

Eđitim Teknolojisi

kuram ve uygulama

Kiř 2019

Cilt 9

Sayı 1

Winter 2019

Volume 9

Issue 1

Educational Technology

theory and practice

ISSN: 2147-1908

Cilt 9, Sayı 1, Kış 2019
Volume 9, Issue 1, Winter 2019

Genel Yayın Editörü / Editor-in-Chief: **Dr. Halil İbrahim YALIN**
Editör / Editor: **Dr. Tolga GÜYER**

Basım Editörü / Publisher Editor: **Dr. Tolga GÜYER**
Redaksiyon / Redaction: **Mertcan ÜNAL, Dr. Burcu BERİKAN, Figen DEMİREL UZUN, Akça Okan YÜKSEL**
Dizgi / Typographic: **Dr. Tolga GÜYER**
Kapak ve Sayfa Tasarımı / Cover and Page Design: **Dr. Bilal ATASOY**
İletişim / Contact Person: **Dr. Tolga GÜYER**

Dizinlenmektedir / Indexed in: **ULAKBİM Sosyal ve Beşerî Bilimler Veritabanı (TR-Dizin), Türk Eğitim İndeksi, Sosyal Bilimler Atıf Dizini**

ETKU Dergisi **2011 yılından itibaren yılda iki defa** düzenli olarak yayınlanmaktadır.
Educational Technology Theory and Practice Journal is published regularly **twice a year since 2011.**

Editör Kurulu / Editorial Board*

Dr. Ana Paula Correia
Dr. Buket Akkoyunlu
Dr. Cem Çuhadar
Dr. Deniz Deryakulu
Dr. Deepak Subramony

Dr. Feza Orhan
Dr. H. Ferhan Odabaşı
Dr. Hafize Keser
Dr. Halil İbrahim Yalın
Dr. Hyo-Jeong So

Dr. Kyong Jee(Kj) Kim
Dr. M. Yaşar Özden
Dr. Özcan Erkan Akgün
Dr. S. Sadi Seferoğlu
Dr. Sandie Waters

Dr. Servet Bayram
Dr. Şirin Karadeniz
Dr. Tolga Güyer
Dr. Trena Paulus
Dr. Yavuz Akpınar
Dr. Yun-Jo An

* Liste isme göre alfabetik olarak oluşturulmuştur. / List is created in alphabetical order

Hakem Kurulu / Reviewers*

Dr. Abdullah Kuzu
Dr. Adile Aşkim Kurt
Dr. Agah Tuğrul Korucu
Dr. Arif Altun
Dr. Aslıhan İstanbullu
Dr. Aslıhan Kocaman Karoğlu
Dr. Ayça Çebi
Dr. Ayfer Alper
Dr. Aynur Kolburan Geçer
Dr. Ayşegül Bakar Çörez
Dr. Bahar Baran
Dr. Barış Sezer
Dr. Berrin Doğusoy
Dr. Betül Özeydin
Dr. Bilal Atasoy
Dr. Burcu Berikan
Dr. Çelebi Uluyol
Dr. Demet Somuncuoğlu Özerbaş
Dr. Deniz Atal Köysüren
Dr. Deniz Mertkan Gezgin
Dr. Ebru Kılıç Çakmak
Dr. Ebru Solmaz
Dr. Ekmel Çetin
Dr. Emin İbili
Dr. Emine Aruğaslan
Dr. Emine Cabı
Dr. Emine Şendurur
Dr. Engin Kurşun
Dr. Erinç Karataş
Dr. Erhan Güneş
Dr. Erkan Çalişkan
Dr. Erkan Tekinarslan
Dr. Erman Yükseltürk

Dr. Erol Özçelik
Dr. Ertuğrul Usta
Dr. Esmâ Aybike Bayır
Dr. Esra Yecan
Dr. Fatma Bayrak
Dr. Fatma Keskinçelik
Dr. Fezile Özdamlı
Dr. Filiz Kalelioğlu
Dr. Filiz Kuşkaya Mumcu
Dr. Funda Erdoğan
Dr. Gizem Karaoğlu Yılmaz
Dr. Gökçe Becit İşçitürk
Dr. Gökhan Akçapınar
Dr. Gökhan Dağhan
Dr. Gülfidan Can
Dr. H. Ferhan Odabaşı
Dr. Hafize Keser
Dr. Halil Ersoy
Dr. Halil İbrahim Akyüz
Dr. Halil İbrahim Yalın
Dr. Halil Yurdugül
Dr. Hanife Çivril
Dr. Hasan Çakır
Dr. Hasan Karal
Dr. Hatice Durak
Dr. Hatice Sancar Tokmak
Dr. Hüseyin Bicen
Dr. Hüseyin Çakır
Dr. Hüseyin Özçınar
Dr. Hüseyin Uzunboylu
Dr. Işıl Kabakçı Yurdakul
Dr. İbrahim Arpacı
Dr. İlknur Resioğlu

Dr. Kerem Kılıçer
Dr. Kevser Hava
Dr. M. Emre Sezgin
Dr. M. Fikret Gelibolu
Dr. Mehmet Akif Ocak
Dr. Mehmet Barış Horzum
Dr. Mehmet Kokoç
Dr. Mehmet Üçgül
Dr. Melih Engin
Dr. Meltem Kurtoğlu
Dr. Muhittin Şahin
Dr. Mukaddes Erdem
Dr. Murat Akçayır
Dr. Mustafa Sarıtepeci
Dr. Mustafa Serkan Günbatar
Dr. Mustafa Yağcı
Dr. Mutlu Tahsin Üstündağ
Dr. Müge Adnan
Dr. Nadire Çavuş
Dr. Necmi Eşgi
Dr. Nezhil Önal
Dr. Nuray Gedik
Dr. Nurettin Şimşek
Dr. Onur Dönmez
Dr. Ömer Faruk İslim
Dr. Ömer Faruk Ursavaş
Dr. Ömür Akdemir
Dr. Özcan Erkan Akgün
Dr. Özden Şahin İzmirli
Dr. Özlem Baydaş
Dr. Özlem Çakır
Dr. Ramazan Yılmaz
Dr. Recep Çakır

Dr. Salih Bardakçı
Dr. Sami Acar
Dr. Sami Şahin
Dr. Selay Arkün Kocadere
Dr. Selçuk Karaman
Dr. Selçuk Özdemir
Dr. Serap Yetik
Dr. Serçin Karataş
Dr. Serdar Çiftçi
Dr. Serkan Şendağ
Dr. Serkan Yıldırım
Dr. Serpil Yalçınalp
Dr. Sibel Somyürek
Dr. Soner Yıldırım
Dr. Şafak Bayır
Dr. Şahin Gökçearslan
Dr. Şeyhmus Aydoğdu
Dr. Tarık Kışla
Dr. Tayfun Tanyeri
Dr. Turgay Alakurt
Dr. Tolga Güyer
Dr. Türkan Karakuş
Dr. Uğur Başarmak
Dr. Ümmühan Avcı Yücel
Dr. Ünal Çakıroğlu
Dr. Veynel Demirer
Dr. Vildan Çevik
Dr. Yalın Kılıç Türel
Dr. Yasemin Demirarslan Çevik
Dr. Yasemin Gülbahar
Dr. Yasemin Koçak Usluel
Dr. Yusuf Ziya Olpak
Dr. Yüksel Göktaş

* Liste isme göre alfabetik olarak oluşturulmuştur. / List is created in alphabetical order.

İletişim Bilgileri / Contact Information

İnternet Adresi / Web: <http://dergipark.gov.tr/etku>
E-Posta / E-Mail: tguyer@gmail.com
Telefon / Phone: +90 (312) 202 17 38

Makale Geçmişi / Article History

Alındı/Received: 26.10.2018

Düzeltilme Alındı/Received in revised form: 25.12.2019

Kabul edildi/Accepted: 10.01.2019

**EĞİTSEL VERİ MADENCİLİĞİ VE ÖĞRENME ANALİTİKLERİ BAĞLAMINDA
E-ÖĞRENME VERİLERİNDE AYKIRI GÖZLEMLERİN BELİRLENMESİ***

Sinan Keskin¹, Furkan Aydın², Halil Yurdugül³

Öz

E-öğrenme teknolojilerinin sağladığı en önemli yararlarından birisi de öğrenme sürecindeki etkileşim verilerinin kayıt edilmesidir. Bu verilere dayalı örüntülerin keşfedilmesi (eğitsel veri madenciliği), keşfedilen örüntülerin öğrenme sürecine ilişkin iyileştirmelerde ve aynı zamanda öğretim tasarımında kullanımı (öğrenme analitikleri) günümüzde önemli bir konu olarak ortaya çıkmıştır. Ancak kayıt edilen her veri sağlıklı bir öğrenme verisi anlamına gelmemektedir. Bu nedenle analiz aşamasından önce aykırı gözlemlerin belirlenmesi ve düzeltmelerin yapılması doğru sonuçlara ulaşılmasında önemli bir yere sahiptir. Aykırı gözlemler, verilerin oluşma sürecinde (real-time) belirlenebileceği gibi süreç sonunda elde edilen veri kümelerinden de belirlenebilmektedir. Bu çalışmada bir e-öğrenme ortamından elde edilen eğitsel veriler üzerinde aykırı gözlem belirleme yöntemlerinin kullanımı ele alınmıştır. Araştırmada bir ders döneminde kullanılan Moodle öğrenme yönetim sistemi (ÖYS) log kayıtları veri kümesi olarak kullanılmıştır. Veri kümesi, 65 öğrencinin hiper-metin, video, değerlendirme, SCORM ve forum etkileşimlerine ilişkin toplam etkileşim süresinden oluşmaktadır. Aykırı gözlem verilerinin belirlenmesinde Z, Grubbs, Rosner, kutu grafiği ve Hampel yöntemi kullanılmıştır. Bu çalışmada aykırı gözlem verileri hazır paket programlar kullanılmadan hesaplama çizelgeleri üzerinden işlemler yapılarak belirlenmiştir. Yapılan analizlerin sonucunda yöntemlere göre aykırı (anormal) gözlem sayılarının değiştiği görülmüştür. Buradan elde edilen deneyimler ve veri tabanı yapısı göz önünde bulundurulduğunda Z yöntemi ve kutu grafiği yöntemlerinin bir e-öğrenme sisteminde uygulama anında aykırı gözlemlerin tespiti amacıyla diğer yöntemlere göre daha kolay uygulanabilir olduğu, bir başka ifadeyle bu yöntemlerin makineye öğretiminin daha işlevsel olduğu görülmüştür. Bununla birlikte diğer yöntemlerin ise bir hipotez sınaması gerektirmesi ve daha duyarlı sonuçlar vermesi yönünden önemli bir avantaja sahip olduğu göz önünde bulundurulmalıdır.

Anahtar Kelimeler: e-öğrenme, aykırı gözlem, veri ön işleme, öğrenme analitikleri, eğitsel veri madenciliği

* Bu çalışmanın bir kısmı, 12. Uluslararası Bilgisayar ve Öğretim Teknolojileri Eğitimi sempozyumunda özet bildiri olarak sunulmuştur.

¹ Araş.Gör., Van Yüzüncü Yıl Üniversitesi/Eğitim Fakültesi/BÖTE Bölümü, sinan.keskin@hacettepe.edu.tr, orcid.org/0000-0003-0483-3897

² Öğr.Gör., Kahramanmaraş Sütçü İmam Üniversitesi/Göksun MYO/Bilgisayar Teknolojileri Bölümü., furkanaydin@live.com, orcid.org/0000-0003-2471-9725

³ Prof.Dr., Hacettepe Üniversitesi/Eğitim Fakültesi/BÖTE Bölümü, yurdugul@hacettepe.edu.tr, orcid.org/0000-0001-7856-4664

THE DETERMINING OF OUTLIERS ON E-LEARNING DATA IN THE CONTEXT OF EDUCATIONAL DATA MINING AND LEARNING ANALYTICS

Abstract

In the process of learning analytics, the determination of outliers and making smoothing before the analysis stage has an important place in reaching the right patterns. The outliers can be determined in the real-time, as well as, at the end of the data collection process. In this study, the use of outlier detection methods is discussed using educational data from an e-learning environment. Also, the methods were tested on a real-time system. The Moodle, Learning Management System (LMS) log records were used as the data set. The study group consists of 65 students. In this study, the total interaction times in hypertext, video, assessment, SCORM, and forum themes were used as data set. Box-plot, Z, Grubbs, Rosner and Hampel methods were used to determine the outliers. Outliers are determined by processing through manual calculations without using the existing packaged software. At the same time, in order to evaluate integrability of these methods into the e-learning environment, some PHP script examples are coded by researchers. As a result of analyzes, it was shown that outlier numbers changed according to the methods. When the experiences obtained therefrom and database structure are considered; Z and Box-Plot methods are easier to implement in e-learning systems, for the real-time outlier detection than other methods. In other words, it has been seen that these methods are more functional in machine teaching. However, it should be noted that other methods have significant advantages, for that they require hypothesis test and give more sensitive results. In the context of machine learning, the positive and negative characteristics of these methods are discussed.

Keywords: e-learning, outlier, data preprocessing, anomaly detection, learning analytics, educational data mining

Summary

One of the most important opportunities provided by e-learning is to keep students' interaction data in the system database. Discovery of patterns based on these data (educational data mining) and using these patterns in instruction and learning design (learning analytics) has emerged as an important subject today. The discovery of patterns consists of a series of processes that are defining the problem, data selection, data preprocessing, transformation, modeling, mining, intervention-evaluation (Han & Kanber, 2006). One of the most important stages of this process is pre-processing and outlier detection. If the outliers that represent abnormal behaviors (anomalies) cannot be determined on time, the patterns obtained from this metrics will contain incorrect knowledge. In this study, the use of outlier detection methods is discussed using educational data from an e-learning environment. Also, the methods were tested on a real-time system. In the context of machine learning, the positive and negative characteristics of these methods are discussed.

In the field of statistics; the data representing the behaviors are called normal as they are close to the average. Abnormal behaviors point out the behaviors which are non-standard and different from expected. From this point of view; outliers can be defined as data that does not fit the data set, that affect the distribution of the data set and the analysis results (Grubbs, 1969; Moore & McCabe, 1999). Numerous methods and techniques have been developed for the anomaly or outlier detection. Some of these are based on statistical methods while others

are based on data mining. Statistical methods can be considered as a) criteria, b) hypothesis testing and c) graphical methods. The methods used in this research; Z and Hampel are criteria based, box-plot is graphic-based, Grubbs and Rosner are based on hypothesis testing. These five outlier detection methods are explained in detail in the following sections.

Moodle learning management system (LMS) log records were used as data set. The online teaching design of the course, which is used within the scope of the study, have been prepared in 5 themes. These themes are called hypertext, video, SCORM, forum, and assessment. Hypertext, video, and SCORM are used to present the contents of the online course. The Forum was used by the learners to interact with friends and lecturer. In the e-assessment theme, self-assessment activities were prepared for each unit. Students have the opportunity to test their knowledge by using these activities. The dataset consists of log records of 65 students using Moodle LMS during a course period. The user interaction data in the Moodle LMS database were examined and it was observed that most of the outliers were in the time variables. According to this, the time spent in five interaction themes is considered as the data set of this research. Z, Grubbs, Hampel, Box-plot and Rosner method were used to determine the outliers. Outliers are determined by making transactions on the calculation charts without the use of package programs. Firstly, learners who are not involved in theme interactions are excluded from the study. For each interaction variable, outlier detection methods, described in detail in the introduction of the research, have been applied.

As a result of the analysis performed, it was seen that by using different statistical outlier detection methods, outliers can be determined successfully in e-learning interaction data. Accordingly, it can be said that these methods commonly used in the field of statistics are also applicable to e-learning data. At the same time, in order to evaluate integrability of these tests into the e-learning environment, some PHP script examples are coded by researchers. As a result of analyzes, it was shown that outlier numbers changed according to the tests. When the experiences obtained therefrom and database structure are considered; Z test and Box-Plot methods are easier to implement in e-learning systems, for the real-time outlier detection than other methods. In other words, it has been seen that these methods are more functional in machine teaching. However, it should be noted that other tests have significant advantages, for that they require hypothesis test and give more sensitive results. This study also discusses possible methods and precautions for preventing outliers.

Giriş

E-öğrenme sürecinde kullanılan teknolojilerin sağladığı en önemli olanaklardan biri öğrencilerin katılım ve/veya etkileşim verilerinin sistem veri tabanında saklanmasıdır. Bu verilere dayalı örüntülerin keşfedilmesi (eğitsel veri madenciliği), keşfedilen örüntülerin öğrenme sürecine ilişkin iyileştirmelerde ve aynı zamanda öğretim tasarımında kullanımı (öğrenme analitikleri) günümüzde önemli bir konu olarak ortaya çıkmıştır. Örüntülerin keşfedilme süreci; problem tanımlama, veri seçimi, veri ön işleme, dönüştürme, modelleme, analiz, değerlendirme ve sunum olarak ele alınan bir dizi işlemde oluşmaktadır (Han & Kanber, 2006). Bu sürecin en önemli aşamalarından birisi de veri ön-işleme (pre-procesing) ve aykırı verilerin keşfedilmesidir. Çünkü öğrencilere ilişkin e-öğrenme ortamından elde edilen veriler (kirli veriler içerdiğinden) doğrudan analize sokulmamalı, bir veri ön-işleme sürecine tabi tutulması gerekmektedir. Özellikle sistemden kaynaklı hesaplama hataları ve sistemi yanıltmayı çalışan kullanıcılar, aykırı gözlem verilerini oluşturmaktadır. Öğrenme analitiklerinin kullanıldığı e-öğrenme uygulamalarında gerek oyunlaştırma öğeleri (örneğin lider tablosu ya da rozetlendirme) ve gerekse gezinim verilerinin başarıyı yordanmasında kullanıldığını bilen öğrenciler aykırı etkileşim verilerinin ortaya çıkmasına neden olmaktadır. Öğrenme ortamında gözlenen sosyal rekabet, öğrencileri aldatmaya yönelik davranışlara yöneltebilmektedir (Cantador & Conde, 2010; Orosz, Farkas, & RolandLevy, 2013). Örneğin; lider tablosu sıralaması oluşturulurken oturum süresi ve video izleme süresinin önemli bir etken olduğunu fark eden bazı öğrenciler sistemi bu yönde yanıltmaya yönelik çeşitli girişimlerde bulunabilmektedir. Anormal davranışlara (anomalies) ilişkin ortaya çıkan aykırı değerlerin (outlier) zamanında belirlenip önlenememesi durumunda metriklerden elde edilecek örüntülerin yanlış bilgiler içermesi söz konusu olacaktır. Bu çalışmada bir e-öğrenme uygulamasında karşılaşılan anormal durumlar ele alınmış, bunların önlenmesi konusunda gerekli yöntemler örneklendirilmiş ve makine öğrenmesi boyutunda bu aykırı gözlemlerle baş etme durumları tartışılmıştır.

E-Öğrenme, Etkileşim Verileri ve Öğrenme Analitikleri

E-öğrenme, bireylere öğrenme fırsatları sunan uygulama, web sitesi gibi teknolojilerden yararlanılarak gerçekleştirilen öğrenme olarak tanımlanabilir (Moore, Seane, & Galyen, 2011). Bu bağlamda öğrenme yönetim sistemleri, (ÖYS) günümüzde en yaygın kullanılan e-öğrenme ortamlarından birisidir. Web 1.0 döneminde yapılandırılan bu tür sistemler ÖYS 1.0 olarak adlandırılmıştır. Web 2.0 ile etkileşimin ön plana çıktığı teknolojiler döneminde tasarlanan sistemler ise ÖYS 2.0 olarak tanımlanmıştır. ÖYS 2.0'ın önemli bir özelliği de; e-öğrenme ortamlarının kullanımı esnasında kullanıcı-sistem etkileşimlerine ilişkin bir dizi verileri saklamasıdır. Saklanan bu verilerin işe koşulmasına yönelik yeni nesil ÖYS'ler ise ÖYS 3.0 olarak tanımlanmaktadır (Şahin, Yurdugül; 2018). Bu veriler; genel olarak kullanıcının ne tür içeriklerde gezindikleri, farklı türdeki sayfalarda ne kadar zaman geçirdikleri, tıklanma sayısı, e-değerlendirme sayısı-süresi, kullanıcı-kullanıcı (öğrenci-öğrenci / öğrenci-öğretmen) etkileşim sayısı vb. metriklerden oluşmaktadır. Ortaya çıkan bu kapsamlı etkileşim verileri alanyazında büyük veri olarak adlandırılmaktadır (Siemens, 2013). Öğrenme ve öğretme süreçlerinin anlaşılması ve iyileştirilmesi amacıyla öğrenme analitiklerine başvurulur. Öğrenme analitikleri ise süreç olarak; öğrencilere ilişkin a) etkileşim verilerine erişim, b) verilerin yapılandırılması, c) verilerin analizleri, d) verilerin görselleştirilmesi ve e) öğrenene veya sisteme müdahale aşamalarını içerir (Ferguson, 2012; LAK 2011; Siemens, 2013; Lal, 2014). Buradan da anlaşılacağı üzere öğrenme analitikleri çok aşamalı ve çok boyutlu bir süreçtir. Bu süreçte

doğru bulgular ve bilgilerin elde edilebilmesi için de temiz ve güvenilir verilerin işleme alınması gereklidir. Bu çalışmada, öğrenme analitiği sürecinde kirli verilerin (aykırı veriler) tespitinde kullanılan yöntemlerin incelenmiş ve aykırı gözlemlerin önlenmesi üzerine yaklaşımlar ele alınmıştır.

Anormal Davranışlar ve Aykırı Veriler

Anormal davranışlar; standart, normal ya da beklenenden farklı olanı işaret eder. İstatistik alanında, davranışları temsil eden veriler ortalamaya yaklaştıkça bu davranışlar normal olarak adlandırılmaktadır. Buna göre davranışlar ya da davranışları temsil eden veriler ortalamadan uzaklaştıkça anormallik yani aykırılık da artmaktadır. Buradan hareketle aykırı gözlem verileri diğer veriler ile karşılaştırıldığında veri kümesine uymayan, veri kümesinin dağılımını ve analiz sonuçlarını etkileyen veriler olarak tanımlanabilir (Grubbs, 1969; Moore & McCabe, 1999). Bu tür veriler ölçeklerin özensiz bir şekilde doldurulması, sistemsel hatalar, kullanıcının yanıltmaya yönelik girişimleri sonucunda oluşabilmektedir. Tanımdan da anlaşılacağı üzere bu tür veriler veri kümesinin dağılımını ve analiz sonuçlarını etkileyen verilerdir. Bu nedenle veri analizi gerçekleştirilmeden önce bu verilerin veri kümesinden çıkarılması veya gerekli düzeltmelerin (smoothing) yapılması gerekmektedir.

Anormal Davranışlar ve Aykırı Gözlem Belirleme Yöntemleri

Anomali ya da aykırı gözlem belirleme konusunda çok sayıda yöntem ve teknik geliştirilmiştir. Bunların bir kısmı istatistiksel yöntemlere dayalı iken diğerleri ise veri madenciliğine dayalı yöntemlerdir. İstatistiğe dayalı yöntemler a) ölçüt, b) hipotez sınaması ve c) grafik yöntemler olarak ele alınabilir. Ölçüte dayalı yöntemlerde her bir gözlem için standart bir değer hesaplaması yapılarak ilgili değer in önceden belirlenen ölçüt aralığında yer alıp almadığı kontrol edilir. Hipotez sınaması gerektiren yöntemlerde ise her bir gözlem için yonteme ilişkin eşitlikler kullanılarak bir değer hesaplaması yapılır. Hesaplanan bu değer tablo değerleri (kritik bölge) ile karşılaştırılarak bir karara ulaşılır. Grafik yönteminde ise çeşitli katsayı hesaplamaları yapılarak veri kümesinin dağılımı, eğilimini, basıklık ve çarpıklığını temsil eden bir grafik çizimi gerçekleştirilir. Grafik üzerinde belirlenen alt ve üst limiti aşan değerler aykırı gözlem olarak işaretlenir. Bu araştırmada aykırı gözlem belirlemeye yönelik her üç yaklaşımdan yöntemler ele alınmıştır. Bu yöntemlerden; Z ve Hampel yöntemi ölçüte; Grubbs ve Rosner hipotez sınamasına; kutu grafikleri ise grafiğe dayalı aykırı gözlem belirleme yöntemleridir. Aykırı gözlem belirlemeye yaklaşımlarının altında yer alan çok sayıda yöntem söz konusudur. Bu çalışmada ilgili yöntemler sınırlandırılıp seçilirken bu yöntemlerin tanınabilirliği, kullanılabilirliği ve kullanım sıklığı ölçüt alınarak indirgenmiştir. Aşağıda aykırı gözlem belirlemede kullanılan bu beş yöntem kısaca açıklanmıştır.

Z Yöntemi

Aykırı gözlem belirlemede çoğu araştırmacının kullandığı en temel yöntemlerden biri olarak ifade edilebilir. Z yönteminde her bir gözleme ilişkin standart bir değer hesaplaması yapılmaktadır (Eşitlik 1). Hesaplanan Z puanı, karar değeri ile karşılaştırılarak gözlemin normal dağılım aralığında yer alıp almadığına karar verilir. Karar değeri ise farklı kaynaklara göre değişmektedir (George, 2011; Tabachnick, & Fidell, 2007). Bu çalışma da ± 1.96 olarak alınmıştır.

$$z_i = \frac{(x_i - \bar{x})}{S} \quad (1)$$

Grubbs Yöntemi

Grubbs aykırı gözlem bulma yöntemi, ortalama ve standart sapmayı temel almaktadır. Yöntemin uygulanabilmesi için veri kümesinin; normal dağılım göstermesi ve en az 3 en fazla 100 gözleme sahip olması gerekmektedir. Grubbs'a dayalı üç alt yöntem bulunmaktadır (Grubbs, 1969, Grubbs, 1972). E-öğrenme ortamlarında aykırı veriler çoğunlukla kullanıcıların sistemi yanıltmaya yönelik sayfayı sürekli yenileme, pencereyi açık tutma vb. davranışları sonucunda sağ kuyrukta oluşmaktadır. Bu nedenle bu çalışma da sadece sağ kuyruktaki aykırı değer(leri) bulmada kullanılan alt yöntem açıklanmıştır.

$$G_{MAX} = \frac{|X_{MAX} - \bar{X}|}{S} \quad (2)$$

Grubbs yöntemi kullanılarak veri kümesi içerisindeki aykırı değerler tek tek belirlenebilmektedir. Sağ kuyrukta yer alan aykırı değerlerin belirlenirken öncelikle veri kümesi içindeki en büyük değer şüpheli değer olarak ele alınır. Şüphelenilen değerden veri kümesinin ortalama değeri çıkartılıp, bulunan değer standart sapmaya bölünerek G_{max} değeri elde edilir (Eşitlik 2). Hesaplanan G_{max} değeri, kritik değer tablosundaki (Grubbs, 1972) değerden büyük ise aykırı değer olarak tanımlanır (Durivage, 2014). Eğer ele alınan gözlem aykırı değer olduğuna karar verilmiş ise bu gözlem veri kümesinden çıkartılır. Yeni veri kümesi için ortalama ve standart sapma değerleri yeniden hesaplanarak veri kümesinde yer alan en büyük değere sahip gözlem, şüpheli gözlem olarak ele alınır. Bu işlem G_{max} değeri tablo değerinden küçük oluncaya dek tekrar edilir.

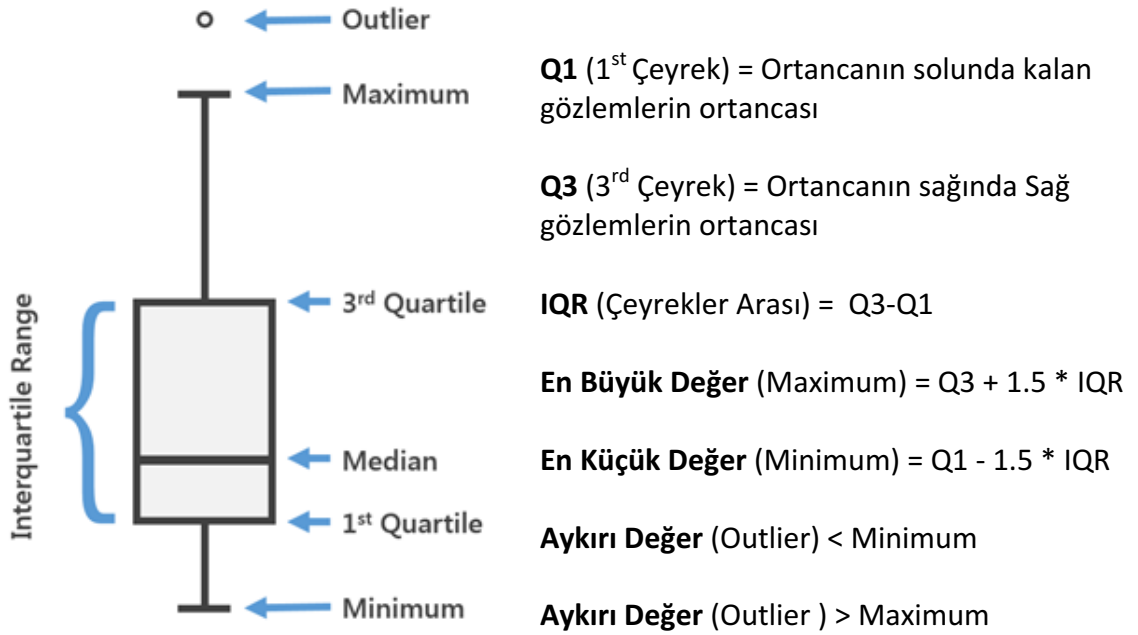
Hampel Yöntemi

Aykırı değerlerin bulunmasında kullanılan bir diğer yöntem, Frank R. Hampel tarafından geliştirilen Hampel yöntemidir (Hampel, 1971; Hampel, 1974). Hampel yönteminin uygulanmasında ortanca (median) ve ortanca mutlak sapma (median absolute deviation, MAD) hesaplamaları kullanılmaktadır. Bu hesaplamalar kullanılarak karar değeri hesaplanır. Hesaplanan karar değerinin üzerinde olan değerler, aykırı gözlem olarak belirlenir (Durivage, 2014).

$$\begin{aligned} MAD &= |x_i - \tilde{x}| \\ \text{Karar Değeri} &= (MAD * 5.2) \end{aligned} \quad (3)$$

Kutu Grafiği (Box-Plot) Yöntemi

Kutu grafikleri (Box plot); veri dağılımı, merkezi eğilim, çarpıklık ve basıklık hakkında araştırmacılara bilgiler sunan ve beş temel hesaplamasının kullanıldığı bir yöntemdir (McGill, Tukey, & Larsen, 1978). Kutu grafiği oluşturulurken en yüksek değer (maximum), en küçük değer (minimum), ortanca (median), 1. çeyrek (1st Quartile) ve 3. çeyrekte (3rd Quartile) hesaplamaları gerçekleştirilir. Veri görselleştirme yöntemlerinden biri olan bu yöntemin en önemli avantajlarından biri araştırmacılara kolay yorumlanabilir görsel bir grafik sunmasıdır. Kutu grafiğinin oluşturulmasına ilişkin hesaplamalar Şekil 1'de verilmiştir.



Şekil 1. Kutu Grafiği Yönteminin Hesaplanması ve Görselleştirilmesi (ArcGIS Pro, 2018)

Kutu grafiği oluşturulurken öncelikle veri kümesi küçükten büyüğe doğru sıralanır. Veri kümesinin ortanca değeri (Q2) belirlenir. Medyanın solunda ve sağında kalan alt veri grupları için de ortanca değerleri bulunur. Bulunan bu ortanca değerleri Q1 (birinci çeyrek) ve Q3 (üçüncü çeyrek) olarak isimlendirilir. Q1 kutunun en alt noktasını Q3 ise en üst noktayı göstermektedir. En büyük ve en küçük değerler Şekil 1’de verilen formül kullanılarak belirlenir. Bu değerlerin ötesinde yer alan değerlere sahip gözlemler aykırı gözlem olarak belirlenir.

Rosner Yöntemi

Rosner yönteminde aykırı değerler gruplar halinde test edilebilir. Rosner yönteminin uygulanabilmesi için veri kümesindeki gözlem sayısının 24’den büyük ve aykırı gözlem(ler) veri kümesinden çıkarıldığında kalan verilerin normal dağılmış olması gerekir (Rosner, 1983). Rosner yöntemi kullanılarak tek seferde 10 gözleme kadar aykırı değer belirlenebilir. Veri kümesindeki her bir şüphelenen gözlem için ayrı ayrı içinde şüpheli değer olmadığı veri setleri oluşturulur. Bu veri setlerine ilişkin standart sapma ve ortalama değerleri hesaplanır. Ardından veri kümesindeki en büyük değerden hesaplanan ortalama değeri çıkarılarak standart sapmaya bölünür (Tablo 1). Yapılan hesaplardan elde edilen değer Rosner’ın kritik değer tablosundaki (Rosner, 1983) değer ile karşılaştırılır. Eğer hesaplanan değer kritik değer tablosundaki değerden büyük ise ilgili gözlem aykırı değer olarak tanımlanır. İlgili gözlem veri kümesinden çıkartılarak bir sonraki en büyük değer için yeniden hesaplamalar yapılır. Bu işlemler veri kümesinde hiçbir aykırı gözlem kalmayana kadar tekrar eder. Hesaplama işlemlerinin formülleri Tablo 1’de verilmiştir.

Tablo 1. Rosner Yönteminin Hesaplanması

Ortalama	$\bar{X}^{(i)} = \frac{X_1 + X_2 + \dots + X_{n-i}}{n-i}$
Standart Sapma	$S^{(i)} = \sqrt{\frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_{n-i} - \bar{X})^2}{n-i}}$
Test İstatistiği	$R_r = \frac{ X^{r-1} - \bar{X} }{S^{r-1}}$

n=gözlem sayısı, i=şüpheli gözlem sayısı

Bu araştırmada, öğrenme analitiği sürecinde e-öğrenme verilerindeki aykırı gözlemlerin belirlenmesinde kullanılacak yöntemlerin ve örnek bir veri kümesi kullanılarak performanslarının incelenmesi amaçlanmıştır. Bu amaca yönelik olarak, bir eğitim öğretim dönemi süresince kullanılan bir e-öğrenme ortamından elde edilen etkileşim verileri ele alınmış ve yukarıda açıklanan aykırı gözlem belirlemede kullanılan istatistiksel beş yöntem ile sınamalar gerçekleştirilmiştir. Gelişen ÖYS 3.0 ile birlikte öğrenme sistemleri öğrenci verilerini anlık olarak işlenmekte ve buna uygun müdahaleler gerçekleştirmektedir. Bu nedenle aykırı gözlemlerin gerçek zamanlı olarak belirlenmesi e-öğrenme sistemleri için önemlidir. Bu araştırmada aykırı gözlemlerin belirlenmesinde kullanılan yöntemlerin performanslarının tartışılmasının yanı sıra bu yöntemlerin birer algoritma olarak makineye öğretilip gerçek zamanlı uygulamalarda kullanımı da ayrıca tartışılmıştır.

Yöntem

Bu çalışmada özellikle e-öğrenme sürecinde ortaya çıkan aykırı gözlem verileri ve bu veriler ile baş etme yöntemleri ele alınıp tartışılmıştır. Bunun için aykırı değer belirleme yöntemlerinin tipik bir e-öğrenme verisi üzerindeki performansları ele alınmış ve daha sonra da makine öğrenmesi bağlamında olumlu ve olumsuz özellikleri tartışılmıştır.

Veri Kümesi

Araştırmada Moodle öğrenme yönetim sistemi (ÖYS) log kayıtları veri kümesi olarak kullanılmıştır. Moodle ÖYS ile öğrencilere farklı türlerde etkinlik ve kaynaklar ders materyali olarak sunulabilmektedir. Örneğin ders içerikleri öğrencilere kitap (hiper-metin), video, SCORM (paketlenmiş öğrenme nesnelere), wiki vb. farklı yöntemler ile sunulabilmektedir. Bu çalışma kapsamında kullanılan dersin öğretim tasarımında beş farklı çevrimiçi öğrenme etkinliği hazırlanmıştır. Bu etkinlikler hiper-metin, video, SCORM, forum ve e-değerlendirme olarak adlandırılmıştır. Hiper-metin, video ve SCORM ders içeriklerinin çevrimiçi ortamda sunulması amacıyla kullanılmıştır. Forum ise öğrenciler tarafından ÖYS üzerinden grup etkileşiminde bulunmak, derse ilişkin tartışmalar yürütmek amacıyla kullanılmıştır. E-değerlendirme ortamında ise her ünite için 10-15 maddelik değerlendirme etkinlikleri hazırlanmıştır. Gönüllü katılımın olduğu e-değerlendirme etkinlikleri ile öğrenciler ünite sonunda bilgilerini sınama fırsatı elde etmektedirler. Araştırmanın veri kümesini bir ders dönemi süresince Moodle ÖYS kullanan 65 öğrencinin log kayıtları oluşturmaktadır. Moodle

ÖYS veri tabanında yer alan kullanıcı etkileşim verileri incelenmiş olup en çok zaman değişkenlerinde aykırılıkların olduğu gözlenmiştir. Nitekim e-öğrenme sistemlerinde öğrenenler tarafından en kolay manipüle edilebilecek değişkenlerin sistemde kalma süreleri olabileceği düşünülerek bu araştırmada özellikle etkileşim ortamlarında harcanan zaman veri seti olarak kullanılmıştır.

Öğrencilerin farklı etkileşim ortamlarında harcadıkları zamana ilişkin betimsel istatistikler Tablo 2’de verilmiştir.

Tablo 2. Etkileşim Ortamlarında Harcanan Zamana İlişkin Betimsel İstatistikler

Etkileşim Ortamı	N	\bar{X}	Ss	Ortanca	Maks
Hiper Metin	61	5454.44	7835.08	2330	37179
Video	64	3437.73	3757.24	1881	15600
Değerlendirme	62	570.27	368.14	509,5	2110
SCORM	56	2703.20	3604.24	1651	18796
Forum	59	933.07	1353.22	427	6733

Tablo 2 incelendiğinde ortalama en çok sürenin içeriğe ilişkin hiper-metin, video ve SCORM ortamlarında harcadığı görülmektedir. Standart sapmalar incelendiğinde ise en küçük standart sapmanın e-değerlendirme ortamı için hesaplandığı görülmektedir. Aynı zamanda e-değerlendirme için hesaplanan ortalama ve ortanca değerlerinin de birbirine yakın olduğu görülmektedir. Buna göre bu ortamdan normal dağılıma daha yakın verilerin elde edildiği sezgisel olarak söylenebilir. Ancak içerik etkileşimlerine ilişkin hesaplanan standart sapma katsayılarının büyük olduğu ve ortanca ile ortalama arasındaki farkın açıldığı görülmektedir.

Aykırı Gözlem Yöntemlerinin Uygulanması

Aykırı gözlem verilerinin belirlenmesinde Z, Grubbs, Hampel, Box-plot ve Rosner yöntemi kullanılmıştır. Bu çalışmada aykırı gözlem verileri hazır paket programlar kullanılmadan hesaplama çizelgeleri üzerinden işlemler yapılarak belirlenmiştir. İlk olarak hiper-metin, video, değerlendirme, SCORM ve forum etkileşimlerine girmeyen öğrenciler çalışma kapsamı dışına alınmıştır. Ardından veriler küçükten büyüğe doğru sıralanmıştır. Ele alınan her bir etkileşim değişkeni için araştırmanın giriş bölümünde detaylı olarak açıklanan aykırı gözlem bulma yöntemleri uygulanmıştır.

Bulgular

Bu bölümde aykırı gözlem belirleme yöntemlerinin ele alınan e-öğrenme verileri üzerinde sınanması sonucunda elde edilen bulgular sunulmuştur. Ele alınan her bir etkileşim değişkeni için Z, Grubbs, Hampel, Box-plot ve Rosner yöntemleri ayrı ayrı uygulanmış ve sırasıyla tablolar halinde raporlanmıştır. E-öğrenme ortamlarında aykırı veriler çoğunlukla kullanıcıların sistemi yanıltmaya yönelik sayfayı sürekli yenileme, pencereyi açık tutma vb. davranışları sonucunda sağ kuyrukta oluşmaktadır. Bu veriler gerçeği yansıtmadığından analiz

sonuçlarının da hatalı olmasına neden olmaktadır. Bu nedenle bu yöntemler özellikle verilerin dağılımının *sağ kuyruğundaki* gözlemler üzerinde test edilmiştir.

Z Yöntemi ile Aykırı Gözlemlerin Tespit Edilmesi

Beş farklı etkileşim değişkenine ilk olarak Z yöntemi uygulanmıştır. Her bir gözleme ilişkin standart z puanı hesaplanmış ve karar değeri ile karşılaştırılarak gözlemin aykırı-anormal olup olmadığına karar verilmiştir. Z yöntemi sonucunda belirlenen aykırı gözlem sonuçları Tablo 3'te verilmiştir.

Tablo 3. Z yönteminin uygulanması sonucunda belirlenen aykırı gözlem sonuçları

Etkileşim Ortamı	N	\bar{X}	S	Karar Değeri	Aykırı Gözlem Sayısı
Hiper Metin	61	5454.44	7835.08	$Z \geq 1.96$	4
Video	64	3437.73	3757.24	$Z \geq 1.96$	4
Değerlendirme	62	570.27	368.14	$Z \geq 1.96$	3
SCORM	56	2703.20	3604.24	$Z \geq 1.96$	4
Forum	59	933.07	1353.22	$Z \geq 1.96$	4

* $Z = Z$ yöntemi sonucunda hesaplanan standart değer

Tablo 3 incelendiğinde 61 öğrencinin hiper-metin etkileşiminde ortalamasının (\bar{X})= 5454.44, standart sapmanın (S)=7835.08 olduğu belirlenmiştir. Uygulanan Z yöntemi sonucunda aykırı gözlem sayısının 4 olduğu gözlemlenmiştir. Video ortamında 64 öğrenciye ilişkin veri kümesinin ortalaması (\bar{X}) = 3437.73, standart sapması (S)=3757.24 olduğu tespit edilmiş ve 4 aykırı gözlemin bulunduğu belirlenmiştir. Değerlendirme etkileşiminde 62 öğrenciye ilişkin veri kümesinin ortalaması (\bar{X})= 570.27, standart sapması (S)=368.14 bulunmuş ve aykırı gözlem olarak 3 değer tespit edilmiştir. SCORM ortamında 56 öğrenciye ilişkin veri kümesinin ortalaması (\bar{X})= 2703.20, standart sapma (S)=3604.24 tespit edilmiş ve 4 aykırı değer gözlemlenmiştir. Son olarak ise Forum etkileşiminde 59 öğrenciye ilişkin veri kümesinin ortalaması (\bar{X})= 933.07, S=1353.22 görülmüş ve 4 aykırı değer tespit edilmiştir.

Grubbs Yöntemi ile Aykırı Gözlemlerin Tespit Edilmesi

İkinci olarak veri kümesine Grubbs yöntemi uygulanmıştır. Veriler analize alınırken küçükten büyüğe doğru sıralandıktan sonra sağ kuyruktan başlanarak en büyük değer şüpheli değer olarak ele alınmış ve her bir şüpheli gözlem için tek tek Grubbs yöntemi uygulanarak G_{max} değeri hesaplanmıştır. Grubbs yönteminde bir seferde tek bir aykırı değer bulunabildiğinden (Grubbs, 1969) aykırı değer olduğuna karar verilen gözlem veri kümesinden çıkarıldıktan sonra tekrar ortalama ve standart sapma hesaplaması yapılmıştır. Hesaplanan G_{max} değeri Grubbs kritik tablosu ($\alpha=0.05$) ile karşılaştırılarak aykırı değer olup olmadığına karar verilmiştir (Grubbs, 1972). Bu işlemlere tüm aykırı gözlemler bulunana kadar devam edilmiş, en son şüpheli gözlemin aykırı gözlem olmadığına karar verildiğinde ise hesaplama işlemi sonlandırılmıştır. Burada dikkat edilmesi gereken bir diğer husus ise Grubbs yönteminde başlangıçta ve son aşamada analize alınan gözlem sayısının (N) farklılık göstermesidir. Başlangıçta analize tüm veri seti alınırken belirlenen her bir aykırı gözlem birer birer veri

setinden çıkarılarak (Tablo 4'te N değerinde verildiği gibi) nihai veri setine ulaşılır. Grubbs yöntemine ilişkin sonuçlar Tablo 4'te verilmiştir.

Tablo 4. Grubbs yöntemi sonucunda belirlenen aykırı gözlem sonuçları

Etkileşim Ortamı	N	X_n	\bar{X}	S	Karar Değeri	Aykırı Gözlem Sayısı
Hiper Metin	61	37179	5454.44	7835.08	$G \geq 3.032$	5
	60	30084	4925.70	6714.63	$G \geq 3.025$	
	59	29227	4499.30	5896.26	$G \geq 3.019$	
	58	21614	4072.95	4946.06	$G \geq 3.013$	
	57	17138	3765.21	4394.26	$G \geq 3.006$	
	56	14940	3526.41	4043.64	$G < 3.000$	
Video	64	15600	3437.73	3757.24	$G \geq 3.049$	3
	63	13775	3244.68	3452.66	$G \geq 3.044$	
	62	12927	3074.84	3204.55	$G \geq 3.037$	
	61	10919	2913.33	2965.81	$G < 3.032$	
Değerlendirme	62	2110	570.27	368.14	$G \geq 3.037$	4
	61	1650	545.03	312.45	$G \geq 3.032$	
	60	1585	526.62	279.72	$G \geq 3.025$	
	59	1276	508.68	244.85	$G \geq 3.019$	
	58	978	495.45	224.71	$G < 3.013$	
SCORM	56	18796	2703.20	3604.24	$G \geq 3.000$	8
	55	10964	2410.60	2889.29	$G \geq 2.992$	
	54	10783	2252.20	2664.50	$G \geq 2.986$	
	53	10471	2091.25	2410.43	$G \geq 2.978$	
	52	8412	1930.10	2126.21	$G \geq 2.971$	
	51	7717	1803.00	1937.62	$G \geq 2.964$	
	50	7148	1684.72	1761.52	$G \geq 2.956$	
	49	6277	1573.22	1591.57	$G \geq 2.948$	
	48	5228	1475.23	1451.35	$G < 2.940$	
Forum	59	6733	933.07	1353.22	$G \geq 3.019$	7
	58	4400	833.07	1123.83	$G \geq 3.013$	
	57	3870	770.49	1026.83	$G \geq 3.006$	
	56	3741	715.14	946.44	$G \geq 3.000$	
	55	3585	660.13	860.06	$G \geq 2.992$	
	54	3477	605.96	767.63	$G \geq 2.986$	
	53	2581	551.79	662.64	$G \geq 2.978$	
	52	2282	512.77	604.49	$G < 2.971$	

* G = Hesaplanan G_{max} değeri

Tablo 4 incelendiğinde hiper-metin ortamında harcanan zamana ilişkin 61 gözlemin bulunduğu görülmektedir. Grubbs yöntemi bu veri kümesine 6 kez uygulanmıştır. İlk beş

uygulamada 61, 60, 59, 58 ve 57 numaralı gözlemlerin aykırı gözlem olduğuna karar verilmiştir ($G_{max} > G_k$). Video değişkeninde 64 gözlem bulunmakta ve bu veri kümesine 4 kez Grubbs yöntemi uygulandığı gözlemlenmektedir. Sağ kuyrukta bulunan 64., 63. ve 62. verilerin aykırı gözlem olduğu tespit edilmiştir ($G_{max} > G_k$). Değerlendirme ortamı ile etkileşimde bulunan 62 öğrenciden 5'inin Grubbs yöntemi sonucunda aykırı gözlem olduğu belirlenmiştir. SCORM ile etkileşimde bulunan 56 öğrenci verisine 9 kez uygulanan Grubbs yöntemi sonucunda 8 gözlemin aykırı değere sahip olduğuna karar verilmiştir ($G_{max} > G_k$). Son olarak ise Forum etkileşimi için 59 gözlemden 8'inin aykırı gözlem olduğu Grubbs yöntemi sonucunda belirlenmiştir ($G_{max} > G_k$).

Hampel Yöntemi ile Aykırı Gözlemlerin Tespit Edilmesi

Üçüncü olarak veri kümesinde yer alan değişkenlere Hampel yöntemi uygulanmıştır. Öncelikle öğrencilere ilişkin etkileşim verileri küçükten büyüğe doğru sıralanmıştır. Ardından değişkenin ortanca (\tilde{x} = median) değeri (1. ortanca) bulunmuştur. Bulunan ortanca değeri her bir gözlem değerinden çıkarılmıştır. Ardından elde edilen değerler küçükten büyüğe sıralanarak ortanca değeri (2. ortanca) belirlenmiştir. İkinci ortanca değeri 5.2 ile çarpılarak karar değeri tespit edilmiştir (Hampel, 1974). Hesaplanan karar değerinin üstünde olan değerler aykırı gözlem olarak belirlenmiştir. Hampel yöntemi sonuçları Tablo 5'te verilmiştir.

Tablo 5. Hampel yöntemi sonucunda belirlenen aykırı gözlem sonuçları

Etkileşim Ortamı	N	\tilde{x}_1	\tilde{x}_2	Karar Değeri ($\tilde{x}_2 * 5.2$)	Aykırı Gözlem Sayısı
Hiper Metin	61	2330	2282	$H \geq 11866.4$	6
Video	64	1881	1627	$H \geq 8460.4$	4
Değerlendirme	62	500	148	$H \geq 769.6$	4
SCORM	56	1640	1502.5	$H \geq 7813$	5
Forum	59	427	392	$H \geq 2038.4$	7

* H = Hampel yöntemi sonucunda hesaplanan değer

Öğrencilerin hiper metin ortamı ile etkileşim sürelerine ilişkin birinci işlem ortancası (\tilde{x}_1) = 2330, ikinci işlem ortancası (\tilde{x}_2)=2282 olarak bulunmuştur. Sağ kuyrukta bulunan 6 veri karar değerinden büyük olduğu için aykırı gözlem olarak belirtilmiştir ($H > H_k = 11886.4$). Video etkileşim sürelerine ilişkin birinci işlem ortancası (\tilde{x}_1) = 1881, ikinci işlem ortancası (\tilde{x}_2)=1627 olarak bulunmuştur. Karar değerinden büyük olan 4 değer aykırı gözlem olarak belirlenmiştir ($H > H_k = 84460.4$). Değerlendirme değişkeni için birinci işlem ortancası (\tilde{x}_1) = 500, ikinci işlem ortancası (\tilde{x}_2)=148 bulunmuştur. Hesaplanan karar değerinden büyük 4 verinin aykırı gözlem olduğuna karar verilmiştir ($H > H_k = 769.6$). SCORM ile etkileşim süresi için birinci işlem ortancası (\tilde{x}_1) = 1640, ikinci işlem ortancası (\tilde{x}_2)=1505.5 bulunmuştur. Sağ kuyrukta bulunan 5 verinin aykırı gözlem olduğu tespit edilmiştir ($H > H_k = 7813$). Son olarak forum değişkeni için birinci işlem ortancası (\tilde{x}_1) = 427, ikinci işlem ortancası (\tilde{x}_2)= 392 bulunmuştur. Forum değişkeni için 7 gözlem karar değerinden büyük olduğundan bu gözlemler aykırı gözlem olarak tespit edilmiştir ($H > H_k = 2038.4$).

Kutu Grafiği Yöntemi ile Aykırı Gözlemlerin Tespit Edilmesi

Dördüncü olarak ele alınan beş değişkene ilişkin kutu grafiği (Box-Plot) hesaplamaları gerçekleştirilmiştir. Hesaplamalar sonucunda en büyük değer üstünde yer alan gözlemler aykırı gözlem olarak belirlenmiştir. Kutu grafiği hesaplamalarına ilişkin sonuçlar Tablo 6'da verilmiştir.

Tablo 6. Kutu grafiği yöntemi sonucunda belirlenen aykırı gözlem sonuçları

Etkileşim Ortamı	N	\tilde{x}	1. çeyrek	3. çeyrek	IQR	Maksimum Değer	Aykırı Gözlem Sayısı
Hiper Metin	61	2330	339	7965.5	7626.5	19405.25	4
Video	64	1881	884	4076	3192	8864	7
Değerlendirme	62	500	376.5	666	289.5	1100.25	4
SCORM	56	1640	170	3995	3825	9732.5	4
Forum	59	427	56	1025	2478.5	2038.4	7

* IQR= Çeyrekler Arası (3. çeyrek – 1. çeyrek)

* Maksimum Değer = (3rd çeyrek + IQR)

Hiper-metin değişkeni için ortanca değeri (\tilde{x})=2330, maksimum değer ise 19405.25 olarak hesaplanmıştır. Sağ kuyrukta bulunan 4 gözlemin maksimum değer üzerinde yer aldığı belirlenmiştir. Video değişkeni için ortanca değeri (\tilde{x})=1881, maksimum değer 8864 olarak hesaplanmıştır. Buna göre maksimum değer üzerinde yer alan 7 gözlem aykırı gözlem olarak belirlenmiştir. Değerlendirme değişkeni için ortanca değeri (\tilde{x})=500, maksimum değer ise 1100.25 olarak hesaplanmıştır. Değerlendirme değişkeni için 4 gözlemin maksimum değeri geçtiği belirlenmiştir. SCORM etkileşimi için ortanca değeri (\tilde{x})=1640, maksimum değeri ise 9732.5 olarak bulunmuştur. SCORM değişkeni için maksimum değer üzerinde olan 4 verinin aykırı gözlem olduğu tespit edilmiştir. Son olarak forum değişkeni için ortanca değeri (\tilde{x})=427, maksimum değer 2038.4 olarak hesaplanmış ve 7 değer aykırı gözlem olarak belirlenmiştir.

Rosner'in Yöntemi ile Aykırı Değerlerin Tespit Edilmesi

Son olarak veri kümesine Rosner yöntemi uygulanmıştır. Veri kümesi küçükten büyüğe doğru sıralanarak sağ kuyrukta yer alan gözlemler birer birer şüpheli gözlem olarak ele alınmıştır. Buna göre her hesaplamanın sonucunda bir değer elde edilmiş ve bu değer Rosner'in kritik değer tablosu ($\alpha=0.05$) (Rosner, 1983) ile karşılaştırılarak ilgili gözlemin aykırı gözlem olup olmadığına karar verilmiştir. Grubbs yöntemi gibi Rosner yöntemi de tekrarlı bir yöntem olduğundan gözlem sayısı her aşamada birer azaltılmış ve nihai veri setine ulaşılmıştır. Rosner yöntemi sonuçları Tablo 7'de verilmiştir.

Tablo 7. Rosner yöntemi sonucunda belirlenen aykırı gözlem sonuçları

Etkileşim Ortamı	N	\bar{X}	S	Karar Değeri	Aykırı Gözlem Sayısı
Hiper Metin	61	5454.44	7835.08	$R \geq 3.21$	4
	60	4925.7	6714.63	$R \geq 3.20$	
	59	4499.29	5896.26	$R \geq 3.19$	
	58	4072.95	4946.06	$R \geq 3.19$	
	57	3765.21	4394.26	$R < 3.18$	
	56	3526.41	4043.64	$R < 3.17$	
Video	64	3437.73	3757.24	$R \geq 3.22$	1
	63	3244.68	3452.66	$R < 3.22$	
	62	3074.84	3204.55	$R < 3.21$	
Değerlendirme	62	570.27	368.14	$R \geq 3.21$	3
	61	545.03	312.45	$R \geq 3.21$	
	60	526.62	279.72	$R \geq 3.20$	
	59	508.68	244.85	$R < 3.19$	
	58	495.45	224.71	$R < 3.19$	
SCORM	56	2703.20	3604.24	$R \geq 3.17$	4
	55	2410.60	2889.29	$R < 3.16$	
	54	2252.20	2664.50	$R \geq 3.15$	
	53	2091.25	2410.43	$R \geq 3.15$	
	52	1930.10	2126.21	$R < 3.14$	
Forum	59	933.07	1353.22	$R \geq 3.19$	6
	58	833.07	1123.83	$R \geq 3.19$	
	57	770.49	1026.83	$R < 3.18$	
	56	715.14	946.44	$R \geq 3.17$	
	55	660.13	860.06	$R \geq 3.16$	
	54	605.96	767.63	$R \geq 3.15$	
	53	551.79	662.64	$R < 3.15$	
	52	512.77	604.49	$R < 3.14$	

* R = Rosner yöntemi sonuçları

Tablo 7 incelendiğinde Hiper-metin değişkeni için 6 kez Rosner yöntemi uygulandığı görülmektedir. Hesaplama sonucunda elde edilen sonuçlar Rosner kritik değer tablosu ile karşılaştırılmış ve 4 gözlemin aykırı gözlem olduğuna karar verilmiştir ($R > R_k$). Video değişkeninde 64 gözlem bulunmaktadır. Rosner yönteminin bu gözlemlere 3 kez uygulanmasının sonucunda sadece en büyük değere sahip gözlem için hesaplanan değer Rosner'ın kritik değer tablosundaki değerden büyük olduğu görülmüştür ($R > R_k$). Değerlendirme için beş kez Rosner yöntemi uygulanmış ve 3 gözlemin aykırı gözlem olduğu belirlenmiştir ($R > R_k$). SCORM değişkeni için beş kez Rosner yöntemi uygulanmıştır. Yapılan hesaplamalar 56. 54. ve 53. gözlem için Rosner'ın kritik değer tablosundaki değerlerden büyük sonuçlara ulaşıldığı belirlenmiştir. Ancak 55. gözlem için hesaplanan değer Rosner'ın kritik değer tablosundaki değerden küçük olduğu görülmüştür. Ancak gözlemden sonra gelen diğer

gözlemler aykırı gözlem olarak belirlendiğinden 55. gözlem de aykırı olarak kabul edilmiştir. Forum değişkeni için veri için 8 kez Rosner yöntemi uygulanmış ve 6 gözlemin aykırı olduğu sonucuna ulaşılmıştır. Benzer şekilde forum değişkeni için 57. gözleme ilişkin hesaplanan değer her ne kadar kritik değerden küçük olsa da kendinden sonraki gözlemler aykırı gözlem olarak belirlendiği için bu gözlem de aykırı gözlem olarak kabul edilmiştir.

Bu çalışmada e-öğrenme ortamlarında gerçekleşen beş farklı etkileşim türü birer değişken olarak ele alınmış ve bu değişkenlere 5 farklı istatistiksel aykırı gözlem bulma yöntemi uygulanmıştır. Analizler sonucunda belirlenen aykırı gözlem sayılarına ilişkin sonuçlar Tablo 8’de toplu olarak sunulmuştur.

Tablo 8. Öğrencilerin ÖYS’ndeki etkileşim sürelerine ilişkin farklı yöntemler sonucunda elde edilen aykırı gözlem sayıları

Etkileşim Ortamı	Z	Grubbs	Hampel	Box-Plot	Rosner
Hiper-Metin	4	5	6	4	4
Video	4	3	4	7	1
Değerlendirme	3	4	4	4	3
SCORM	4	9	5	5	5
Forum	4	7	7	7	6

Farklı e-öğrenme etkileşim verilerine aykırı gözlem belirleme yöntemlerinin uygulanması sonucunda bulunan aykırı gözlemlerin sayıları Tablo 8’de özetlenmiştir. Farklı yöntemler sonucunda birbirinden farklı kesme noktaları oluşmuştur. Bu kesme noktalarının üzerindeki tüm veriler aykırı gözlem olarak işaretlenmiştir. Tablo 8 incelendiğinde farklı etkileşim değişkenlerine uygulanan beş farklı yöntem sonucunda değişen sayılarda aykırı gözlemlerin tespit edildiği görülmektedir. Farklı yöntemler sonucunda her ne kadar farklı sayıda aykırı gözlem belirlense de sonuçlar veri setinde yer alan en büyük değerlere sahip gözlemleri işaret etmektedir. Örneğin Örneğin Z, Box-plot ve Rosner yönteminin hiper-metin verilerine uygulanması sonucunda 4’er aykırı gözlem bulmuştur. Bunların dördü de sağ kuyrukta yer alan aynı gözlemlerdir. E-değerlendirme değişkeni için farklı yöntemlerin yaklaşık olarak benzer sonuçlar ürettiği görülmektedir. Nitekim bu değişkene ilişkin Tablo 1’de verilen betimsel istatistikler göz önünde bulundurduğunda bu değişkenin diğerlerine kıyasla normal dağılıma daha yakın olduğu gözlenmiştir. Özellikle video ve SCORM etkileşimlerinde, ele alınan yöntemlerin daha tutarsız sonuçlar ürettiği söylenebilir.

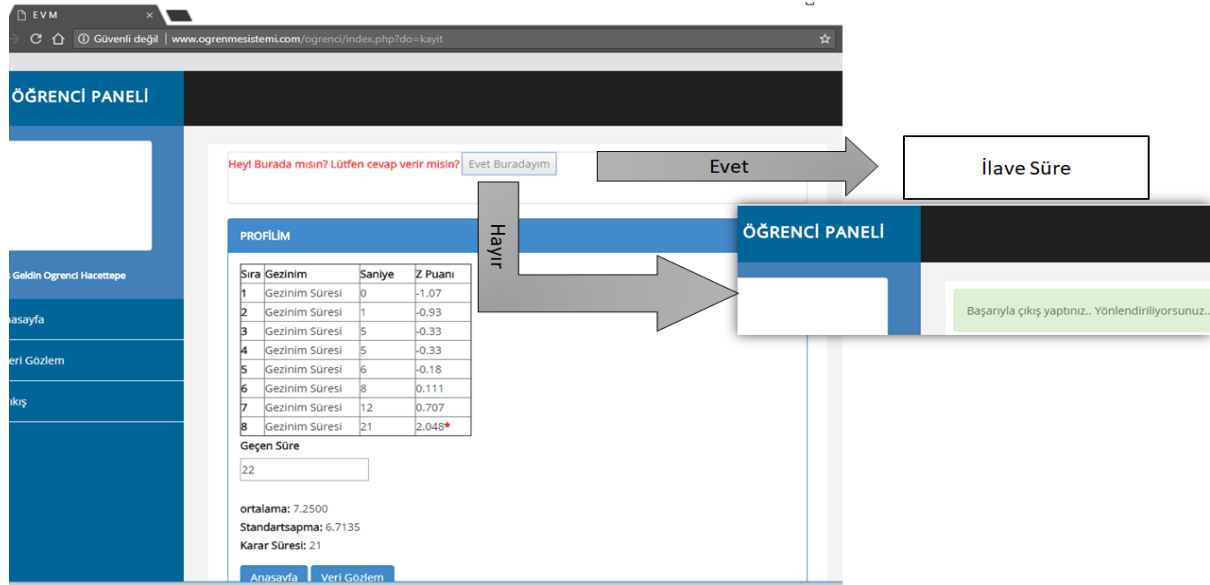
Sonuç ve Öneriler

Bu çalışmada beş farklı e-öğrenme etkileşimine ilişkin sistemde geçirilen zaman değişkenleri ele alınmıştır. Bu değişkenler içerisinde dağılımı bozan şüpheli gözlemlere beş farklı istatistiksel aykırı gözlem bulma yöntemi uygulanmıştır. Gerçekleştirilen uygulamalar sonucunda farklı aykırı gözlem bulma yöntemleri kullanılarak, e-öğrenme etkileşim verileri içinde başarılı bir şekilde aykırı gözlemlerin belirlenebileceği görülmüştür. Buna göre istatistik alanında yaygın bir şekilde kullanılan aykırı gözlem bulma yöntemlerinin e-öğrenme verileri için de uygulanabilir olduğu söylenebilir.

Bu araştırmada ele alınan aykırı gözlem bulma yöntemlerinin farklı sayılarda aykırı gözlemler bulduğu görülmektedir. Ancak bu tabloya (Tablo 8) bakarak hangi yöntemin aykırı gözlemleri bulmada daha başarılı olduğu söylemek mümkün değildir. Bu çalışma ele alınan

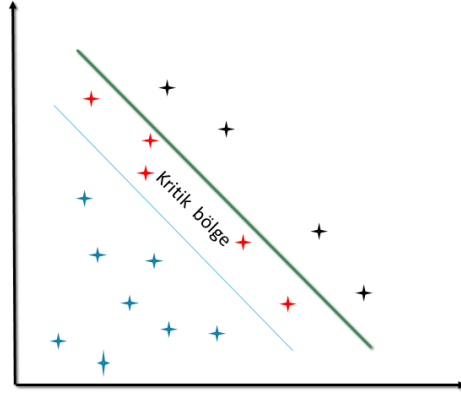
yöntemlerin e-öğrenme verilerine uygulanabilirliğinin sınırlanması ile sınırlı olduğundan bu yöntemlerin aykırı gözlemleri bulmadaki performanslarının kıyaslandığı çalışmalara da ihtiyaç duyulduğu söylenebilir.

Araştırma kapsamında ayrıca ele alınan aykırı gözlem bulma yöntemlerinin e-öğrenme verilerine uygulanmasının yanı sıra bu yöntemlerin çalışan bir e-öğrenme ortamına entegre edilebilirliğini değerlendirebilmek adına örnek PHP betik (Hypertext Preprocessor) kodlamaları yapılmıştır. Hazırlanan örnek sistem tasarımına ilişkin ekran görüntüsü Şekil 2’de verilmiştir.



Şekil 2. Aykırı gözlemleri gerçek zamanlı bulmaya yönelik örnek sistem tasarımı

Şekil 2’de sunulan örnek sistem tasarımı, kullanıcı-sistem etkileşimlerini anlık olarak izleyerek gerçek zamanlı olarak aykırı verilerin belirlenmesi amacıyla hazırlanmıştır. Sistemde makalenin giriş bölümünde detaylı olarak açıklanan farklı aykırı gözlem bulma yöntemleri kullanılmıştır. Bu yöntemler, gerçek zamanlı olarak aykırı gözlemlerin belirlenmesinde iki şekilde kullanılabilir. İlk olarak bir eğitim dönemi süresince kullanılmış olan bir e-öğrenme sisteminden elde edilen etkileşim verileri, eğitim verisi olarak ele alınır. Bu sisteme ilişkin her etkileşim değişkeni için bir kritik değer (değişkenin alabileceği maksimum değer) belirlenir. Buradan elde edilen bilgiler bir sonraki dönemde yeni kullanıcı etkileşimlerini kontrol etmek amacıyla kullanılır. Yeni dönemde gerçekleşen etkileşimlere ilişkin gözlem değeri, önceden belirlenen kritik değeri aştığı durumlarda sistem kullanıcıyı uyarma veya veri tabanında bu veriyi düzeltme şeklinde bir müdahalede bulunabilir. İkinci yöntem ise bir eğitim verisi olmadan sistem kullanıldıkça oluşan etkileşim verileri üzerinden aykırı değerlerin bulunmasını esas almaktadır. Bir önceki yöntemden farklı olarak bu tür bir uygulamada veri tabanına veri eklendikçe aykırı gözlemler için belirlenen kritik değer dinamik olarak değişebilmektedir. Bir diğer ifade ile makine öğrenmesine dayalı olarak gerçek zamanlı aykırı gözlem belirleme-önleme gerçekleştirilebilir. Şekil 3’te veri dağılımında aykırı gözlemlere karar vermeye ilişkin örnek bir gösterim verilmiştir.



Şekil 3. Veri dağılımında aykırı gözlemlere ilişkin kritik bölge

Şekil 3'te verilen kullanıcı etkileşimlerine ilişkin veri dağılımında yer alan mavi işaretli gözlemler, dağılımı bozmayan normal davranışları temsil etmektedir. Gözlemlere ilişkin değerler anormal bir şekilde büyüme göstermesi halinde ise bu değerler kritik bölge olarak işaretlenen alana yaklaşacaktır. Anormal olarak ifade edilen e-öğrenme davranışı ise siyah olarak dağılım grafiğinde gösterilmiştir. Buna göre gerçek zamanlı e-öğrenme sistemlerinin, kırmızı olarak işaretlenmiş alana düşen gözlemlere müdahale etmesi gerekmektedir. Bu müdahale yukarıda da ifade edildiği gibi veri düzeltme veya kullanıcıyı uyarma şeklinde gerçekleştirilebilir.

Buradan elde edilen deneyimler ve veri tabanı yapısı göz önünde bulundurulduğunda Z yöntemi ve kutu grafiği yöntemleri bir e-öğrenme sisteminde uygulama anında aykırı gözlemlerin tespiti amacıyla diğer yöntemlere göre daha kolay uygulanabilir olduğu görülmüştür. Bir başka ifadeyle bu yöntemlerin makineye öğretiminin daha işlevsel olduğu söylenebilir. Bununla birlikte diğer yöntemlerin ise bir hipotez sınaması gerektirmesi ve daha duyarlı sonuçlar vermesi yönünden önemli bir avantaja sahip olduğu göz önünde bulundurulmalıdır. Ele alınan bu istatistiksel yöntemlere ek olarak makina öğrenmesine dayalı aykırı gözlem belirlemede; KNN, kümeleme, SVM vb. gibi yöntemlerden sıkça faydalandığı görülmektedir (Choudhary, 2017; Hogo, 2010). Bu yöntemler özellikle veri dağılımının sürekli değiştiği ve veri setinde gürültü verilerinin ve aykırı gözlemlerin bir arada bulunduğu durumlarda tercih edilmektedir. Bu yöntemler verilerin yoğunluk noktalarını, gruplandırmaları ve veri setindeki kesme noktalarını kullanarak aykırı gözlemleri belirlemektedir.

Kaynakça

ArcGIS Pro (2018). *Box Plot*. Erişim Tarihi: 24.04.2018, <https://pro.arcgis.com/en/pro-app/help/analysis/geoprocessing/charts/box-plot.htm>.

Cantador, I., & Conde, J. M. (2010). *Effects of competition in education: A case study in an e-learning environment*. Proceedings of the IADIS International Conference E-learning 2010, Retrieved from

<https://pdfs.semanticscholar.org/95a0/4babb8841f3f644e2d7d497c98807eac3595.pdf>

Chouldary, P. (2017) Introduction to Anomaly Detection.

<https://www.datascience.com/blog/python-anomaly-detection> Adresinden 12.10.2018 tarihinde alınmıştır.

Durivage, M. A. (2014). Practical engineering, process, and reliability statistics. ASQ Quality Press.

Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304-317.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics*, 11(1), 1–21. <https://doi.org/10.2307/1266761>

Grubbs, F. E., & Beck, G. (1972). Extension of sample sizes and percentage points for significance tests of outlying observations. *Technometrics*, 14(4), 847-854.

Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, 42, 1887-1896.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346), 383-393.

Han, J., Kanber, M. (2006) *Data Mining: Concepts and Techniques*, Morgan Kaufmann.

Hogo, M. A. (2010). Evaluation of e-learners behaviour using different fuzzy clustering models: a comparative study. *arXiv preprint arXiv:1003.1499*.

LAK. (2011) Learning Analytics & Knowledge. Retrieved from: <https://tekri.athabasca.ca/analytics/>

Lal, P. (2014). Designing online learning strategies through analytics. In *Online Tutor 2.0: Methodologies and Case Studies for Successful Learning* (pp. 1-15). IGI Global.

McGill, R., Tukey, J. W., & Larsen, W. A. (1978). Variations of box plots. *The American Statistician*, 32(1), 12-16.

Moore, D. S. and McCabe, G. P. (1999) *Introduction to the Practice of Statistics*, 3rd ed. New York: W. H. Freeman, 1999.

Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same?. *The Internet and Higher Education*, 14(2), 129-135.

Orosz, G., Farkas, D., & Roland-Levy, C. (2013). Are competition and extrinsic motivation reliable predictors of academic cheating? *Frontiers in Psychology*, 4(87), 1e16. <http://dx.doi.org/10.1080/10508422.2013.877393>.

Rosner, B. (1983). Percentage points for a generalized ESD many-outlier procedure. *Technometrics*, 25(2), 165-172.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400.

Şahin, M. & Yurdugül, H. (2018). Öğrenme Yönetim Sistemi 3.0 ve Öğrenen Beklentileri. EDUCOON 2018, September, Ankara, Turkey.

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, M.A.