

Eđitim Teknolojisi

kuram ve uygulama

Yaz 2016

Cilt 6

Sayı 2

Summer 2016

Volume 6

Issue 2

Educational Technology

theory and practice

ISSN: 2147-1908

Cilt 6, Sayı 2, Yaz 2016
Volume 6, Number 2, Summer 2016

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

Sorumlu Yazı İşleri Müdürü / Publisher Editor: **Dr. Sami ŞAHİN**
Redaksiyon / Redaction: **Dr. Tolga GÜYER**
Dizgi / Typographic: **Dr. Tolga GÜYER**
Sayfa Tasarımı / Page Design: **Dr. Tolga GÜYER**
Kapak Tasarımı / Cover Design: **Dr. Bilal ATASOY**
İletişim / Contact Person: **Dr. Aslıhan KOCAMAN KAROĞLU**

Taranmaktadır / Indexed in: **ULAKBİM Sosyal ve Beşeri Bilimler Veritabanı, Türk Eğitim İndeksi, ASOS Sosyal Bilimler İndeksi**

Editör Kurulu / Editorial Board*

Dr. Abdullah Kuzu
Dr. Akif Ergin
Dr. Ana Paula Correia
Dr. Aytekin İşman
Dr. Buket Akkoyunlu
Dr. Cem Çuhadar
Dr. Deniz Deryakulu
Dr. Deepak Subramony

Dr. Eralp H. Altun
Dr. Feza Orhan
Dr. H. Ferhan Odabaşı
Dr. Hafize Keser
Dr. Halil İbrahim Yalın
Dr. Hyo-Jeong So
Dr. İbrahim Gökdaş
Dr. Kyong Jee(Kj) Kim

Dr. M. Oğuz Kutlu
Dr. M. Yaşar Özden
Dr. Mehmet Gürol
Dr. Michael Evans
Dr. Michael Thomas
Dr. Özcan Erkan Akgün
Dr. Özgen Korkmaz
Dr. S. Sadi Seferoğlu

Dr. Sandie Waters
Dr. Scott Warren
Dr. Servet Bayram
Dr. Şirin Karadeniz
Dr. Tolga Güyer
Dr. Trena Paulus
Dr. Yasemin Gülbahar Güven
Dr. Yavuz Akpınar
Dr. Yun-Jo An

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

Hakem Kurulu / Reviewers*

Dr. Adile Aşkın Kurt
Dr. Agah Tuğrul Korucu
Dr. Arif Altun
Dr. Aslıhan Kocaman Karoğlu
Dr. Ayça Çebi
Dr. Ayfer Alper
Dr. Aynur Kolburan Geçer
Dr. Ayşegül Bakar Çörez
Dr. Aytekin İşman
Dr. Bahar Baran
Dr. Berrin Doğusoy
Dr. Bilal Atasoy
Dr. Deniz Atal Köysüren
Dr. Ebru Kılıç Çakmak
Dr. Ebru Solmaz
Dr. Emin İbili
Dr. Emine Şendurur
Dr. Erinç Karataş
Dr. Erhan Güneş
Dr. Erkan Çalışkan
Dr. Erkan Tekinarslan
Dr. Ertuğrul Usta
Dr. Fatma Keskinlik

Dr. Fezile Özdamlı
Dr. Fikret Gelibolu
Dr. Filiz Kalelioğlu
Dr. Gizem Karaoğlan
Dr. Gökçe Becit İşçitürk
Dr. Gökhan Dağhan
Dr. Gülfidan Can
Dr. Halil Ersoy
Dr. Halil İbrahim Yalın
Dr. Halil Yurdugül
Dr. Hasan Çakır
Dr. Hasan Karal
Dr. Hatice Durak
Dr. Hüseyin Bicen
Dr. Hüseyin Özçınar
Dr. Işıl Kabakçı Yurdakul
Dr. İbrahim Gökdaş
Dr. İlknur Resioğlu
Dr. Kevser Hava
Dr. Mehmet Akif Ocak
Dr. Mehmet Barış Horzum
Dr. Mehmet Kokoç
Dr. Melih Engin

Dr. Meltem Kurtoğlu
Dr. Muharrem Aktümen
Dr. Mukaddes Erdem
Dr. Mustafa Serkan Günbatır
Dr. Mutlu Tahsin Üstündağ
Dr. Nadire Çavuş
Dr. Necmettin Teker
Dr. Necmi Eşgi
Dr. Nezhil Önal
Dr. Nuray Gedik
Dr. Nurettin Şimşek
Dr. Onur Dönmez
Dr. Ömer Faruk Ursavaş
Dr. Ömür Akdemir
Dr. Özcan Erkan Akgün
Dr. Özden Şahin İzmirli
Dr. Özgen Korkmaz
Dr. Özlem Çakır
Dr. Ramazan Yılmaz
Dr. Recep Çakır
Dr. Sami Şahin
Dr. Selay Arkün Kocadere
Dr. Selçuk Özdemir

Dr. Serap Yetik
Dr. Serdar Çiftçi
Dr. Serçin Karataş
Dr. Serpil Yalçınalp
Dr. Sibel Somyürek
Dr. Şafak Bayır
Dr. Şeyhmus Aydoğdu
Dr. Şirin Karadeniz
Dr. Tayfun Tanyeri
Dr. Tolga Güyer
Dr. Tolga Kabaca
Dr. Türkan Karakuş
Dr. Uğur Başarmak
Dr. Ümmühan Avcı Yücel
Dr. Ünal Çakıroğlu
Dr. Veysel Demirel
Dr. Yalın Kılıç Türel
Dr. Yasemin Deminarslan Çevik
Dr. Yasemin Gülbahar Güven
Dr. Yasemin Koçak Usluel
Dr. Yavuz Akbulut
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.ulakbim.gov.tr/etku/>

E-Posta / E-Mail: tguyer@gmail.com

Telefon / Phone: +90 (312) 202 17 38

Belgegeçer / Fax: +90 (312) 202 83 87

Adres / Address: Gazi Üniversitesi, Gazi Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü,
06500 Teknikokullar - Ankara / Türkiye

Makale Geçmişi / Article History

Alındı/Received: 25.02.2016

Düzeltilme Alındı/Received in revised form: 03.03.2016

Kabul edildi/Accepted: 06.03.2016

ÇEVİRİMİÇİ ÖĞRENME ORTAMINDAKİ FARKLI ÖĞRENCİ PROFİLLERİNİN KÜMELEME YÖNTEMİ İLE BELİRLENMESİ*

Gökhan AKÇAPINAR¹, Arif ALTUN², Petek AŞKAR³

Öz

Bu çalışmanın amacı, çevrimiçi öğrenme ortamında benzer davranış örüntüsü sergileyen farklı öğrenci gruplarının kümeleme yöntemi ile belirlenmesidir. Çalışma, Türkiye’de bir devlet üniversitesinde Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü’nde Bilgisayar Donanımı dersine kayıtlı 76 üniversite ikinci sınıf öğrencisi ile yürütülmüştür. Öğrenciler 14 hafta süresince yüz yüze derslere ek olarak çevrimiçi öğrenme ortamında, yansıma yazma, tartışma, soru-cevap, kaynak takibi vb. aktiviteler gerçekleştirmişlerdir. Analizlerde kullanılan veriler bu ortamın veri tabanından elde edilmiştir. Veriler iki farklı kümeleme algoritması ile analiz edilmiş ve sonuçları karşılaştırılmıştır. Aynı zamanda elde edilen farklı öğrenci gruplarının akademik performansları incelenerek etkileşim düzeyi ile akademik performans arasındaki ilişki analiz edilmiştir. Araştırma sonuçları öğrencilerin çevrimiçi öğrenme ortamındaki davranış örüntülerine göre üç farklı kümeye ayrılabilceğini göstermiştir. Bu kümelerin isimlendirilmesi konusunda ise iki farklı yaklaşım izlenmiştir. Kümeler, her bir kümede yer alan öğrencilerin aktivite düzeylerine göre (Çok aktif, Aktif, Aktif olmayan) ve öğrencilerin akademik performanslarına göre (Yüksek öğrenenler, Orta öğrenenler, Düşük öğrenenler) tanımlanmıştır.

Anahtar Kelimeler: Öğrenci profili belirleme; çevrimiçi öğrenme; eğitsel veri madenciliği; kümeleme.

IDENTIFYING DIFFERENT STUDENT PROFILES IN AN ONLINE LEARNING ENVIRONMENT WITH CLUSTER ANALYSIS

Abstract

The purpose of this study is to explore how to group students who exhibit similar behavior patterns in an online learning environment via clustering. Relationships between these clusters and students’ academic performance in each cluster were examined, as well. The

* Bu çalışma birinci yazarın doktora tezinden üretilmiştir.

¹ Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: gokhana@hacettepe.edu.tr

² Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: altunar@hacettepe.edu.tr

³ Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, emekli öğretim üyesi, e-posta: paskar@hacettepe.edu.tr

study group consisted of 76 university students studying at the Computer Education and Instructional Technology Department at a state university in Turkey. In addition to the face to face classes during the course of 14 weeks, the students conducted several activities (writing reflections, participating in discussions, following course resources, etc.) in the online learning environment. The results of the cluster analysis indicated that students were ideally divided into three different groups according to the activities carried out in the online learning environment. Regarding the labeling of these clusters, two different methods were used. Firstly, the finalized clusters were named according to the students' activity levels (Non-active, Active, Very active). Then, these clusters were named with regards to students' academic performance (Low learners, Moderate learners, High learners).

Keywords: Student profiling; online learning; educational data mining; clustering.

Summary

Online learning environments can store significant amount of interaction data in a database regarding students' usage. However, analyzing these data to improve teaching and learning is restricted to simple statistics and graphics (Ali, Asadi, Gašević, Jovanović, & Hatala, 2013). On the other hand, data mining (DM) methods which has been used in various fields for many years to find hidden patterns and trends in the data, carry significant potentials regarding the analysis of educational data. Investigation of grouping similar students in online learning environments is one of these application areas. This can be used for grouping students who has similar attributes such as personal characteristics, learning preferences, individual differences, usage patterns, etc. (Cristobal Romero & Ventura, 2013).

In this study, it was intended to profile students who exhibit similar behavior in an online learning environment via two different clustering algorithms. Relationships between these clusters and students' academic performance in each cluster were examined, as well. The following research questions were addressed:

1. How many groups can be identified according to the X-Means clustering algorithm and how they can be defined based on the cluster centers?
2. How many groups can be identified according to the EM clustering algorithm and how they can be defined based on the cluster centers?
3. Can these clusters obtained by two different algorithms be defined based on students' academic success?
4. How students who failed and passed the course were distributed across these clusters?

The study group consisted of 76 university students studying at Computer Education and Instructional Technology Department at a state university in Turkey. In addition to the face to face classes during the course of 14 weeks, the students completed several tasks in the online learning environment. These tasks included writing reflections about the concepts which they learned in the course, reading, commenting and assessing (like/dislike) reflections written by other students, asking questions in discussions, writing answers to questions, assessing questions and answers which were written by others, following announcements and course

resources. Using the data collected from the environment, 28 variables were targeted representing students' behavior in the environment.

As a preprocessing step, feature selection algorithm was applied to data and then the best ranked 10 attributes associated with academic performance (t_QuestionAssessment, t_UniqueSession, t_UniquesPost, t_Post, t_Navigation, t_Session, t_TagUsage, t_Tag, t_QuestionAnswer, t_SessionDuration) was selected for cluster analysis. As a next step, all attributes were standardized by converting them into z scores. Data was analyzed with two different clustering algorithms (X-Means Clustering, EM Clustering) and the results were compared in terms of clustering performance. In both analyses, optimum cluster numbers were determined by algorithms. Clustering analysis were conducted by RapidMiner data mining software with the help of W-Xmeans and W-EM packages.

At the end of the analysis, it was seen that both clustering algorithms exhibited similar results. When examining the results of the cluster analysis, it was observed that students ideally were divided into three different groups according to the tasks they carried out in the online learning environment. When cluster means are examined, it is seen that students in Cluster 3 made more logins into the environment, spent more time, wrote more posts, wrote posts on more different days, used more tags in their posts, visit more posts written by other students, visited the discussion section more often, wrote more answers to the questions, made more comments about the posts and assessing more questions and posts compared to students in other clusters. Students in Cluster 1 however, are seen to have made much fewer logins into the environment, wrote much fewer posts, used fewer tags in their posts, visited the discussion section much less, wrote almost no answers to the questions, made fewer comments about the posts written by others and assessing less questions and posts. Finally, Cluster 2 is characterized by students with values somewhat smaller than Cluster 3 but greater than Cluster 1. The resulting clusters were named according to the students' activity levels: Non-active (Cluster 1), Active (Cluster 2) and Very Active (Cluster 3).

When relationships between these clusters and students' academic performance in each cluster were examined (see Table 6), it was found that students with lower activities in the environment (Cluster 1) had a lower course success (Low learners); students with moderate activities in the environment (Cluster 2) had a moderate course success (Moderate learners); and students with higher activities in the environment (Cluster 3) had a higher course success (High learners).

The results might be considered as an expected conclusion that there was a positive correlation between students' success and their activity levels in online learning environment. However, modeling this relationship based on data is important. It should be noted that this study displays how data mining could be used for educational purposes. Secondly, researchers, teachers and instructional designers can benefit from using such data-driven models. These models can be used to classify students automatically (Valdiviezo, Reátegui, & Sarango, 2013) or to make an automatic adaptation based on their activity level in an adaptive learning environment (Amershi & Conati, 2009). Teachers can use these models to monitor their students' learning progress in an online learning environment. At the same time, they can use these clusters to form workgroups in collaborative activities (Cristóbal Romero, Ventura & García, 2008).

Giriş

Çevrimiçi öğrenme ortamları öğretme ve öğrenmeyi destekleyecek çok sayıda araca sahiptir. Öğretmenler bu araçları kullanarak ders materyallerini paylaşabilirler, ödevler ve testler hazırlayabilirler, tartışmalar yürütebilirler (Cristobal Romero, Espejo, Zafra, Romero ve Ventura, 2013); öğrenciler ise bu ortamlarda viki, blog, mesajlaşma vb. araçları kullanarak işbirlikli çalışmalar yapabilirler (Moreno, Gonzalez, Castilla, Gonzalez ve Sigut, 2007) ya da kendi öğrenme etkinliklerini planlayabilirler. Geleneksel sınıf ortamlarından farklı olarak çevrimiçi öğrenme ortamları, öğrenciler tüm bu etkinlikleri gerçekleştirirken geride bıraktıkları izlerin kayıt edilmesine olanak sağlamaktadır. Bu izler; öğrencinin öğrenme materyali ile etkileşimi, tartışma ortamında bir soruya cevap yazması, ya da sadece ortama giriş – çıkış yapması ile ilgili olabilir. Son yıllarda çevrimiçi öğrenci sayısındaki artış da dikkate alındığında öğrenme süreçleri ile ilgili önemli miktarda verinin çevrimiçi veri tabanlarında kayıt edildiğinden söz edilebilir; fakat veri tabanlarında tutulan bu verilerin eğitim - öğretimin iyileştirilmesi amacıyla kullanımı tanımlayıcı istatistikler ve basit grafiklerle sınırlı kalmaktadır (Ali ve diğerleri, 2013).

Diğer taraftan uzun yıllardır çeşitli alanlarda veri tabanlarındaki gizli örüntü ve eğilimleri belirlemek amacıyla kullanılan veri madenciliği (VM) yöntemleri, eğitsel ortamlardan elde edilen bu tür verilerin analizi konusunda da önemli bir potansiyele sahiptir. Eğitsel ortamlarda VM yöntemleri; öğrenme kuramlarının değerlendirilmesinde, öğrenci destek ve geri bildirimlerinde, performans kestirimlerinde, erken uyarı sistemlerinde, öğrenme teknolojilerinde ve gelecekteki öğrenme uygulamalarının geliştirilmesinde kullanılabilir (Greller ve Drachler, 2012). Eğitsel Veri Madenciliği (EVM) araştırmacılarının üzerinde çalıştığı önemli bir konu ise verideki gizli örüntülerin ortaya çıkartılmasıdır. Benzer davranış örüntüsü sergileyen öğrenci gruplarının belirlenmesi bunun bir örneğidir ve burada amaç; kişisel özellikleri, öğrenme tercihleri, bireysel farklılıkları vb. açılardan benzer öğrenci gruplarının belirlenmesidir (Cristobal Romero ve Ventura, 2013). Bu sayede öğretim elemanları/öğretmenler bu bilgileri kullanarak öğrencilerin gelişimini izleme ve sorun yaşayan öğrencilere uygun müdahale yöntemleri geliştirme konusunda değerli bilgiler üretebilirler. Aynı zamanda bu veriler, uyarlanabilir öğrenme ortamlarında öğrencilerin otomatik olarak sınıflandırması amacıyla ve benzer öğrenci grupları için otomatik uyarlamalar yapılması amacıyla kullanılabilir.

Bu çalışmanın amacı, çevrimiçi öğrenme ortamında benzer davranış örüntüsü sergileyen farklı öğrenci gruplarının kümeleme yöntemi ile belirlenmesidir. Elde edilen farklı grupların akademik performansları da ayrıca analiz edilmiştir. Bu bağlamda cevap aranan araştırma soruları aşağıdaki gibidir.

1. X-Ortalamalar kümeleme algoritmasına göre öğrenciler kaç farklı gruba ayrılmaktadır ve küme merkezlerine göre bu gruplar nasıl tanımlanabilir?
2. EM kümeleme algoritmasına göre öğrenciler kaç farklı gruba ayrılmaktadır ve küme merkezlerine göre bu gruplar nasıl tanımlanabilir?
3. Her iki kümeleme algoritmasına göre elde edilen farklı kümeler bu kümelerde yer alan öğrencilerin akademik performanslarına göre tanımlanabilir mi?

4. Her iki kümeleme algoritmasına göre dersi geçen ve kalan öğrencilerin kümelere göre dağılımları nasıldır?

Çalışmanın devamı şu şekilde organize edilmiştir: İlk olarak VM yöntemleri kullanılarak benzer öğrenci gruplarının belirlenmesi ile ilgili çalışmalara yer verilmiştir. Daha sonra araştırma yönteminin de temelini oluşturan EVM sürecine yer verilmiştir. Yöntem bölümünde, araştırmada kullanılan değişkenler, veri toplama süreci, veri kaynakları ve veri analiz süreci anlatılmıştır. Son bölümde ise elde edilen bulgular, ilgili literatür eşliğinde tartışılmış ve sonuçlar yorumlanmıştır.

Benzer Öğrenci Gruplarının Belirlenmesi ile İlgili Çalışmalar

Amershi ve Conati (2009) çevrimdışı ve çevrimiçi olmak üzere iki bölümden oluşan araştırmalarında yarı otomatik bir uyarlanabilir öğrenme ortamı tasarlamışlardır. Araştırmacılar ilk olarak çevrimdışı bölümde öğrencilerin öğrenme ortamındaki etkileşim davranışlarını temsil eden değişkenleri kümeleme yöntemiyle analiz ederek benzer özellikteki öğrencileri gruplamışlardır. Analiz sonucuna göre de öğrencilerin etkileşim verilerine göre iki kümeye ayrıldığını tespit etmişlerdir. Bu kümeleri ise Yüksek Öğrenme (YÖ) ve Düşük Öğrenme (DÖ) şeklinde adlandırmışlardır. Araştırmanın çevrimiçi bölümünde ise araştırmacılar, kümeleme analizi sonucu elde edilen bu küme bilgilerini kullanarak bir tahmin modeli oluşturmuşlardır ve oluşturulan bu tahmin modelini ortama yeni giren öğrencilerin öğrenme çıktılarını (yüksek, düşük) tahmin etmek amacıyla kullanmışlardır. Araştırmacılar geliştirdikleri modeli iki farklı öğrenme uygulamasında yeni öğrencilerle test etmiş ve geliştirilen modelin öğrenci davranışlarını sırasıyla %88,3 ve %86,3 oranlarında doğru olarak sınıfladığını belirtmişlerdir. Araştırmacılar elde ettikleri bulguların öğretmenlerin çevrimiçi ortamların pedagojik faydaları konusundaki farkındalıklarının artırılmasında kullanılabileceğini ve yazılım geliştiricilere bu ortamların iyileştirilmesi konusunda yardımcı olabileceğini belirtmişlerdir.

Cristóbal Romero ve diğerleri (2008) öğrencilerin Moodle kullanım verilerine göre profillerini belirlemek amacıyla kümeleme analizinden yararlanmışlardır. Öğrenci verilerini k-Ortalamlar (k-Means) kümeleme yöntemi ile analiz etmişler ve küme sayısını ön tanımlı üç olarak belirtmişlerdir. Analiz sonucu elde edilen kümeleri nitelendirmek içinse küme merkezlerini (ortalama / mod) ve standart sapmalarını dikkate almışlardır. Buna göre; Küme 1'de yer alan öğrencilerin hiçbir ödevi yapmadığını, mesaj okuma ve quizlere katılım oranlarının çok düşük olduğunu ve ortamda geçirdikleri sürenin de düşük olduğunu belirtmişlerdir. Küme 2'de yer alan öğrencilerin ise daha fazla mesaj gönderdiklerini, daha fazla ödev yaptıklarını, daha fazla mesaj okuduklarını, quizlerde daha başarılı olduklarını ve ortamda çok daha fazla zaman harcadıklarını belirtmişlerdir. Küme 3'de yer alan öğrencilerin değerlerinin ise Küme 2'de yer alanlardan daha az Küme 1'de yer alan öğrencilerden ise daha fazla olduğu belirtilmiştir. Araştırmacılar elde edilen bu bilgileri kullanarak öğretmenlerin öğrencileri sınıflayabileceğini (çok aktif, aktif, aktif olmayan) ve işbirlikli çalışmalarda çalışma gruplarının oluşturulması konusunda bu bilgilerden yararlanabileceklerini belirtmişlerdir.

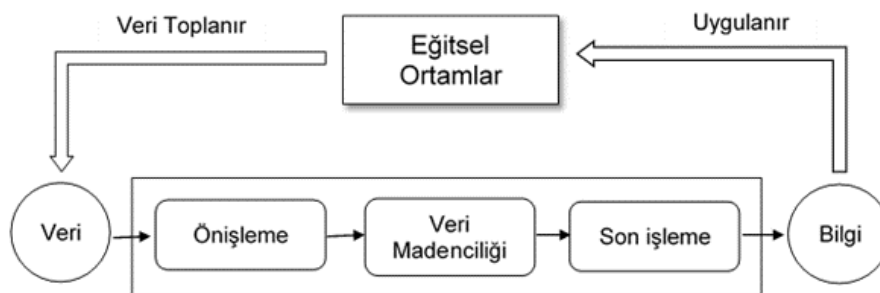
Beal, Qu ve Lee (2006), 83 lise öğrencisi ile yaptıkları çalışmada öğrencilerin matematik dersine karşı motivasyonları ile zeki öğrenme sistemindeki kullanım örüntüleri arasındaki ilişkiyi araştırmışlardır. Bu amaçla ilk olarak öğrencilere matematik motivasyon ölçeği uygulamışlar ve öğrencileri ölçeğin alt boyutlarından aldıkları puanlara göre hiyerarşik kümeleme analizi yardımı ile üç gruba ayırmışlardır. Aynı kümede yer alan öğrencilerin ölçeğin alt

boyutlarından aldıkları ortalama puanlarına göre ise grup profili çıkartmışlardır. Buna göre Küme 1 (n=50)'de yer alan öğrencileri matematiğin gelecek yaşamlarında önemli olabileceğini düşünen fakat diğer derslerden daha az seven, matematiğe karşı ortalama motivasyona sahip olarak; Küme 2 (n = 21)'de yer alan öğrencileri, matematiği sevmediğini belirten ve matematik konusunda çok da yeteneklerinin olmadığını düşünen öğrenciler olarak, Küme 3 (n=12)'de yer alan öğrencileri ise yetenekleri konusunda inançları yüksek olan, matematiği diğer derslerden daha çok sevdiklerini belirten ve matematik öğrenmenin önemli olduğunu düşünen öğrenciler olarak tanımlamışlardır. Araştırmacılar aynı zamanda farklı kümelerde yer alan öğrencilerin zeki öğrenme sistemindeki kullanım örüntülerini de analiz etmişlerdir. Buna göre Küme 1'de yer alan öğrencilerin daha çok, problemlerin cevabını tahmin ederek bulmaya çalıştıklarını, Küme 2'de yer alan öğrencilerin problemlerin çözümü ile ilgili öğrenme amaçlı olarak yardım materyallerinden yararlandıklarını Küme 3'de yer alan öğrencilerin ise yardım almadan problemleri doğru olarak çözdüklerini belirlemişlerdir. Bu bulgular öğrenci davranış verilerinin pedagojik modellerin oluşturulmasında ve öğrenme stratejilerinin tahmin edilmesindeki potansiyel kullanımına dikkat çekmektedir.

Yukarıdaki çalışmalardan da görüldüğü üzere çevrimiçi öğrenme ortamlarından elde edilen verilerin kümeleme yöntemleri ile analiz edilmesi, öğrenci, öğretmen, uyarlanabilir ortamlar ve ortam geliştiriciler için önemli çıktılar üretebilmektedir.

Eğitsel Veri Madenciliği

EVM; öğrenme ortamlarından elde edilen verilerin ortaya çıkartılması için yöntemler geliştirilmesi ve bu yöntemlerin öğrencilerin ve öğrenme ortamlarının daha iyi anlaşılmasında kullanılması ile ilgilenen bir disiplin olarak tanımlanmaktadır (Siemens ve Baker, 2012). García, Romero, Ventura ve de Castro (2011) Şekil 1'de görsel olarak verilen EVM sürecini eğitsel sistemlerden elde edilen ham verinin eğitim yazılımlarının, geliştiricilerin, öğretmenlerin ve araştırmacıların kullanabileceği bilgiye çevirme süreci olarak tanımlamaktadır.



Şekil 1. Eğitsel Veri Madenciliği Süreci (García, Romero, Ventura ve de Castro, 2011)

EVM çalışmalarının önemli bir bölümünde eğitsel ortam olarak, bilgisayar ve web tabanlı eğitim sistemleri kullanılmaktadır. Peña-Ayala (2014), 2010 – 2013 yılları arası yapılan EVM çalışmalarını incelediği çalışmada, incelediği makalelerin yarısına yakınında eğitsel ortam olarak Zeki Öğrenme Sistemlerinin (ZÖS) ve Öğrenme Yönetim Sistemlerinin (ÖYS) kullanıldığını belirtmiştir. Bunun dışında uyarlanabilir ortamlar, çevrimiçi değerlendirme sistemleri (Cristobal Romero ve Ventura, 2013), geleneksel sınıf ortamları (Márquez-Vera, Cano, Romero ve Ventura, 2013) ve Web 2.0 araçları (viki, forum, blog vb.) da EVM

çalışmalarında verinin elde edildiği ortamlar olarak kullanılmaktadır. Çalışmanın amacına göre bu ortamlar aynı zamanda analiz sonuçlarını uygulamak için de kullanılabilir.

EVM çalışmalarında kullanılan veriler ise, verinin elde edildiği ortama (geleneksel, bilgisayar destekli, web destekli vb.) ve araştırma problemlerine göre değişiklik göstermektedir. Bu veriler; log verileri, etkileşim verileri, davranış verileri, gözlem verileri, ölçek verileri ve kontrollü deneylerde ölçülen ölçümlerden elde edilen veriler olabilmektedir (Cristóbal, Sebastián, Mykola ve Ryan, 2010). Bunun dışında göz hareketleri ile ilgili metrikler (Akçapınar, Cosgun ve Altun, 2011; Amershi ve Conati, 2009), ya da öğrenciler tarafından yazılan mesajlar (Cristóbal Romero ve diğerleri, 2008) ve açık uçlu sınavlarda öğrencilerin verdiği cevaplar (Cavalcanti, Pires, Cavalcanti ve Pires, 2012) gibi metinler de EVM çalışmalarında veri kaynağı olarak kullanılabilir.

VM çalışmalarında olduğu gibi EVM çalışmalarında da analiz aşamasına geçilmeden önceki en önemli adım verilerin ön işleme sürecidir. Ön işlemenin amacı verinin kalitesini artırmak ve seçilen analiz için ideal değişkenleri elde etmektir (Hämäläinen ve Vinni, 2010). Aynı zamanda anlaşılabilir sonuçların üretilmesi açısından da önem arz etmektedir. Ön işleme süreci, verinin türüne ve karmaşıklığına göre değişiklik göstermekle birlikte VM çalışmalarında en çok kullanılan ön işleme adımları veri temizleme, dönüştürme, birleştirme ve özellik seçme olarak özetlenebilir.

EVM sürecinin en önemli bölümü, eğitsel ortamlardan elde edilen ve ön işleme sürecinden geçmiş veriye VM yöntemlerinin uygulandığı aşamadır. Burada seçilen problem durumuna göre uygun VM yöntemi belirlenir ve analizler yapılır. Yapılan tarama çalışmaları EVM araştırmalarında en çok tercih edilen VM yöntemlerinin sırasıyla: sınıflama; kümeleme; regresyon ve birliktelik kuralları olduğuna işaret etmektedir (Peña-Ayala, 2014; C. Romero ve Ventura, 2010). Bu yöntemler EVM çalışmalarında kullanılan VM yöntemlerinin %91'ini oluşturmaktadır (Peña-Ayala, 2014).

EVM sürecinin son aşaması ise analiz sonucu elde edilen bilgilerin, model veya örüntülerin yorumlanarak karar verme süreçlerinde ya da eğitsel ortamın iyileştirilmesi amacıyla kullanılması aşamasıdır (García ve diğerleri, 2011; Cristobal Romero ve Ventura, 2013).

Yöntem

Araştırmada yöntem olarak Şekil 1'de özetlenen EVM süreci izlenmiştir ve çalışmanın yöntem kısmı bu doğrultuda yazılmıştır.

Veri Toplama Süreci

Araştırma Türkiye'de bir devlet üniversitesinde, Bilgisayar ve Öğretim Teknolojileri Eğitimi Bölümü'nde "Bilgisayar Donanımı" dersine kayıtlı 76 üniversite ikinci sınıf öğrencisi (41 Kız, 35 Erkek) ile yürütülmüştür. Öğrenciler 14 Hafta süresince yüz yüze derslere ek olarak, ders dışı tüm etkinlikleri gerçekleştirmek amacıyla ilk yazar tarafından geliştirilen çevrimiçi öğrenme ortamını kullanmışlardır. Öğrencilerin öğrenme süreçleri ile ilgili olarak daha fazla veriye ulaşmak için mevcut çevrimiçi öğrenme ortamları yerine bu ortam tercih edilmiştir. Haftada iki saat olan yüz-yüze derslerde bilgisayarın temel çalışma mantığı ve bilgisayarını oluşturan donanımlarla ilgili olarak öğrencilere sunumlar yapılmıştır. Şekil 2'de ekran

görüntüsü sunulan çevrimiçi ortam ise öğrencilerin, sınıf ortamında öğrendikleri bilgiler ve kavramlar ile ilgili olarak yansıma yazmaları, yazdıkları yansımaları etiketlemeleri, tartışmalara katılmaları, ders ile ilgili kaynak ve duyuruları takip etmeleri amacıyla yapılandırılmıştır. Öğrencilere haftalık olarak o derste ele alınacak kavramlar verilmiş ve bu kavramlarla ilgili olarak yansıma yazmaları istenmiştir. Yansımaların kısa ve öz olması istendiği için içerik alanı 500 karakter ile sınırlandırılmıştır ve öğrencilerin başka bir kaynaktan doğrudan alıntı yapmasının önüne geçebilmek için bütün sistemde metin girişi yapılan alanlarda “kopyala – yapıştır” özelliği devre dışı bırakılmıştır.

puan	cevap	görüntüleme	konu başlığı	etiketler	gönderen	gönderme tarihi
0	0	10	RAS, RAS-TO-RAS, CAS		Arayışkan	23.12.2013
0	0	4	HHD, SSD		Arayışkan	23.12.2013
0	0	6	Genişleme yuvaları (External buses)	- Anakarta harici donanım elemanlarını bağlamak için kullanılan veri...	Arayışkan	23.12.2013
1	1	17	SCSI Kartı	SCSI, Small Computer System Interface'ın (Küçük Bilgisayar Sistem Arab...	Arayışkan	18.12.2013
0	1	4	TV-Kartı	TV kartı; Bilgisayarlarda TV izlemeye olanak sağlayan bilgisayar parça...	Arayışkan	18.12.2013
0	1	3	Fax-Modem Kartları	Bilgisayarların telefon kabloları üzerinden lokal yada genel ağlara ba...	Arayışkan	18.12.2013
1	1	6	RJ-45 Konnektörü	Üzerinde 8 adet metal bağlantı noktası bulunan konnektör. Bilgisayarla...	Arayışkan	18.12.2013
0	1	6	Ethernet Kartı	Ethernet kartı, bilgisayar ağlarında bilgisayarlar arasında iletişim...	Arayışkan	18.12.2013
0	1	2	Mikrofon Girişi	Joystick/MIDI port : Joystick ve MIDI aygıtlarının bağlanmasını sağlar...	Arayışkan	18.12.2013

Şekil 2. Çevrimiçi öğrenme ortamı

Öğrenciler aynı zamanda diğerleri tarafından yazılan iletileri okuyabilmekte, birbirlerinin iletilerine yorum yapabilmekte ve yazılan ileti ve yorumlarla ilgili değerlendirmede (katılıyorum - katılmıyorum) bulunabilmektedirler. Sistem tarafından öğrencinin ortama her girişi ayrı bir oturum olarak ele alınmaktadır. Öğrencinin çevrimiçi ortamda yaptığı her türlü aktivite ise bu oturum ile ilişkili olarak kayıt edilmektedir. Bu kayıtlar sadece sayfa gezinimlerini değil aynı zamanda sayfa içerisindeki tüm etkileşimleri de kapsamaktadır (içerik ile ilgili yorumların açılması vb.). Araştırmada kullanılan değişkenler bu verilerin analiz edilmesi ile üretilmiştir.

Öğrencilere sistemi kullanmaları yönünde bir zorunluluk belirlenmemiştir ancak haftalık olarak verilen kavramlarla ilgili yazacakları yansımaların final notlarına katkı sağlayacağı belirtilmiştir. Araştırma kapsamında öğrencilerin akademik performansları hesaplanırken bu katkılar elimine edilmiştir, sadece yazılı sınavlardan aldıkları notlar dikkate alınmıştır.

Akademik Performans

Üçüncü ve dördüncü araştırma problemleri kapsamında, elde edilen kümeler akademik performans açısından analiz edilmiştir. Bu bağlamda öğrenci performansının göstergesi olarak öğrencilerin Bilgisayar Donanımı dersine ait dönem sonundaki ders geçme notları dikkate alınmıştır. Öğrenciler dönem içerisinde iki ara sınav, bir final sınavı olmak üzere toplam üç adet yazılı sınav olmuştur. Öğrencilerin ders geçme notları, ara sınavların %25'i ve final sınavının %50'si alınarak hesaplanmıştır. Tablo 1'de bu puanlara ilişkin tanımlayıcı istatistikler verilmiştir. Kümelerde yer alan öğrencilerin geçme – kalma durumlarını incelemek için de geçme notları, üniversitenin lisans yönetmeliğinde yer alan göstergeler dikkate alınarak “Kaldı” (n = 27, ders geçme notu ≤ 50) ve “Geçti” (n = 49, ders geçme notu > 50) şeklinde kategorik hale dönüştürülmüştür.

Tablo 1. Yazılı sınavlara ilişkin tanımlayıcı istatistikler

Sınav	N	Ortalama	SS
1. Ara sınav	75	54,20	21,77
2. Ara sınav	73	52,33	17,92
Final sınavı	74	60,59	17,51
Ders geçme notu	76	55,49	18,20

Veri Seti

Veri analizine başlamadan önce çevrimiçi öğrenme ortamının veri tabanındaki bilgilerin ve öğrencilerin der geçme notlarının bir arada tutulduğu bir analiz veri tabanı oluşturulmuştur. Bu veri tabanı, 76 öğrencinin 14 haftalık kullanım verisini ve öğrencilerin akademik performansını gösteren iki değişkeni (sürekli ve kategorik) içermektedir. Bu veriler: 3803 oturum, 119921 gezinim, 4566 yazılı metin, tartışma ortamında 62 adet soru, 297 cevap, toplam 8601 değerlendirme ve yazılan iletileri etiketlemek için kullanılan 3937 etiket verisini içermektedir.

Araştırmada Kullanılan Değişkenler

Yukarıda ayrıntıları verilen ham verinin doğrudan analizlerde kullanılması olanaklı değildir, bu nedenle bir takım dönüşümlerden geçerek analizlerde kullanılacak değişkenlerin elde edilmesi gerekmektedir. Analizlerde kullanılacak bu değişkenlerin belirlenmesinde öğrencilerin öğrenme ortamında gerçekleştirdiği etkileşimler göz önüne alınmış ve bu etkileşimleri yansıtacak, ölçülebilir metrikler oluşturulmaya çalışılmıştır. Veri tabanındaki ham veriden bu değişkenlere ilişkin verileri elde etme işi geliştirilen bir araç yardımıyla otomatik olarak yapılmıştır. C# programlama dili kullanılarak geliştirilen araç, öğrenme ortamına ilişkin verilerin tutulduğu MS SQL veri tabanına bağlanarak SQL sorguları aracılığı ile bu verileri almak için kullanılmıştır. Tablo 2'de bu değişkenler ve açıklamalarına yer verilmiştir. Bu değişkenlerden hangilerinin analizlerde kullanılacağına ise veri ön işleme sürecinde karar verilmiştir.

Tablo 2. Ham veriden üretilen değişkenler ve açıklamaları

No	Değişken	Açıklama
1	t_BenzersizIleti*	İleti yazılan farklı gün sayısı
2	t_SoruDeğerlendirme*	Tartışma ortamında soru değerlendirme sayısı
3	t_BenzersizOturum*	Ortama giriş yapılan farklı gün sayısı
4	t_Gezinim*	Ortamda yapılan toplam gezinim sayısı
5	t_EtiketKullanım*	Yazılan iletilerde kullanılan etiket sayısı
6	t_Ileti*	Öğrenci tarafından yazılan ileti sayısı
7	t_Etiket*	Öğrenci tarafından ortama eklenen etiket sayısı
8	t_IletiGezinim	İletilerin bulunduğu sayfada yapılan toplam gezinim
9	t_CevapOkuma	Tartışma ortamında yazılan cevapları açma sayısı
10	t_Cevap*	Tartışma ortamında yazılan toplam cevap sayısı
11	t_BildirimGezinim	Kullanıcılar tarafından yapılan aktivitelerin yer aldığı sayfada gezinim sayısı
12	t_Oturum*	Gerçekleştirilen toplam oturum sayısı
13	t_CevapDeğerlendirme	Tartışma ortamında değerlendirilen cevap sayısı
14	t_IletiDeğerlendirme	Diğer öğrenciler tarafından yazılan iletileri değerlendirme sayısı
15	t_KopyalaYapıştır	İleti yazarken kopyala – yapıştır kullanım sayısı
16	t_IletiOkuma	Diğerleri tarafından yazılan iletileri açma sayısı
17	t_OturumSuresi*	Ortamda kalınan toplam süre
18	o_TusKullanım	İleti yazarken ortalama tuş kullanım sayısı
19	o_SilmeTusuKullanım	İleti yazarken ortalama silme tuşu kullanım sayısı
20	t_KaynakGezinim	Kaynakların yer aldığı sayfada gezinim sayısı
21	t_DuyuruGezinim	Duyuruların yer aldığı sayfada gezinim sayısı
22	t_TartismaGezinim	Tartışma sayfasında gezinim sayısı
23	o_SayfaDeaktif	İleti yazarken ortalama sayfanın arka planda kalma sayısı
24	t_Yorum	Yazılan iletilere yapılan yorum sayısı
25	t_YorumDeğerlendirme	Yazılan yorumları değerlendirme sayısı
26	o_IletiYazmaSuresi	Ortalama ileti yazma süresi
27	t_KopyaIletiSayisi	Benzerlik oranı yüksek ileti sayısı
28	t_Soru	Öğrencinin tartışma ortamında yazdığı toplam soru sayısı

* İşaretili değişkenler ön işleme sürecinde "özellik seçme" işlemi sonucu seçilen ve analizde kullanılan değişkenlerdir.

Veri Ön İşleme Süreci

Bu aşamada ilk olarak değişkenlere ilişkin tanımlayıcı istatistikler incelenmiş ve öğrenciler tarafından katılım oranı düşük olduğu için t_Soru değişkeni analizden çıkartılmıştır. Yirmi yedi değişkenin tamamını kullanmak yerine daha az sayıda değişkenle analizi gerçekleştirmek için özellik seçme yönteminden yararlanılmıştır. Özellik seçme (feature selection) yöntemi veri madenciliği çalışmalarında (özellikle tahmin analizlerinde) önemli değişkenleri belirlemek amacıyla ya da sonuçların güvenilirliğini azaltmadan analize girecek değişken sayısını azaltmak amacıyla kullanılan bir yöntemdir (Lopez, Luna, Romero ve Ventura, 2012; Márquez-Vera ve diğerleri, 2013). Bu sayede, elde edilen sonuçların yorumlanması ve uzman olmayan kişiler tarafından anlaşılması daha kolay olmaktadır. Burada da benzer bir yaklaşım izlenerek 27 değişkenden akademik performans ile ilişkili 10 değişken seçilmiştir. Değişkenlerin belirlenmesinde Gini indeks kriteri dikkate alınmıştır. Seçilen değişkenler ve Gini indeksleri Tablo 3'te sunulmuştur. Ön işlemenin son adımında analize girecek tüm değişkenler z puanlarına çevrilerek standartlaştırılmıştır.

Tablo 3. Gini indeksine göre en önemli 10 değişken

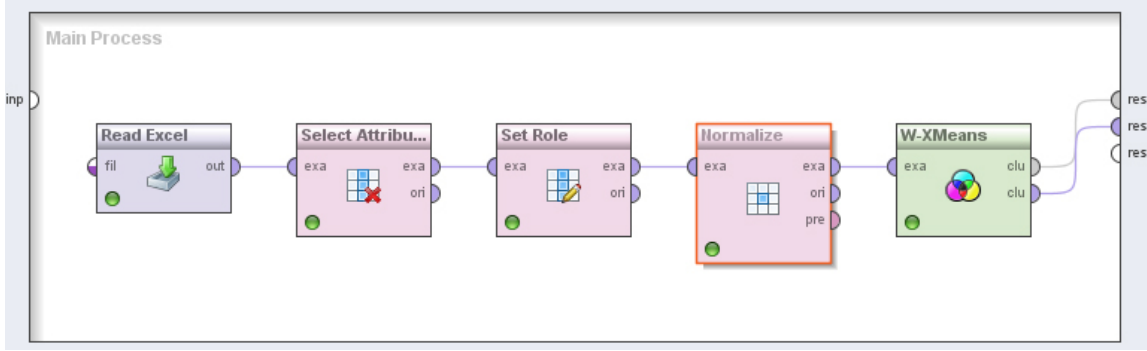
No	Değişken	Gini indeksi
1	t_SoruDeğerlendirme	0,083
2	t_BenzersizOturum	0,078
3	t_BenzersizIleti	0,076
4	t_Ileti	0,056
5	t_Gezinim	0,042
6	t_Oturum	0,037
7	t_EtiketKullanım	0,033
8	t_Etiket	0,031
9	t_Cevap	0,028
10	t_OturumSuresi	0,024

Veri Analizi

Çalışma kapsamında çevrimiçi öğrenme ortamında benzer kullanım davranışları sergileyen farklı öğrenci gruplarının belirlenmesi amacıyla kümeleme analizinden yararlanılmıştır. Kümeleme analizinin amacı, küme içi uzaklıkları azaltıp, kümeler arası uzaklıkları artırarak, kendi içerisinde benzerlik gösteren elemanlardan oluşan farklı kümeler elde etmektedir (Chien-Sing ve Singh, 2004). Kümeleme analizinde verinin kaç farklı kümeye ayrılacağına belirlenmesi konusunda ise iki farklı yaklaşım bulunmaktadır. Bu sayı ön tanımlı olarak belirtilmektedir ya da çeşitli algoritmalar kullanılarak veriye en uygun küme sayısı bulunmaktadır (Baker, Corbett, Koedinger ve Roll, 2006; Bienkowski, Feng ve Means, 2012).

Kümeleme analizi tanımlayıcı bir yöntem olduğu için sınıflama analizinde olduğu gibi farklı algoritmaları performans metrikleri açısından karşılaştırarak en iyi performans gösteren algoritmayı seçmek gibi bir durum olanaklı olmamaktadır. Bu nedenle, veriler literatürde çok kullanılan iki kümeleme yöntemi X-Ortalamlar (X-Means) ve EM (Expectation Maximization) ile analiz edilmiş ve sonuçları karşılaştırılmıştır. Her iki yöntem için de ideal küme sayısı algoritmalar tarafından belirlenmiştir. Kümeleme analizleri Rapidminer VM yazılımının Weka eklentisinde bulunan W-XMeans ve W-EM algoritmaları kullanılarak Şekil 3 ve Şekil 4'te verilen analiz süreçlerine göre gerçekleştirilmiştir. Aşağıda bu kümeleme yöntemleri ile ilgili ayrıntılı bilgi verilmiştir.

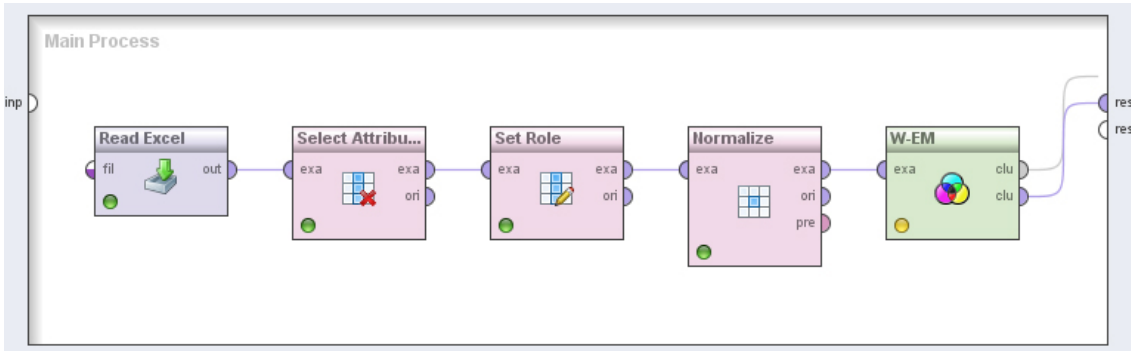
X-Ortalamlar: X-Ortalamlar kümeleme algoritması, Pelleg ve Moore (2000) tarafından geliştirilmiştir ve K-Ortalamlar (K-Means) kümeleme algoritmasının zayıf yönlerinin geliştirilmesi ile ortaya çıkmış bir kümeleme algoritmasıdır. MacQueen (1967) tarafından bulunan K-Ortalamlar kümeleme algoritmasının amacı n sayıdaki elemanı ön tanımlı olarak belirtilen k tane kümeye ayırmaktır. Bunu yaparken ilk olarak k adet hayali nokta belirlenir. Bu noktalar rastgele olarak seçilir ve küme merkezi olarak adlandırılır. Daha sonra çeşitli uzaklık metrikleri kullanılarak her bir eleman en yakın olduğu küme merkezine atanır. Bu işlem tüm elemanlar bir kümeye atanana kadar devam eder. Yeni oluşan kümelerin merkezi ilk başta rastgele olarak seçilen küme merkezlerinden farklı olacağı için, yeni küme merkezleri hesaplanır ve her bir eleman tekrar bu küme merkezlerine atanır. Bu işlem küme merkezleri sabitlenene kadar devam eder ve bu sayede veri k tane kümeye ayrılmış olur.



Şekil 3. X-Ortalamlar kümeleme analiz süreci

X-Ortalamlar kümeleme algoritmasının yukarıda anlatılan sürece getirdiği en önemli katkı k değerinin (küme sayısının) otomatik olarak belirlenmesidir. Algoritma Bayes Bilgi Kriteri'ni (Bayesian Information Criterion) kullanarak optimal küme sayısını belirlemektedir. Diğer bir katkı ise büyük veri setlerinde hızlı çalışması konusunda olmuştur.

EM Kümeleme: EM kümeleme algoritmasının adı Expectation Maximization (Maksimum Beklenti) kelimelerinin baş harfinden oluşmaktadır. X-Ortalamlar kümeleme algoritmasından farklı olarak burada bir eleman bir kümeye uzaklık ölçütleri ile kesin bağlı değildir, bunun yerine her bir kümeye bağlı olma olasılığı vardır (Nisbet, Elder ve Miner, 2009). Bağımsız değişkenler kullanılarak küme sayısı kadar farklı ortalama ve standart sapmada dağılımlar oluşturulur ve elemanlar yakın oldukları dağılıma göre kümelere yerleştirilir (Witten, Frank ve Hall, 2011). Burada küme sayısı kullanıcı tarafından belirlenebileceği gibi çapraz geçerlilik yöntemi kullanılarak algoritma tarafından da belirlenebilir.



Şekil 4. EM Kümeleme analiz süreci

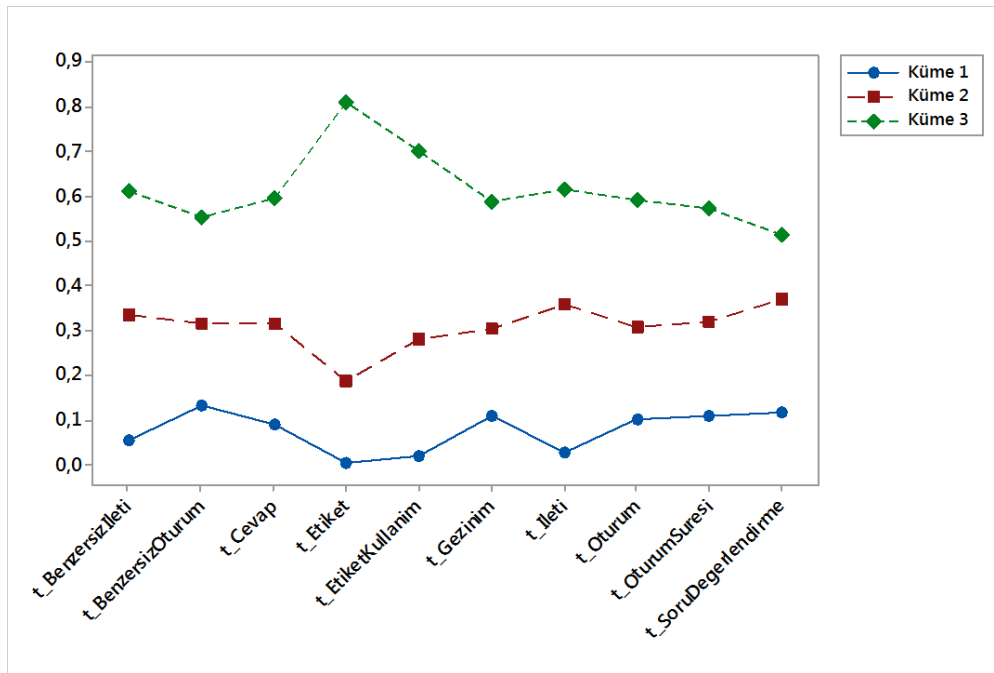
Bulgular

Bu bölümde ilk olarak birinci ve ikinci araştırma problemleri kapsamında X-Ortalamlar ve EM kümeleme algoritmaları kullanılarak gerçekleştirilen kümeleme analizlerine ilişkin bulgulara yer verilmiştir. X-Ortalamlar algoritması ile yapılan analiz sonucu verinin ideal olarak 3 farklı kümeye ayrılabilceği tespit edilmiştir. Tablo 4'te bu üç kümede yer alan öğrencilerin analizde kullanılan değişkenler açısından ortalama değerlerine yer verilmiştir.

Tablo 4. X-Ortalamalar algoritmasına göre küme ortalamaları

Değişkenler	Küme 1 (n = 21)	Küme 2 (n = 32)	Küme 3 (n = 23)
<i>t_BenzersizIleti</i>	0,74	4,57	8,41
<i>t_BenzersizOturum</i>	8,26	19,52	34,31
<i>t_Cevap</i>	0,96	3,38	6,38
<i>t_Etiket</i>	0,09	4,57	19,97
<i>t_EtiketKullanım</i>	2,52	38,24	96,13
<i>t_Gezinim</i>	463,35	1310,86	2554,25
<i>t_Ileti</i>	3,78	53,29	91,38
<i>t_Oturum</i>	13,74	42,14	81,31
<i>t_OturumSuresi</i>	521,22	1536,71	2777,78
<i>t_SoruDeğerlendirme</i>	10,61	33,33	46,75

Şekil 5'te küme ortalamalarının 0 - 1 aralığında normalleştirilmesi ile elde edilen grafik incelendiğinde Küme 1'de yer alan öğrencilerin, ortama az sayıda giriş yapan, ortamda az zaman geçiren, tartışmalara katılmayan, öğrendikleri kavramlarla ilgili az sayıda ileti yazan öğrenciler oldukları görülmektedir.

**Şekil 5.** X-Ortalamalar algoritmasına göre standartlaştırılmış küme ortalamaları

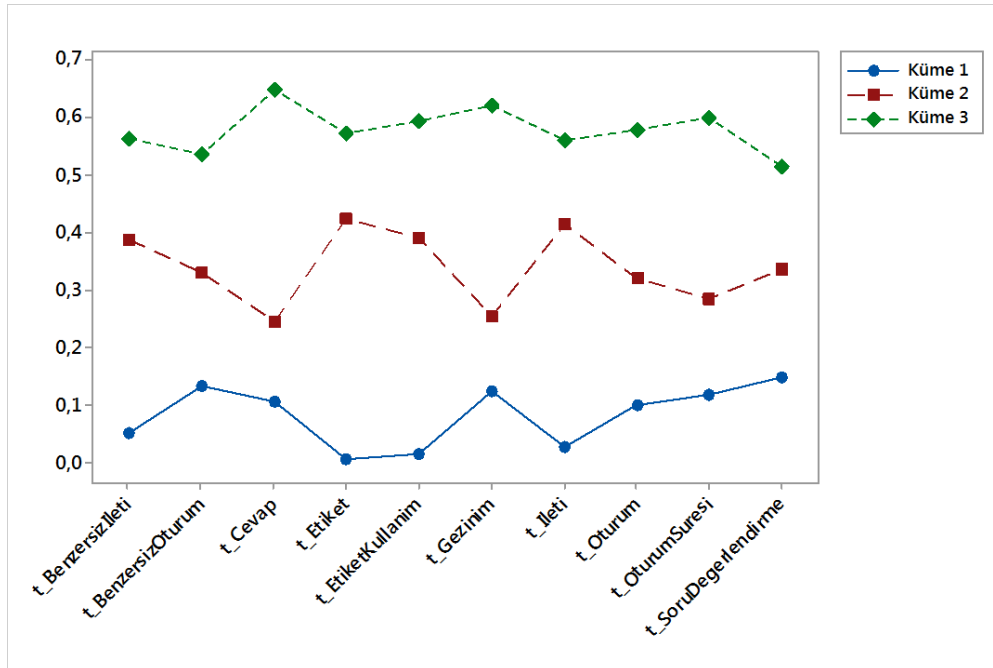
Küme 3 ise tam tersi şekilde ortama düzenli olarak giriş yapan, tartışmalara aktif olarak katılan, ortamda diğerlerine göre daha fazla zaman geçiren öğrencilerden oluşmaktadır. Küme 2'de yer alan öğrenciler ise ortamda orta düzeyde aktivite sergileyen, ortalama değerleri Küme 1'de yer alan öğrencilerden daha yüksek ama Küme 3'te yer alan öğrencilerden daha düşük olan öğrencilerin yer aldığı küme olarak tanımlanabilir.

EM algoritması kullanılarak yapılan kümeleme analizi sonucu da verinin ideal olarak 3 farklı kümeye ayrılacağı tespit edilmiştir. Tablo 5'te bu kümelerde yer alan öğrencilere ilişkin değişken ortalamaları yer almaktadır.

Tablo 5. EM Algoritmasına göre küme ortalamaları

Değişkenler	Küme 1 (n=17)	Küme 2 (n = 27)	Küme 3 (n = 32)
<i>t_BenzersizIleti</i>	0,82	6,34	9,24
<i>t_BenzersizOturum</i>	9,67	24,13	39,12
<i>t_Cevap</i>	1,44	3,34	8,88
<i>t_Etiket</i>	0,11	13,38	18,00
<i>t_EtiketKullanım</i>	2,41	66,97	101,71
<i>t_Gezinim</i>	671,26	1379,72	3390,94
<i>t_Ileti</i>	4,33	73,13	98,41
<i>t_Oturum</i>	16,59	53,69	96,29
<i>t_OturumSuresi</i>	692,67	1679,88	3570,00
<i>t_SoruDeğerlendirme</i>	15,37	34,94	53,35

Şekil 6'da ise normalleştirilmiş küme ortalamalarının grafiği verilmiştir. Değişken ortalamaları incelendiğinde burada da elde edilen kümelerin isimlendirilmesinin X-Ortalamlar algoritması ile aynı şekilde yapılabileceği görülmektedir.



Şekil 6. EM Algoritmasına göre normalleştirilmiş küme ortalamaları

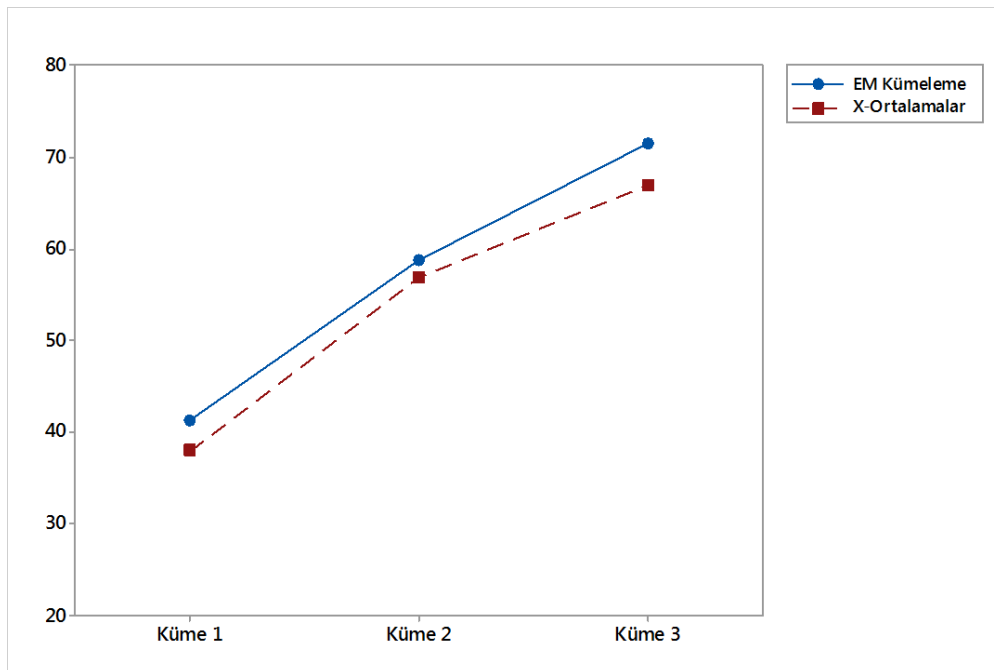
Farklı algoritmalarla yapılan kümeleme analizi sonucu elde edilen küme ortalamaları incelendiğinde kümelerde yer alan öğrenci sayılarında ve küme ortalamalarında farklılıklar olmakla birlikte her iki algoritmanın da öğrencileri benzer özellikte üç farklı kümeye ayırdığı görülmektedir.

Üçüncü araştırma sorusu kapsamında “Her iki kümeleme algoritmasına göre elde edilen farklı kümeler bu kümelerde yer alan öğrencilerin akademik performanslarına göre tanımlanabilir mi?” sorusuna cevap aranmıştır. Tablo 6’da farklı kümeleme algoritmalarına göre kümelerde yer alan öğrencilerin ortalama ders geçme notları verilmiştir.

Tablo 6. EM ve X-Ortalamlar kümeleme algoritmalarına göre kümelerde yer alan öğrencilerin ortalama ders geçme notları

Küme	EM Kümeleme	X-Ortalamlar
Küme 1	41,33	38,00
Küme 2	58,88	56,95
Küme 3	71,59	67,09

Buna göre her iki kümeleme algoritmasının da benzer sonuçlar ürettiği görülmektedir. Şekil 7’de görsel olarak verilen ortalamalar incelendiğinde her iki algoritma için de Küme 1’de yer alan öğrencilerin derste düşük performans gösteren öğrenciler olduğu, Küme 2’de yer alan öğrencilerin ortalama performansa sahip öğrenciler olduğu ve Küme 3’de yer alan öğrencilerin de diğerlerine göre daha yüksek performansa sahip öğrenciler olduğu görülmektedir.



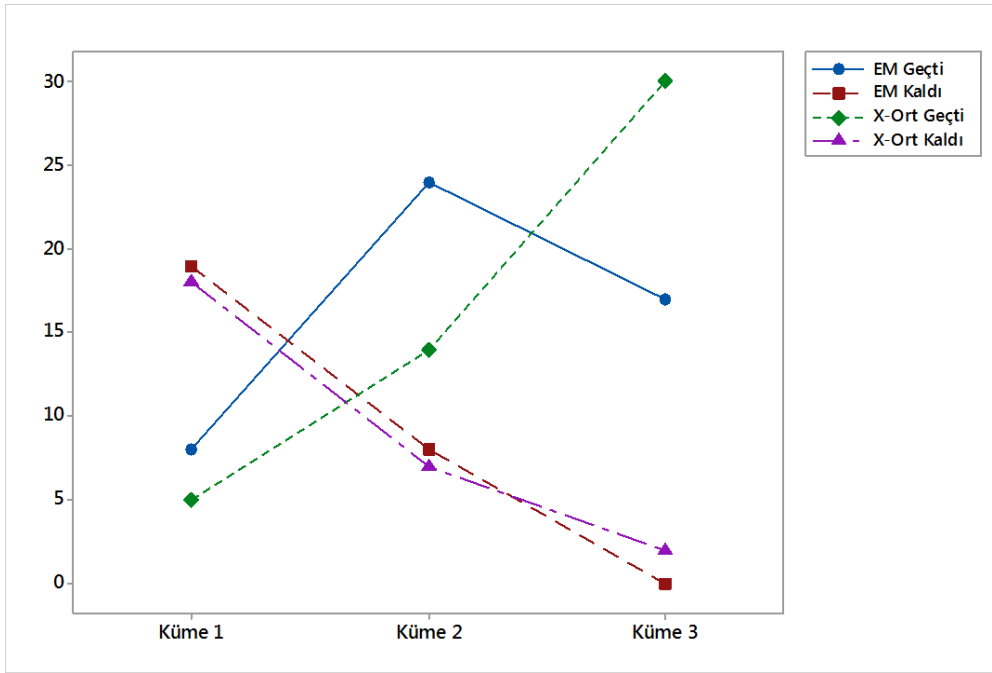
Şekil 7. EM ve X-Ortalamlar kümeleme algoritmalarının ders geçme notlarına göre karşılaştırması

Dördüncü araştırma sorusu kapsamında dersi geçen ve dersten kalan öğrencilerin kümelerine göre dağılımları incelenmiştir (Tablo 7). EM Kümeleme algoritmasına göre Küme 1’de yer alan 27 öğrenciden 8’inin dersi geçtiği 19’unun kaldığı, Küme 2’de yer alan 32 öğrenciden 24’ünün dersi geçtiği, 8’inin kaldığı, Küme 3’de yer alan 17 öğrencinin ise tamamının dersi geçtiği görülmektedir. X-Ortalamlar kümeleme algoritmasına göre ise Küme 1’de yer alan 23 öğrenciden 5’inin dersi geçtiği 18’inin kaldığı, Küme 2’de yer alan 21 öğrenciden 14’ünün dersi geçtiği, 7’sinin kaldığı, Küme 3’de yer alan 32 öğrenciden ise 30’unun

dersi geçtiği, 2'sinin kaldığı görülmektedir. Şekil 8'de görsel olarak da verilen bu sonuçlar incelendiğinde algoritmaların benzer sonuçlar ürettiği görülmektedir.

Tablo 7. Dersi Geçen ve dersten kalan öğrencilerin kümelere göre dağılımı

Küme	EM Kümeleme			X-Ortalamalar		
	<i>n</i>	<i>Geçti</i>	<i>Kaldı</i>	<i>n</i>	<i>Geçti</i>	<i>Kaldı</i>
Küme 1	27	8	19	23	5	18
Küme 2	32	24	8	21	14	7
Küme 3	17	17	0	32	30	2



Şekil 8. Dersi geçen ve dersten kalan öğrencilerin kümelere göre dağılımı

Sonuçlar ve Öneriler

Bu çalışmada öğrencilerin çevrimiçi öğrenme ortamındaki etkileşim verileri uzaklık tabanlı (X-Ortalamalar) ve olasılık tabanlı (EM Kümeleme) iki farklı kümeleme algoritması ile analiz edilerek benzer kullanım örüntüsü sergileyen öğrenci gruplarının ortaya çıkartılması amaçlanmıştır. Her iki kümeleme algoritmasında da verinin kaç farklı kümeye ayrılacağı (ideal küme sayısı) algoritmalar tarafından belirlenmiştir. Analiz sonuçları öğrencilerin ideal olarak üç farklı kümeye ayrılabilceğini göstermiştir. Kümelerin isimlendirilmesi konusunda etkileşim düzeyi ve akademik performansa göre iki farklı yaklaşım izlenmiştir. İlk olarak kümeler, analizde kullanılan değişken ortalamaları dikkate alınarak Çok aktif öğrenciler (Küme 3), Aktif öğrenciler (Küme 2) ve Aktif olmayan öğrenciler (Küme 1) şeklinde isimlendirilmiştir. Daha sonra elde edilen kümelerdeki öğrencilerin ders geçme notları incelenmiş ve elde edilen kümeler ders geçme notları ile ilişkili olarak Yüksek öğrenenler (Küme 3), Orta öğrenenler (Küme 2) ve Düşük öğrenenler (Küme 1) şeklinde isimlendirilmiştir.

Yapılan çalışmalar çevrimiçi öğrenme ortamlarındaki etkileşim verileri ile öğrencilerin akademik performansları arasındaki ilişkiyi ortaya koymaktadır (Cristobal Romero ve diğerleri,

2013; Cristóbal Romero, López, Luna ve Ventura, 2013). Bu nedenle öğrencilerin çevrimiçi öğrenme ortamındaki aktivite düzeyi ile akademik performansları arasındaki pozitif ilişki beklenen bir durumdur; fakat bu durumun veriye dayalı olarak modellenmesi veriye dayalı olarak karar verme mekanizmalarının oluşturulması açısından önemlidir.

Burada elde edilen küme bilgileri ile oluşturulan sınıflama modelleri, ileriki araştırmalarda ortama yeni giren öğrencilerin otomatik olarak sınıflandırılmasında kullanılabilir. Ortamda düşük düzeyde aktivite gösteren öğrenciler haftalık olarak belirlenebilir ve yapılacak müdahalelerle dönem sonundaki olası başarısızlıkların önüne geçilebilir. Bu veriler, öğretim tasarımcılarına, sisteme yapacakları müdahalelere karar verme ve etkilerini izleme sürecinde yardımcı olabilir. Öğrenciler de bu bilgileri kendi öğrenme süreçlerini izlemek amacıyla kullanabilirler.

Bu tür modeller uyarlanabilir öğrenme ortamlarında kullanılan otomatik sınıflama modellerine girdi olabilecek değişkenleri üretmek amacıyla kullanılabilir (Amershi ve Conati, 2009). Bu sayede gerçek zamanlı uyarlamalar mümkün olabilir (Bouchet, Harley, Trevors ve Azevedo, 2013). Bu modeller aynı zamanda çevrimiçi öğrenme ortamlarında benzer davranış gösteren öğrencileri tanımlamak için kullanılabilir (Valdiviezo ve diğerleri, 2013); öğretmenler bu bilgileri yeni gelen öğrencileri sınıflandırmak için (Lopez ve diğerleri, 2012) ya da işbirlikli etkinliklerde çalışma gruplarını oluşturmak için kullanabilirler (Cristóbal Romero ve diğerleri, 2008).

Seçilen kümeleme algoritmaları verileri kümeye atarken ve ideal küme sayısını bulurken farklı yöntemler kullanmaktadır. Bu nedenle, verinin her iki algoritma tarafında da üç kümeye ayrılması ve bu kümelerin de aktivite düzeyine ve akademik performansa göre benzer şekilde ayrılması elde edilen sonuçların güvenilirliğini artırmaktadır. Ancak; veri madenciliği çalışmalarında tüm durumlarda kullanılacak tek bir model üretmek mümkün olmamaktadır (Gašević, Dawson, Rogers ve Gasevic, 2016; Cristobal Romero ve diğerleri, 2013). Bu nedenle, burada sunulan sonuçlar ele alınan donanım dersi, dersin işleniş biçimi (yüz-yüze / çevrimiçi), kümelemede kullanılan değişkenler ve akademik performans göstergesi olarak alınan ders geçme notları ile sınırlıdır.

Kaynakça

- Akçapınar, G., Cosgun, E. ve Altun, A. (2011). Prediction of Perceived Disorientation in Online Learning Environment with Random Forest Regression. Paper presented at the Proceedings of the 4th International Conference on Educational Data Mining., July 6-8, Eindhoven, Netherlands.
- Ali, L., Asadi, M., Gašević, D., Jovanović, J. ve Hatala, M. (2013). Factors influencing beliefs for adoption of a learning analytics tool: An empirical study. *Computers & Education*, 62(0), 130-148. doi: <http://dx.doi.org/10.1016/j.compedu.2012.10.023>
- Amershi, S. ve Conati, C. (2009). Combining Unsupervised and Supervised Machine Learning to Build User Models for Exploratory Learning Environments. *Journal of Educational Data Mining*, 1(1), 71-81.
- Baker, R. J. d., Corbett, A., Koedinger, K. ve Roll, I. (2006). Generalizing Detection of Gaming the System Across a Tutoring Curriculum. In M. Ikeda, K. Ashley ve T.-W. Chan (Eds.), *Intelligent Tutoring Systems (Vol. 4053, pp. 402-411)*: Springer Berlin Heidelberg.

- Beal, C. R., Qu, L. ve Lee, H. (2006). Classifying learner engagement through integration of multiple data sources. Paper presented at the Proceedings of the 21st national conference on Artificial intelligence - Volume 1, Boston, Massachusetts.
- Bienkowski, M., Feng, M. ve Means, B. (2012). Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief. Washington, D.C.
- Bouchet, F., Harley, J., Trevors, G. ve Azevedo, R. (2013). Clustering and profiling students according to their interactions with an intelligent tutoring system fostering self-regulated learning. *Journal of Educational Data Mining*, 5(2).
- Cavalcanti, E. R., Pires, C. E. S., Cavalcanti, E. P. ve Pires, V. F. (2012). Detection and Evaluation of Cheating on College Exams using Supervised Classification. *Informatics in Education*, 11(2), 169-190.
- Chien-Sing, L. ve Singh, Y. P. (2004, 30 Aug.-1 Sept. 2004). Student modeling using principal component analysis of SOM clusters. Paper presented at the Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on.
- Cristóbal, R., Sebastián, V., Mykola, P. ve Ryan, S. J. D. B. (2010). Introduction Handbook of Educational Data Mining (pp. 1-6): CRC Press.
- García, E., Romero, C., Ventura, S. ve de Castro, C. (2011). A collaborative educational association rule mining tool. *The Internet and Higher Education*, 14(2), 77-88. doi: 10.1016/j.iheduc.2010.07.006
- Gašević, D., Dawson, S., Rogers, T. ve Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68-84. doi: <http://dx.doi.org/10.1016/j.iheduc.2015.10.002>
- Greller, W. ve Drachler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*, 15 %6(3), 42-57.
- Hämäläinen, W. ve Vinni, M. (2010). Classifiers for Educational Data Mining Handbook of Educational Data Mining (pp. 57-74): CRC Press.
- Lopez, M. I., Luna, J. M., Romero, C. ve Ventura, S. (2012). Classification via clustering for predicting final marks based on student participation in forums. Paper presented at the 5th International Conference on Educational Data Mining, EDM 2012, Chania, Greece.
- MacQueen, J. (1967, 1967). Some methods for classification and analysis of multivariate observations. Paper presented at the Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, Berkeley, Calif.
- Márquez-Vera, C., Cano, A., Romero, C. ve Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data. *Applied Intelligence*, 38(3), 315-330. doi: 10.1007/s10489-012-0374-8
- Moreno, L., Gonzalez, C., Castilla, I., Gonzalez, E. ve Sigut, J. (2007). Applying a constructivist and collaborative methodological approach in engineering education. *Computers & Education*, 49(3), 891-915. doi: <http://dx.doi.org/10.1016/j.compedu.2005.12.004>
- Nisbet, R., Elder, J. ve Miner, G. (2009). Chapter 7 - Basic Algorithms for Data Mining: A Brief Overview. In R. Nisbet, J. Elder ve G. Miner (Eds.), *Handbook of Statistical Analysis and Data Mining Applications* (pp. 121-150). Boston: Academic Press.
- Pelleg, D. ve Moore, A. W. (2000). X-means: Extending K-means with Efficient Estimation of the Number of Clusters. Paper presented at the Proceedings of the Seventeenth International Conference on Machine Learning.

- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, 41(4, Part 1), 1432-1462. doi: <http://dx.doi.org/10.1016/j.eswa.2013.08.042>
- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R. ve Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, 21(1), 135-146. doi: 10.1002/cae.20456
- Romero, C., López, M.-I., Luna, J.-M. ve Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68, 458-472. doi: <http://dx.doi.org/10.1016/j.compedu.2013.06.009>
- Romero, C. ve Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. *Systems, Man, and Cybernetics, Part C: Applications and Reviews*, IEEE Transactions on, 40(6), 601-618. doi: 10.1109/TSMCC.2010.2053532
- Romero, C. ve Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27. doi: 10.1002/widm.1075
- Romero, C., Ventura, S. ve García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384. doi: <http://dx.doi.org/10.1016/j.compedu.2007.05.016>
- Siemens, G. ve Baker, R. S. J. d. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. Paper presented at the Proceedings of the 2nd International Conference on Learning Analytics and Knowledge.
- Valdiviezo, P., Reátegui, R. ve Sarango, M. (2013). Student Behavior Patterns in a Virtual Learning Environment. Paper presented at the Eleventh LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI'2013), August 14 - 16, 2013 Cancun, Mexico.
- Witten, I. H., Frank, E. ve Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*: Morgan Kaufmann Publishers Inc.