Porcu, M. (2015). Student background determinants of reading achievement in Italy. A quantile regression analysis. *International Journal of Educational Development*, 44(C), 95-107.

- Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning*. Massachusetts: Addison-Wesley.
- Grabe, W. ve Stoller, F. L. (2011). *Teaching and researching reading*. New York: Pearson Education Limited.
- Gumus, S. ve Atalmis, E. H. (2011). Exploring the relationship between purpose of computer usage and reading skills of Turkish students: Evidence from PISA 2006. *Turkish Online Journal Of Educational Technology-TOJET*, 10(3), 129-140.
- Harju-Luukkainen, H., Vettenranta, J., Ouakrim-Soivio, N. ve Bernelius, V. (2016). Differences between students' PISA reading literacy scores and grading for mother tongue and literature at school: A geostatistical analysis of the finnish PISA 2009 data. *Education Inquiry*, 7(4), 29413.
- Haupt, R. L. ve Haupt, S. E. (1998). *Practical genetic algorithms*. USA: Willey-Interscience Publication.
- Karakoca, A. (2009). *Çok değişkenli lineer olmayan modellerde genetik algoritma* (Yayımlanmamış doktora tezi). Selçuk Üniversitesi, Konya.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182.
- Kuhn, M., Wickham, H. ve RStudio. (2021). Preprocessing and feature engineering steps for modeling. R package version 0.1.17. <https://cran.r-project.org/web/packages/recipes/index.html> adresinden erişildi.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C. ve Hunt, T. (2021). Classification and regression training. R package version 6.0-90. <https://cran.r-project.org/web/packages/caret/index.html> adresinden erişildi.
- Kurnaz, H. ve Yıldız, N. (2015). Ortaokul öğrencilerinin okuma motivasyonlarının çeşitli değişkenlere göre değerlendirilmesi. *Türkiye Sosyal Araştırmalar Dergisi*, 19(3), 53-70.
- Kushchu, I. (2002). Genetic programming and evolutionary generalization. *IEEE Transactions on Evolutionary Computation*, 6(5), 431-442.
- Kutlu, Ö. (2004, Mayıs). *Türkiye’de demokrasi anlayışının gelişmesini sağlayacak bir yol: Okuduğunu anlama becerilerinin geliştirilmesi*. Uluslararası Demokrasi Eğitimi Sempozyumu’nda sunulan bildiri, On sekiz Mart Üniversitesi, Çanakkale.
- Kutlu, Ö., Yıldırım, Ö., Bilican, S. ve Kumandaş, H. (2011). İlköğretim 5. sınıf öğrencilerinin okuduğunu anlamada başarılı olup olmama durumlarının kestirilmesinde etkili olan değişkenlerin incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2(1), 131-139.
- Lazarus, K. (2020). Socio-demographic factors affecting reading comprehension achievement among secondary school students with learning disabilities in Ibadan, Nigeria. *IAFOR Journal of Education: Language Learning in Education*, 8(1), 145-157.
- Leardi, R., Boggia, R. ve Terrile, M. (1992). Genetic algorithms as a strategy for feature selection. *Journal of Chemometrics*, 6(5), 267-281.
- Leisch, F. ve Dimitriadou, E. (2021). Machine learning benchmark problems. R package version 2.1-3. <https://cran.r-project.org/web/packages/mlbench/index.html> adresinden erişildi.
- Michalewicz, Z. 1996. *Genetic algorithms + data structures = evolution programs* (3. bs.). USA: Springer.
- Minghua, S., Qingxian, X., Benda, Z. ve Feng, Y. (2017). Regression modelling based on improved genetic algorithm. *Tehnicky vjesnik/Technical Gazette*, 24(1), 63-70.
- Mustadi, A. ve Amri, F. (2020, December). Factors affecting reading interest of elementary school students. *2nd Yogyakarta International Conference on Educational Management/Administration and Pedagogy (YICEMAP 2019)* içinde (s. 15-21). Atlantis Press.
- OECD. (2017). *PISA 2015 Technical report*. Paris: OECD Publishing.
- Örkcü, H. H. (2009). *Ayırma analizine matematiksel programlama ve yapay sinir ağları yaklaşımları* (Yayımlanmamış doktora tezi). Gazi Üniversitesi, Ankara.
- Özçelik, D. A. (1987). *Eğitim programları ve genel öğretim yöntemi*. Ankara: ÖSYM Eğitim Yayını.
- Özdemir, E. (1993). *Türkçe öğretimi*. İstanbul: İnkılap Kitabevi.

- Özdemir, M. (2017). Genetik algoritma ile doğrusal regresyonda tahmin amaçlı model seçimi. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 28, 213-233.
- Pahmi, S., Saepudin, S., Maesarah, N., Solehudin, U. I. ve Wulandari (2018). Implementation of CART (classification and regression trees) algorithm for determining factors affecting employee performance. *2018 International Conference on Computing, Engineering, and Design (ICCED) 6-8 September 2018* içinde (s. 57-62). Bangkok: Institute of Electrical and Electronics Engineers (IEEE).
- Paterlini, S. ve Minerva, T. (2010, June). Regression model selection using genetic algorithms. *Proceedings of the 11th WSEAS International Conference on Neural Networks and 11th WSEAS International Conference on Evolutionary Computing and 11th WSEAS International Conference on Fuzzy Systems* içinde (s. 19-27). World Scientific and Engineering Academy and Society (WSEAS).
- Pitoyo, A. (2020). A meta-analysis: Factors affecting students' reading interest in Indonesia. *International Journal of Multicultural and Multireligious Understanding*, 7(7), 83-92.
- Rajchert, J. M., Żużtak, T. ve Smulczyk, M. (2014). Predicting reading literacy and its improvement in the polish national extension of the PISA study: The role of intelligence, trait and state-anxiety, socio-economic status and school-type. *Learning and Individual Differences*, 33, 1-11.
- Scrucca, L. (2021). Genetic algorithms. R package version 3.2.2. <https://cran.r-project.org/web/packages/GA/index.html> adresinden erişildi.
- Sever, S. (1995). *Türkçe öğretimi ve tam öğrenme*. Ankara: Anı Yayıncılık.
- Silge, J., Chow, F., Kuhn, M., Wickham, H. ve RStudio. (2021). General resampling infrastructure. R package version 0.1.1. <https://cran.r-project.org/web/packages/rsample/index.html> adresinden erişildi.
- Şen, Z. ve Öztopal, A. (2001). Genetic algorithms for the classification and prediction of precipitation occurrence. *Hydrological Sciences Journal*, 46(2), 255-267.
- Tebekana, J. ve Cishe, E. N. (2015) Socio-cultural background factors affecting the grade 3 learners' acquisition of english literacy (reading) skills in Mthatha Education District of South Africa. *International Journal of Educational Sciences*, 8(3), 557-562.
- Temizyürek, F., Çolakoğlu, B. K. ve Coşkun, S. (2013). Dokuzuncu sınıf öğrencilerinin okuma alışkanlıklarının bazı değişkenler açısından incelenmesi. *Türk Eğitim Bilimleri Dergisi*, 11(2), 114-150.
- Tolvi, J. (2004). Genetic algorithms for outlier detection and variable selection in linear regression models. *Soft Computing*, 8(8), 527-533.
- Torppa, M., Eklund, K., Sulkunen, S., Niemi, P. ve Ahonen, T. (2018). Why do boys and girls perform differently on PISA reading in Finland? The effects of reading fluency, achievement behaviour, leisure reading and homework activity. *Journal of Research in Reading*, 41(1), 122-139.
- Torres, L. R., Ordóñez, G. ve Calvo, K. (2021). Teacher and student practices associated with performance in the PISA reading literacy evaluation. *Frontiers in Education*, 6, 167.
- Torres, R. C. (2019). Factors affecting the reading comprehension of intermediate level learners: Basis for an intervention program. https://www.researchgate.net/profile/Rossana-Torres/publication/351451654_Factors_Affecting_the_Reading_Comprehension_of_Intermediate_Level_Learners_Basis_for_An_Intervention_Program/links/609918dc299bf1ad8d8e381b/Factors-Affecting-the-Reading-Comprehension-of-Intermediate-Level-Learners-Basis-for-An-Intervention-Program.pdf adresinden erişildi.
- Trejos, J., Villalobos-Arias, M. A. ve Espinoza, J. L. (2016). Variable selection in multiple linear regression using a genetic algorithm. *Handbook of research on modern optimization algorithms and applications in engineering and economics* içinde (s. 133-159). IGI Global.
- Uysal, P. K. ve Bilge, H. (2018). An investigation on the relationship between reading fluency and level of reading comprehension according to the type of texts. *International Electronic Journal of Elementary Education*, 11(2), 161-172.

- Vasant, P. M. (2013). *Meta-heuristics optimization algorithms in engineering, business, economics, and finance*. Hershey, PA: IGI Global.
- Wickham, H. ve RStudio. (2020). Modelling functions that work with the pipe. R package version 0.1.8. <https://cran.r-project.org/web/packages/modelr/index.html> adresinden erişildi.
- Williams, J. P. (2003). Teaching text structure to improve reading comprehension. H. L. Swanson, K. R. Harris, & S. Graham (Eds.). *Handbook of learning disabilities* içinde (s. 293–305). New York: The Guilford Press.
- Yalçın, S. (2018). 21. yüzyıl becerileri ve bu becerilerin ölçülmesinde kullanılan araçlar ve yaklaşımlar. *Ankara University Journal of Faculty of Educational Sciences (JFES)*, 51(1), 183-201.
- Yang, C. Y., Jeng, J. T., Chuang, C. C. ve Tao, C. W. (2011, June). Constructing the linear regression models for the symbolic interval-values data using PSO algorithm. *Proceedings 2011 International Conference on System Science and Engineering* içinde (s. 177-181). IEEE.
- Yıldız, M. ve Akyol, H. (2011). İlköğretim 5. Sınıf öğrencilerinin okuduğunu anlama, okuma motivasyonu ve okuma alışkanlıkları arasındaki ilişki. *Gazi Eğitim Fakültesi Dergisi*, 31(3), 793-815.
- Žegklitz, J. ve Pošík, P. (2015, July). Model selection and overfitting in genetic programming: Empirical study. *Proceedings of the Companion Publication of the 2015 Annual Conference on Genetic and Evolutionary Computation* içinde (s. 1527-1528).

Extended Abstract

Introduction

People acquire the knowledge and experience they need both in school and out-of-school educational processes and in their business life through reading (Yalçın, 2018). Reading skills appear as a factor that significantly affects the academic, social and individual development of individuals, as well as how information is accessed and developed (Temizyürek, Çolakoğlu & Coşkun, 2013; Yıldız & Akyol, 2011). While the criterion for the development level of societies was the literacy rate until the recent past, today it has been replaced by the phenomena of reading comprehension and making inferences. This can be proved by the fact that the measurement of reading skills is included in the large-scale exams (PISA, PIRLS) held in international standards (Kurnaz & Yıldız, 2015).

Given that reading skills are an important phenomenon for both individuals and societies, in the current study, it is aimed to investigate the extent to which the variables of gender, grade level, educational status of parents, internet use, language spoken at home, response time to items measuring reading skills, the number and type of books at home, amount of time allocated for reading skills at school and use of digital tools significantly predict students' reading skills by means of the data obtained from the PISA application in which students from different countries participate and the number of observations is high. In this context, it is intended to optimize the best linear regression model in which the predicted variable is reading skills and the predictor variables are the ones mentioned above through the genetic algorithms model, which has yielded good results in the previous studies focused on the selection of the best model (variables) in regression analyses (Leardi, Boggia, & Terrile, 1992; Minghua, Qingxian, Benda, & Feng, 2017; Paterlini & Minerva, 2010; Tolvi, 2004; Trejos, Villalobus-Arias & Espinoza, 2016). In other words, it is intended to perform a feature selection process for the regression model, which consists of the variables that best predict reading skills. In this connection, an answer to the question "What are the variables best predicting students' reading skills and the features of these variables for the study group constituted by using the student data derived from the PISA 2015 application?" was sought.

Method

The study group of the current study was formed on the basis of choosing students who answered the multiple-choice items related to reading skills in the same order, using the data obtained from the PISA 2015 application. For this process, a form of the 36th test booklet, which was answered with a computer-based test application, was used, as it contained conditions suitable for

the purpose of the research, such as being the booklet that students answered the most and recording of the response times. A total of 9396 students answered this form. As a result of deleting the missing data with the listwise deletion method in the data set in which the predicted and predictor variables of the study were brought together, the research continued with the data of a total of 5232 students, of which 2699 were female (51.6%) and 2533 were male (48.4%).

Reading skills, which are the predicted variable of the current study, represent the sum of the students' raw scores obtained from their answers given to the 19 multiple-choice items in the 36th booklet in the PISA 2015 application. While calculating the raw scores, each correct answer was assigned 1 point while each wrong or blank answer was assigned 0 point. The predictor variables of the current study were listed as gender, use of the internet at home, the number of e-book readers, variety and number of books at home, language used at home, class hours allocated to the teaching of reading skills at school, mother's education level, father's education level, grade level and amount of total time allocated for answering the questions.

The regression model, which was created to produce a solution to the problem situation of the current study, was analyzed on the basis of the genetic algorithm method. Since the genetic algorithms method does not require the assumptions such as normality, linearity and homogeneity required by parametric statistical methods (Altunkaynak, 2009; Şen & Öztopal, 2001), no assumption tests were applied for the analysis. The data were transferred to the R program environment and the analysis processes were carried out on this platform.

Result and Discussion

In machine learning-based methods such as the genetic algorithms method, the data set is divided into two subgroups as training and testing in order to avoid overfitting problems, which are explained as the problem of producing biased errors for predictions (Kushchu, 2002; Žeglitz & Pošik, 2015). For this reason, the analyses were carried out by dividing the data into two; 70% as training set ($n=3663$) and 30% as testing set ($n=1569$). The regression model was analyzed using the training set data, then the testing and the whole data set were used to determine the performance level of the model for different data groups. In the analysis performed by the genetic algorithms method, genetic operations based on binary coding were applied with the initial population size of 50, the maximum number of generations of 100, the probability of mutation operation of .1, the probability of crossover operation of .8 and the number of elitism of 2. As the fit (objective) function, the function that makes the BIC criterion the least was used as the model fit criterion in order to select the variables that enable the determination of the best regression model (Minghua et al., 2017; Trejos et al., 2016).

According to the results obtained, the variables that predicted the students' reading skills significantly were listed as follows; the number of books at home, total time allocated to answering the items, use of the internet at home, the number of book types at home, gender, language spoken at home, father's education level and the number of e-book readers. The performance of the regression model created on the training data set was investigated with R^2 , RMSE and MAE metrics for the training set, testing set and the whole data set. It was observed that the performance metrics values obtained for the training set were close to the values obtained for the testing data set and for the whole data set with different characteristics, so the model created was found to have exhibited similar performances in different data structures. In addition, considering the value obtained on the training set, it was understood that the variables that predicted the students' reading skills significantly explained 15.27% ($R^2=.1527$) of the variance in reading skills.

It has been observed that the results obtained in the current study are largely consistent with the results of similar studies in the literature. In this regard, it can be argued that the variation in the variables (such as gender, use of the internet at home, language spoken at home, the number of e-book readers, father's education level, the speed at which they answered the items measuring reading skills and the number and types of books at home) can be said to lead to a variation in students' reading skills.