

## Programlama eğitiminde öğrenci performansının makine öğrenmesi algoritmaları ile tahminlenmesi

Aykut DURAK<sup>1\*</sup>, Vahide BULUT<sup>2</sup>

<sup>1</sup> Yazılım Mühendisliği, İzmir Katip Çelebi Üniversitesi, İzmir, Türkiye

<sup>2</sup> Yazılım Mühendisliği, İzmir Katip Çelebi Üniversitesi, İzmir, Türkiye

\*(aykdur78@gmail.com)

**Özet** – Bu çalışmanın amacı, bilgisayar programlama eğitimi alan öğrencilerin programlama performans düzeylerini tahmin ve sınıflamaya dayalı makine öğrenmesi modelleri oluşturularak tahmin edilmesi amaçlanmaktadır. Bu çalışmanın katılımcılarını, programlama eğitimi alan bireyler oluşturmaktadır. Bu çalışmada kullanılan veri seti öz bildirim dayalı veri toplama araçlarıyla elde edilmiştir. Veri seti, Türkiye'nin tüm bölgelerinden geniş bir temsil düzeyi gözetilerek çevrimiçi veri toplama araçları aracılığıyla elde edilmiştir. Çeşitli algortimalar denenerek elde edilen en yüksek doğruluk Lojistik Regresyon ile %79 olmuştur.

**Anahtar Kelimeler** – Programlama Eğitimi, Programlama Performansı, Bilgi İşlemsel Kimlik, Bilgi İşlemsel Düşünme, Makine Öğrenmesi Algoritmaları

### I. GİRİŞ

Öğrencilerin bilgisayar bilimi ile ilgili kendilerini nasıl gördüklerini açıklayan bilgiişlemsel kimlik, bilgiişlemsel düşünme perspektifi, programlama yetkilendirme ve programlama kaygısının programlama performansı ile ilişkisinin araştırılması önemlidir. Yüksek ve düşük programlama performansı ile ilişkili yapıların açıklanması ve bu faktörlerden hangilerinin en önemli olduğunun belirlenmesi programlama performansının kimler için desteklenebileceğini anlamak ve bu deneyimleri geliştirmeyi kolaylaştırmak için önemlidir ([1], [2], [3]). Bu çalışma, öğrencilerin mevcut durumunu belirlemek, kimlik gelişimini teşvik etmek amacı taşır. Bu çalışma nihayetinde bilgisayar bilimi kariyerleri için ilgi yaratmak için eğitim müdahalelerinin nasıl tasarlanıp uygulanacağına odaklanan gelecekteki çalışmalara ışık tutabilir. Öte yandan, öğrencilerin programlama yeterliklerini geliştirmek, dünyanın çeşitli ülkelerinde bir eğitim hedefi haline gelse de öğrencilerin programlama performansının düzeyini değerlendirmek için yapılmış çalışmalar çok kısıtlıdır.

Tüm bu gerekçeler bağlamında bu araştırmada, bilgisayar programlama eğitimi alan öğrencilerin öz bildirim dayalı veri toplama aracına verdikleri yanıtların makine öğrenmesi algoritmaları kullanarak analiz edilmesi yoluyla programlama performansını etkileyen faktörlerin tahmin edilmesi amaçlanmaktadır.

### II. MATERYAL VE YÖNTEM

#### A. Araştırma modeli, hipotezler ve katılımcılar

Bu araştırmada, programlama eğitimi alan öğrencilerin öz bildirim dayalı veri toplama aracına verilen yanıtların makine öğrenmesi algoritmaları ile analiz edilmesine odaklanılmaktadır. Bu çalışmada programlama performansının tanımlanması ile ilgili olarak makine öğrenmesi algoritmaları yöntemleri üzerinde durulmuştur. Bu çalışmada, düşük ve yüksek programlama performansını etkileyen yapıları incelemek için bir model önerilmektedir.

Bu çalışmanın katılımcılarını, Türkiye'de bulunan lise ve üniversitede öğrenim gören ve en az bir bilgisayar programlama dersi almış olan öğrenciler oluşturmaktadır. Sosyal medya ve hazır

mesaj sistemleri kullanılarak programlama dersi alan öğrencilere ulaşılmıştır.

### B. Veri Toplama Araçları

Bu tez çalışmasında kişisel bilgi formu ve dört ölçek kullanılarak veriler toplanmıştır. Veri toplama araçları programlama dersi alan öğrencilere online ortamda uygulanmıştır.

İlk bölümde kişisel bilgi formu kullanılmıştır. Bu formda 9 soru yer almaktadır. Sorular cinsiyet, yaş, sınıf düzeyi, akademik başarı, öğrencilerin genel akademik not ortalamaları alınarak, programlama dersi başarısı ile ilgilidir.

İlk bölümde Bilgişlemsel Kimlik Ölçeği kullanılmıştır. Bu ölçeğin orijinali Kong ve Wang [4] tarafından geliştirilmiş ve Türkçeye uyarlaması Yıldız-Durak [5] tarafından yapılmıştır. Üçüncü bölümde Programlama Yetkilendirme Ölçeğine yer verilmiştir. Bu ölçek Kong, Chiu, ve Lai [6] tarafından oluşturulan ve Kong ve Lai [7] tarafından yapılandırılmıştır. Türkçeye uyarlaması Yıldız-Durak [5] tarafından yapılmıştır. Dördüncü bölümde, Bilgişlemsel Düşünme Perspektifi Ölçeği kullanılmıştır. Bu ölçek, Kong ve Wang [4] tarafından geliştirilmiştir. Türkçeye uyarlaması Yıldız-Durak [5] tarafından yapılmıştır. Beşinci bölümde Programlama Kaygısı Ölçeği kullanılmıştır. Bu ölçek Yildirim ve Ozdener [8] tarafından geliştirilmiştir.

### C. Veri Seti ve Analizi

Veriler, veri toplama aracı geçerliği ve güvenilirliği sağlanmış ölçeklerdeki Likert tipi maddeler içeren elektronik bir form ile toplanmıştır. Veri toplama araçları için ölçek sahiplerinden ölçek kullanım izni alınmıştır. Veri toplama sürecinde Google online formlar aracılığıyla toplanan veriler elektronik tablolama programına aktarılmıştır.

Bu çalışma, geliştirilen araştırma modelini test etmek için Knime 4.7 yazılımında makine öğrenmesi algoritmaları kullanılarak öğrencilerin programlama performansının tahminlenmesine çalışmıştır. Sınıflandırma algoritmaları kullanılmadan önce verilerin ön inceleme aşaması gerçekleştirilmiştir. Bu aşamada katılımcıların ölçeklerden aldıkları puanlar z puana dönüştürülmüştür. Daha sonra cinsiyet ve eğitim durumu gibi değişkenler kategorik değişkenler olduğundan dolayı dummy değişkene dönüştürülmüştür. Naive bayes sınıflandırıcılar ve

lojistik regresyon kullanılarak modeller oluşturulmuştur.

### III. BULGULAR

Modelin doğruluğu (accuracy), hassasiyeti (precision) ve duyarlılığı (recall) için bulgular Tablo 1 'te sunulmuştur.

Tablo 1. Çıktı değişkenine göre algoritma doğruluk sonuçları

	Doğruluk (accuracy)		Hassasiyet (precision)		Duyarlılık (recall)	
	Düşük	Yüksek	Düşük	Yüksek	Düşük	Yüksek
Alg oritm a /Değışkenler	0.69	0.69	0.77	0.61	0.64	0.74
Nai ve bayes sınıflandırıcı	0.79	0.79	0.78	0.79	0.76	0.81
Lojistik regresyon	0.79	0.79	0.78	0.79	0.76	0.81

Tablo 1 incelendiğinde düşük programlama performans durumlarına göre en yüksek doğruluk sonucuna sahip algoritmanın lojistik regresyon (0.790) algoritması olduğu görülmektedir. Yüksek programlama performans durumlarına göre hesaplanan doğruluk sonuçlarında yine lojistik regresyon (0.790) en doğru sonucu verdiği bulgusuna ulaşılmıştır.

Düşük programlama performans durumlarına göre en yüksek hassasiyet sonucuna sahip algoritmanın lojistik regresyon algoritması (0.789) olduğu görülmektedir. Yüksek programlama performans durumlarına göre yapılan hassasiyet hesaplamalarında en yüksek hassasiyete sahip algoritmanın yine lojistik regresyon algoritması (0.791) olduğu sonucuna ulaşılmıştır.

Düşük programlama performans durumlarına göre en yüksek duyarlılık sonucuna sahip algoritmanın lojistik regresyon algoritması (0.763) olduğu

görülmektedir. Ayrıca yüksek programlama performans durumlarına göre kullanılan algoritmanın duyarlılık sonuçları incelendiğinde en yüksek duyarlılığa sahip algoritmanın lojistik regresyon (0.815) olduğu bulgusuna ulaşılmıştır.

#### IV. TARTIŞMA VE SONUÇLAR

Günümüzde programlama hem bilişsel becerilerin geliştirilmesi hem de kendi başına bir beceri olarak çok önemli hale geldi ([9], [10], [11]).

Dijital çağda her bireyin teknoloji ile etkileşimi programlama öğretime ihtiyacın artmasına sebep olmaktadır [11]. Programlama performansı ile ilişkili yapıları anlamaya yönelik alanyazında çok sayıda çalışma yapılmıştır. Ancak yetkilendirme, kimlik, kaygı gibi değişkenlerle bütüncül bakışa sahip bir çalışmaya alanyazında rastlanmamıştır.

Makine öğrenmesi modelleriyle bu değişkenlerin etkisi değerlendirilerek programlama performansını artırıcı çıkarımlar yapılabilir. Bu nedenle bu araştırmada, programlama dersi alan öğrencilerinin programlama düzeylerini tahmin ve sınıflamaya dayalı makine öğrenmesi modelleri oluşturularak tahmin edilmesi amaçlanmaktadır. Bu amaç doğrultusunda, öğrencilerin çeşitli değişkenlere göre durumları çeşitli sınıflandırma algoritmalarıyla tahminlenmiştir. Sınıflandırma algoritmaları çeşitli parametrelerle çalıştırılarak en yüksek performans elde edilmeye çalışılmıştır. En yüksek doğruluk oranı %79 oranı ile lojistik regresyon algoritması ile elde edilmiştir.

#### TEŞEKKÜR

Bu çalışma ikinci yazarın danışmanlığında ilk yazarın yüksek lisans tezi bağlamında gerçekleştirilmiştir.

#### KAYNAKLAR

- [1] Y. B., Kafai, & Q.Burke, "Computational participation: Teaching kids to create and connect through code". Emerging research, practice, and policy on computational thinking, 2017.393-405.
- [2] S. C., Kong, & T. Lin, "High achievers' attitudes, flow experience, programming intentions and perceived teacher support in primary school: A moderated mediation analysis". *Computers & Education*, 190, 2022, 104598.
- [3] H. Y., Wang, I., Huang, & G. J. Hwang, "Comparison of the effects of project-based computer programming activities between mathematics-gifted

students and average students". *Journal of Computers in Education*, 3, 2016. 33-45.

- [4] S. C., Kong, & Y. Q. Wang, "Formation of computational identity through computational thinking perspectives development in programming learning: A mediation analysis among primary school students". *Computers in Human Behavior*, 106, 2020, 106230.
- [5] H. Yildiz-Durak, "The Model for Middle School Students' Computational Identity ". *Education and Information Technologies*. 2023.
- [6] S. C., Kong, M. M., Chiu, & M. Lai, "A study of primary school students' interest, collaboration attitude, and programming empowerment in computational thinking education". *Computers & Education*, 127, 2018, 178-189.
- [7] S. C., Kong, & M. Lai, "Computational identity and programming empowerment of students in computational thinking development". *British Journal of Educational Technology*. 2021.
- [8] O. G. Yıldırım, & N.Özdener, "The Development and Validation of the Programming Anxiety Scale". *International Journal of Computer Science Education in Schools*, 5(3), 2022. 17-34.
- [9] H.,Yildiz Durak, & M.Saritepeci, "Analysis of the relation between computational thinking skills and various variables with the structural equation model". *Computers & Education*, 116, 2018, 191-202.
- [10] H. Yildiz Durak, "The effects of using different tools in programming teaching of secondary school students on engagement, computational thinking and reflective thinking skills for problem solving". *Technology, Knowledge and Learning*, 25, 2020,179-195.
- [11] H.,Yildiz Durak, M., Saritepeci, & A. Durak, "Modeling of relationship of personal and affective variables with computational thinking and programming". *Technology, Knowledge and Learning*, 28(1), 2023, 165-184.