

Büyük Veriden Öğrencilerin Öğretim İçerik Tercihlerinin Başarıya Etkisinin Belirlenmesine Yönelik Veritabanlarından Bilgi Keşfi Yöntemi: OULAD Veri Seti Örneği ¹

Can MEŞE ²

Başvuru Tarihi: 30.10.2022

Kabul Tarihi: 26.01.2023

Makale Türü: Araştırma Makalesi

Öz

Sanal öğrenme ortamlarında öğrencilerin öğrenme içerikleri ile etkileşimlerinden dolayı çok sayıda veri ortaya çıkmaktadır. Bu veriler eğitsel veri madenciliği algoritmaları ile sanal öğrenme ortamlarının tasarımında yol gösterici nitelikte olmaktadır. Ancak öğrencilerin sanal öğrenme ortamında etkileşimde buldukları öğretim materyallerinin başarı durumlarına etkisinin incelenmesine ihtiyaç duyulmaktadır. Bu tür ihtiyaçlara yönelik yapılan eğitsel veri madenciliği uygulamalarına yönelik izlenmesi gereken bilimsel süreçlerin de ortaya konulması gerekmektedir. Bu çalışmada öğrencilerin sanal öğrenme ortamında farklı öğrenme materyalleri ile etkileşimlerinin başarı durumuna etkisinin ortaya konulmasında veritabanından bilgi keşfi yönteminin kullanılması amaçlanmıştır. Çalışmada büyük veri olarak The Open University Öğrenme Analitiği Veri Seti kullanılmıştır. Bu veri seti ile veritabanından bilgi keşfi yöntemine göre süreç izlenmiş ve karar ağacı algoritmalarından CART algoritması ile araştırma sorularına yanıt aranmıştır. Çalışmanın bulgularına göre öğrencilerin öğretim materyalleri ile etkileşimlerinin başarının bir belirleyicisi olduğu görülmüştür. Bu doğrultuda öğretim tasarımcılarının uzaktan eğitim ortamlarında standartların oluşturulmasına ve etkili bir öğretim tasarımı için tercih edilecek öğretim materyallerinin belirlenmesine yönelik alacakları kararda araştırma bulgularının yol gösterici nitelikte olması öngörülmektedir.

Anahtar Kelimeler: Veritabanlarından Bilgi Keşfi, Öğrenci-İçerik Etkileşimi, Öğretim Materyali, CART Analizi, OULAD Veri Seti

Atıf: Meşe, C. (2023). Büyük veriden öğrencilerin öğretim içerik tercihlerinin başarıya etkisinin belirlenmesine yönelik veritabanlarından bilgi keşfi yöntemi: OULAD veri seti örneği. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 23(1), 121-138.

¹ Bu çalışma etik kurul izin belgesi gerektirmemektedir.

² Kahramanmaraş İstiklal Üniversitesi, Mühendislik, Mimarlık ve Tasarım Fakültesi, Dijital Oyun Tasarımı Bölümü, canmese@gmail.com, ORCID: 0000-0002-9759-7055

Knowledge Discovery in Databases Method for Determining the Effect of Students' Teaching Content Preferences on Success from Big Data: OULAD Data Set Example

Can MEŞE³

Submitted by: 30.10.2022

Accepted by: 26.01.2023

Article Type: Research Article

Abstract

In virtual learning environments, much data is generated due to the interaction of students with learning content. These data are effective in the design of educational virtual learning environments using data mining algorithms. However, there is a need to examine the effect of teaching materials that students interact with from virtual learning environments on their success. Revealing the scientific processes should be clarified to follow for educational data mining applications made for such conducting educational data mining. This research aims to use the knowledge discovery in database method to reveal the effect of students' interactions with different learning materials in the virtual learning environment on their success. The Open University Learning Analytics Dataset (OULAD) as a big data was used in the study. With this dataset, answers to the research questions were sought with the CART algorithm, one of the decision tree algorithms. According to the findings of the study, it was seen that the interaction of the students with the teaching materials was a determinant of success. In this direction, it is foreseen that the research findings will be guiding in the decision of instructional designers to establish standards in distance education environments and to determine the teaching materials to be preferred for an effective instructional design.

Keywords: Knowledge Discovery In Databases, Student-Content Interaction, Teaching Material, CART Analysis, OULAD Dataset

³ Kahramanmaraş İstiklal University Faculty of Engineering, Architecture and Design, Department of Digital Game Design, canmese@gmail.com, ORCID: 0000-0002-9759-7055

Giriş

Öğrenciler sanal öğrenme ortamları ile etkileşimleri sayesinde sanal öğrenme ortamlarındaki davranışlarını yansıtan birçok veri bırakırlar. Bu veriler, öğretim içeriklerinin hazırlanmasından öğrenme ortamının tasarlanmasına kadar sanal eğitim hizmeti veren eğitim kurumların kararlarında etkili olmaktadır. Bu noktada elde edilen verilerin analizi büyük önem kazanmaktadır. Verilerin analizinde veri madenciliği teknikleri işe koşulmaktadır. Veri madenciliği teknikleriyle elde edilen bulgular, kurumsal ve öğretimsel kararların alınmasında etkili olmaktadır (Adak ve Yurtay, 2013).

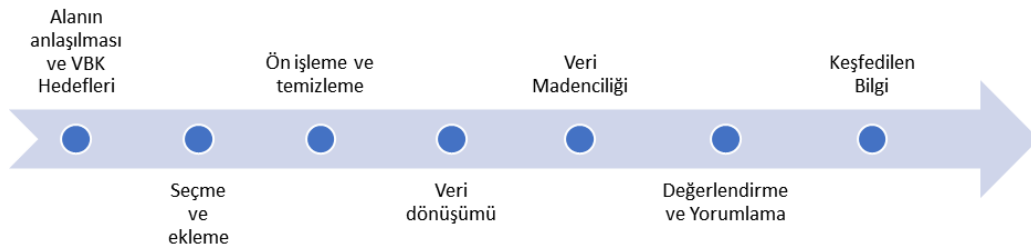
Öğrencilerin sanal öğrenme ortamlarındaki davranışları sayesinde sanal öğrenme ortamları, öğrencilerin çevrimiçi deneyimlerinin izlenmesine olanak sağlamaktadır (Zacharis, 2018). Öğrenciler sanal öğrenme ortamlarında öğrenme süreçlerine ilişkin gezinme davranışları, okuma ve yazma alışkanlıkları gibi birçok veri üretirler (Siemens, 2013). Bu verilerden yola çıkarak öğrencilerin davranışları hakkında çıkarımda bulunulabilir. Yapılan araştırmalara göre öğrencilerin öğrenme davranışlarının tespiti, tanımlanması ve modellenmesinde eğitsel veri madenciliği işe koşulmaktadır (Papamitsiou ve Economides, 2014). Eğitsel veri madenciliği (EDM) sayesinde öğrencilerin e-öğrenme ortamlarındaki etkileşimleri sonucuyla bıraktıkları eğitim verilerine veri madenciliği tekniklerinin uygulanmasına olanak sağlanmaktadır (Calders ve Pechenizkiy, 2012). EDM ile sınıflandırma, tahmine dayalı modelleme, kümeleme, ikili kümeleme, sık örüntü madenciliği, işbirlikçi filtreleme ve öneriler, görsel analitik gibi birçok veri madenciliği yöntemi sanal öğrenme ortamlarından elde edilen eğitim verilerine uyarlanabilmektedir (Calders ve Pechenizkiy, 2012).

Öğrenme analitikleri sayesinde istatistiksel analiz, sinir ağları vb. yöntemlerle veri tabanında keşifler yapılır ve öğrenci davranışlarına ilişkin öngörü kazanılır (Siemens, 2013). Öğrenme analitikleri, öğrencilerin sanal öğrenme ortamlarındaki ihtiyaçlarını belirlemek ve bu ihtiyaçlarına daha iyi çözümler üretmek için yeni yöntemler ve araçlar sunmaktadır (Greller ve Drachsler, 2012). Öğrencilerin sanal öğrenme ortamındaki davranışları ile öğrenme performansı arasındaki ilişkiyi araştıran çalışmalar sayesinde öğrencilerin sanal öğrenme ortamındaki davranışlarının öğrenme performansı üzerinde önemli bir etkisi bulunmaktadır (Qiu ve diğerleri, 2022). Bir diğer ifadeyle öğrencilerin derste sunulan öğrenme materyallerine tıklama sayıları ya da sanal öğrenme ortamlarında gerçekleştirdikleri etkinlik sayıları dersi bırakma oranı gibi başarı durumlarını ölçmeye yardımcı olabilmektedir (Feldman-Maggor ve diğerleri, 2021). Bu doğrultuda öğrencilerin sanal öğrenme ortamları ile etkileşimlerinden yola çıkarak başarılarının incelenmesi önemli görülmektedir. Bu noktada öğrencilerin sanal öğrenme ortamlarında bırakmış oldukları büyük verilerin (big data) analiz edilmesine ihtiyaç duyulmaktadır. Bu analiz sürecinde eğitsel veri madenciliği, öğrenme analitikleri gibi süreçler işe koşulmaktadır. Ancak bu analizlerin bilimsel araştırma süreci temelinde ele alınması için veritabanından bilgi keşfi (Knowledge Discovery in Databases) yöntemi gibi yöntemlerin takip edilmesi gerekmektedir. Bu araştırmada eğitsel verilerde veritabanından bilgi keşfi yönteminin kullanımına ilişkin açıklamalara yer verilmiştir.

Veritabanından Bilgi Keşfi (Knowledge Discovery in Databases) Yöntemi

Kurumların veriye dayalı karar almalarında büyük ölçekli veriler analiz edilerek karar destek sistemleri oluşturulabilir (Özkan, 2013). Bu doğrultuda kurumlar, bireylerin davranışlarından elde edilen verilerin analiz edilmesine ilişkin bir bakış açısına sahiptirler. Bu sayede kurumlar bireylerin davranışlarını anlayacak ve bireylere uygun bir şekilde hizmet sunacaklardır. Sanal öğrenme ortamlarında da öğrencilerin davranışlarının analiz edilip belirli bir karar alınmasına yönelik çalışmalar yapılmaktadır. Bilgilerin veritabanlarına kaydedildiği düşüldüğünde bu veritabanındaki verilerden bilgi yapılarının keşfedilmesine ihtiyaç duyulmaktadır. Bu noktada veritabanından bilgi keşfinin önemi gündeme gelmektedir.

Çoğu araştırmacı veri madenciliği ile bilgi keşfi kavramlarını benzer anlamda kullanmalarına rağmen aslında veri madenciliği, veritabanından bilgi keşfi sürecinin bir aşamasında kullanılmaktadır (dos Santos ve diğerleri, 2021; Fayyad, 2001; Özdemir ve diğerleri, 2010; Rokach ve Maimon, 2015). Bu doğrultuda veri madenciliğine yönelik yapılacak çalışmalarda veritabanından bilgi keşfi sürecine göre metodolojik sürecin izlenmesi önem kazanmaktadır. Veritabanından bilgi keşfi, verilerin hazırlanması, verilerin seçilmesi, verilerin temizlenmesi, ön bilgilerin dikkate alınması ve veri madencilik sonuçlarının yorumlanması gibi süreçlerin işe koşulması ile elde edilen verilerden faydalı bilgilere dönüştürülmesi sağlanır (Fayyad ve diğerleri, 1996). Dolayısıyla veritabanından bilgi keşfinde belirli bir sistematik süreç takip edilir. Rokach ve Maimon (2015) tarafından önerilen veritabanından bilgi keşfi süreci (Şekil 1), hedeflerin belirlenmesi ile başlar ve keşfedilen bilginin uygulanmasıyla sona erer.



Şekil 1. Veritabanından Bilgi Keşfi Süreci (Rokach ve Maimon, 2015)

Şekil 1'de görüleceği üzere alanın anlaşılması ve veritabanından bilgi keşfi hedeflerinin belirlenmesi veritabanından bilgi keşfi (VBK) sürecinin ilk aşamasını oluşturur. Bu aşamada veri madenciliği sürecinde araştırmacılar, son kullanıcının hedeflerini ve bilgi keşif sürecinin gerçekleşeceği ortamı anlar ve tanımlamalar yaparlar (Rokach ve Maimon, 2015). Seçme ve ekleme aşamasında; hedefler doğrultusunda belirlenen veri setinin oluşturulması ve bu veri setinin üzerinde bilgi keşfinin gerçekleştirileceği değişkenler belirlenir (Azevedo ve Santos, 2008). Dolayısıyla veritabanında hangi verilerin bulunduğu, ek verilerin gerekli olup olmadığı ve bilgi keşfi için gerekli olan değişkenlerin belirlendiği bu aşamada analiz yapılacak veri seti tek bir veri kümesi halinde hazırlanır (Rokach ve Maimon, 2015; Uluyardımcı ve Zontul, 2019). Eğitsel verilerin uygun bir şekilde ön işleme tabi tutulması, araştırmaların bulgularındaki büyük yanlışlıkları önlenmesinin yanı sıra sonuçların gerçekliğini ve güvenilirliğini önemli ölçüde artırabilir (Feldman-Maggor ve diğerleri, 2021). Bu nedenle ön işleme ve temizleme aşamasında kayıp verilerin ele alınması ve gürültülü veya aykırı değerlerin kaldırılması gibi veri temizleme işlemlerinin yanı sıra daha kompleks veri madenciliği teknikleri kullanılarak veriler ön işleme tabi tutulur (Özdemir ve diğerleri, 2010). Veri dönüşümü aşamasında ise daha iyi bir veri madenciliği yapılması için veriler hazırlanır ve geliştirilir (Han ve Kamber, 2012; Rokach ve Maimon, 2015). Ayrıca bu aşamada boyut azaltma veya dönüştürme yöntemleri tercih edilebilir (Azevedo ve Santos, 2008). Bir diğer ifadeyle çok boyutlu olan veri seti anlamlı olarak daha az boyutlarla ifade edilebilir. Ayrıca veri madenciliğinin uygulanması için gerekli olan veri dönüştürme işlemleri bu aşamada gerçekleştirilir. Araştırmanın bir diğer aşaması olan veri madenciliği aşamasında ise ilgili veri seti üzerinde gerçekleştirilecek veri madenciliği görevlerinden tahminleme ya da açıklama görevinin hangisinin tercih edileceği belirlenir, uygun veri madenciliği algoritması seçilir ve veri setine bu algoritma uygulanır (Fayyad, 2001; Rokach ve Maimon, 2015). Değerlendirme ve yorumlama aşamasında, veri madenciliği algoritmalarının sonuçları doğrultusunda elde edilen yapılar değerlendirilir ve yorumlanır (Azevedo ve Santos, 2008). VBK sürecinin son aşamasında veri madenciliği algoritmasının uygulanmasının sonucunda elde edilen bilginin kullanılması ve bu bilgiden yola çıkarak kararların alınması ve uygulanması gerçekleştirilir (Rokach ve Maimon, 2015). Böylelikle VBK ile elde edilen bilginin farklı sistemlere uygulanabilirliği sağlanır.

VBK sürecinin ilk aşamasında alanın anlaşılması ve veritabanından bilgi keşfi (VBK) hedeflerinin belirlenmesi yer almaktadır. Bu doğrultuda araştırmada veritabanından bilgi keşfi yöntemi kullanılarak büyük veriden öğrencilerin öğretim içerik tercihlerinin başarıya etkisinin belirlenmesi amaçlanmıştır. Bu amaca yönelik araştırma sürecinde VBK yöntemi tercih edilmiştir. VBK yönteminin ilk aşamasında alanın anlaşılması ve veritabanından bilgi keşfi hedeflerinin belirlenmesine yönelik açıklamalar araştırmanın önemi doğrultusunda aşağıda açıklanmıştır.

Dünyanın farklı noktalarında araştırmacıların eğitsel veri madenciliği veya öğrenme analitikleri temelinde öğrencilerin dersi bırakmasına yönelik çalışmalar yürüttüğü görülmektedir (de Andrade ve diğerleri, 2021). Buna ek olarak öğrencilerin dersten başarısızlık durumunu önceden tespit eden çalışmalar da yapılmaktadır (Aljohani ve diğerleri, 2019; Costa ve diğerleri, 2017). Bunun yanı sıra sanal öğrenme ortamlarında öğrenci davranışlarını inceleyen çalışmalar da bulunmaktadır. Sarıtaş ve diğerleri (2022) tarafından yapılan çalışmada bir devlet üniversitesinde öğrenci ve öğretim elemanlarının sistem raporları ve kullanıcıların uzaktan eğitimdeki davranış ve alışkanlıkları görsellerle sunulmuştur. Ayrıca Nkomo ve Nat (2021) tarafından yapılan çalışmada, SimpleKmeans kümeleme tekniği ile öğrencilerin harmanlanmış öğrenme ortamından elde edilen öğrenme verilerinden yola çıkarak öğrencilerin derslere katılımlarının belirlenmesi amaçlanmış ve öğrencilerin düşük, orta ve yüksek katılım seviyelerine dayalı olarak öğrenme kaynaklarıyla farklı şekilde etkileşime geçtiği bulunmuştur. Bu noktada öğrencilerin öğrenme yönetim sistemi üzerinden sunulan öğrenme içerikleri ile etkileşimlerine yönelik davranışlarını inceleyen çalışmalar yürütülmektedir (Akçapınar ve Hasnine, 2022; Xing, 2015). Dolayısıyla öğrencilerin öğrenme yönetim sistemi gibi sanal öğrenme ortamları ile etkileşiminin keşfedilmesi yararlı görülmektedir (Agudo-Peregrina ve diğerleri, 2014; Riestra-González ve diğerleri, 2021). Qiu ve diğerleri (2022) tarafından yapılan çalışmada öğrencilerin e-öğrenme davranış kategorileri e-öğrenme performans tahminini, makine öğrenmesi algoritmalarından SVC (R), SVC (L), Naïve Bayes, KNN (U), KNN (D) ve softmax ile incelenmiştir. Dolayısıyla yapılan araştırmalarda öğrenci başarısının tahminlenmesi ve öğrenci davranışlarının incelenmesine yönelik çalışmaların alanyazında yoğunlaştığı söylenebilir. Öğrencilerin sanal öğrenme ortamlarındaki etkileşimlerinin öğrencilerin başarı durumlarına yansımalarının incelenmesi ise önemli görülmektedir. Bu doğrultuda öğrencilerin sanal öğrenme ortamlarında öğrenme materyalleriyle etkileşimlerinden yola çıkarak öğrencilerin öğrenme materyalleri tercihlerinin başarı durumlarında nasıl etkili olduğunun incelenmesine ihtiyaç duyulmaktadır. Dolayısıyla öğrencilerin sanal öğrenme ortamında farklı öğrenme materyalleri ile etkileşimlerinin başarı durumuna etkisinin ortaya konulması amaçlandığı bu araştırmada aşağıdaki araştırma sorularına yanıt aranmıştır:

- Öğrencilerin sanal öğrenme ortamları ile etkileşimleri doğrultusunda başarılarında etkili olan sanal öğrenme içerikleri nelerdir?
- Öğrencilerin sanal öğrenme ortamları ile etkileşimlerinin başarı durumlarına etkisi nasıldır?

Araştırma sorularının yanıtlanması sayesinde öğrencilerin sanal öğrenme ortamları ile etkileşimlerinin başarı durumlarına yansımaları keşfedilmiştir. Böylelikle araştırma kapsamında öğrencilerin başarı durumlarında etkili olan sanal öğrenme ortamındaki öğretim materyalleri belirlenmiştir. Araştırmanın bulgularının uzaktan öğretim tasarımcılarına uzaktan eğitim ortamlarında standartların oluşturulmasında, etkili bir öğretim tasarımı için öğretim materyallerinin belirlenmesinde ve uzaktan öğretim tasarımına ilişkin alınacak kararlarda yol göstermesi öngörülmektedir.

Yöntem

Büyük veriden öğrencilerin öğretim içerik tercihlerinin başarıya etkisinin belirlenmesine yönelik gerçekleştirilen bu çalışmada nicel araştırma yöntemleri arasında yer alan ilişkisel araştırma yöntemi işe koşulmuştur. İlişkisel araştırmalar, iki veya daha fazla nicel değişkenin ilişkili olma derecesi temelinde insan davranışlarını açıklamaya yardımcı olmak ya da olası sonuçları tahmin etmek için kullanılır (Fraenkel ve diğerleri, 2012). Bunun yanı sıra iki veya daha fazla değişkenin ilişkilendirilmesinde birbirini etkileyip etkilemediğini belirlemek için ilişkisel araştırma yöntemi tercih edilir (Creswell, 2012). Bu doğrultuda araştırma VBK süreci temelinde yürütülmüştür. VBK, ham verilerden bilgi kalıpları çıkarmak için özel algoritmaların uygulanması olarak ele alınmaktadır (Fayyad ve diğerleri, 1996). Veri madenciliği yöntemi, eldeki sorunlara geleneksel istatistiksel yöntemlerin yerine gelişmiş teknik ve algoritmalar kullanarak yaklaşımlar oluşturmak olarak ifade edilebilir (Eygü ve Bayhan, 2020). Veri madenciliği ile veritabanlarından bilgi keşfi eş anlamlı olarak kabul edilse de veri madenciliği, veritabanlarından bilgi keşfi sürecinin bir aşaması olarak ele alınmaktadır (Han ve Kamber, 2012; Sachin ve Vijay, 2012). Bu doğrultuda araştırmanın süreci Şekil 1’de gösterildiği gibi veritabanlarından bilgi keşfi süreci (Rokach ve Maimon, 2015) temel alınarak eğitsel veri madenciliği süreci doğrultusunda yürütülmüştür. VBK sürecinin ilk aşamasını oluşturan alanın anlaşılması ve VBK hedeflerinin belirlenmesi araştırmanın giriş kısmında açıklanmış ve araştırma soruları ile ifade edilmiştir. Bu doğrultuda araştırma sürecinde işe koşulan diğer VBK aşamaları aşağıda açıklanmıştır.

Verileri Seçme ve Ekleme

Eğitsel verilerin kamuya açık hale getirilmesi dataTEL challenge, dataTEL workshop, PSLC dataShop, LinkedEducation.org gibi girişimlerle 2010 yılından itibaren paylaşılmaktadır (Greller ve Drachler, 2012). Açık veri paylaşımlarından biri de The Open University tarafından gerçekleştirilmiştir. The Open University Öğrenme Analitiği Veri Seti (The Open University Learning Analytics Dataset- OULAD), öğrencilerin demografik verilerini ve sanal öğrenme ortamındaki etkileşim verilerini içermesinin yanı sıra e-öğrenme veri çeşitliliği açısından en kapsamlı uluslararası açık veri setlerinden biridir (Qiu ve diğerleri, 2022). Buna ek olarak The Open University, öğrenme analitiklerine ilişkin en çok yayın üreten kurum olarak görülmektedir (Azevedo ve Azevedo, 2021). Kuzilek ve diğerleri (2017) tarafından bu üniversitede (The Open University-OU) sunulan derslerin verilerini içeren bir veri seti hazırlanmıştır. Bu veri seti, öğrencilerin demografik bilgilerini ve sanal öğrenme ortamında (Virtual Learning Environment- VLE) etkileşimlerini temsil eden öğretim materyallerine tıklama sayılarını içermektedir. Veri setinde 2013-2014 öğretim yıllarında toplam 22 derse kayıtlı 32.593 öğrencinin başarı değerlendirme sonuçları ve sanal öğrenme ortamında öğretim materyallerine günlük tıklama sayıları (10.655.280 tıklama sayısı) bulunmaktadır (Kuzilek ve diğerleri, 2017). Veri setinde öğrencilerin demografik özelliklerine (cinsiyet, eğitim seviyesi, yaş, öğrencinin yaşadığı bölge, engel durumu vb.), performansına (öğrencilerin çalışmaları sırasında elde ettikleri sonuçları ve başarıları) ve öğrenme davranışına (öğrenci sanal öğrenme ortamındaki etkinliklerinin log kayıtları) ilişkin veriler ilişkisel tablo yapısında sunulmuştur. Bu veri setinde öğrencilere ait bilgiler gizli tutulmuş ve etik açıdan gerekli önlemler alınmıştır (Kuzilek ve diğerleri, 2017). Bu araştırma, kapsamı gereği etik kurul onayı gerektirmemektedir. Dolayısıyla bu çalışmada araştırma sorularını yanıtlamak için The Open University Öğrenme Analitiği Veri Seti araştırmanın verileri olarak tercih edilmiştir.

Ön işleme ve Temizleme

OULAD veri seti ilişkisel veri tabanı üzerine kurgulanmıştır. Veriler assessments.csv, courses.csv, studentAssessment.csv, studentInfo.csv, studentRegistration.csv, studentVle.csv ve vle.csv veri dosyalarında tutulmuştur. Araştırma kapsamında verilerin ön işlenmesinde RapidMiner yazılımı kullanılmış olup OULAD veri setindeki dosyalar bu yazılıma aktarılmıştır. Daha sonra verilerin anlamlı bir bütünlük oluşturması adına

bu veri setindeki tabloların birleştirilmesi (join) yapılmıştır. Öğrencilerin demografik bilgileri studentInfo.csv dosyasında tutulurken sanal öğrenme ortamındaki etkileşimleri studentVle.csv ve vle.csv dosyasında tutulmaktadır. Dolayısıyla join yöntemi ile veri dosyalarındaki veriler eşleştirilmiştir. Böylelikle öğrencilere ait verilerle öğrencilerin öğretim materyalleriyle etkileşim verileri bir araya getirilmiştir.

Veri Dönüşümü

Araştırmanın bu aşamasında öğrencilerin sanal öğrenme ortamı ile etkileşimleri ve demografik bilgileri RapidMiner yazılımında Pivot Operatörü kullanılarak hazırlanmıştır. Bu işlemde veriler öğrenci bilgileri ve öğretim materyali açısından gruplar halinde yeniden düzenlenmiştir. Her grup için öğrencilerin ilgili içerikteki toplam tıklama sayıları hesaplanmıştır. Böylelikle daha büyük bir tablodaki verileri özetleyen bir pivot tablo oluşturulmuştur. Pivot tablo dönüşümünden sonra öğrencilerin tıklama yapmadıkları içerikler kayıp değer olarak belirlenmiş ve bu tıklama yapılmayan kayıp değerler sıfır (0) ile doldurulmuştur. Bu sayede veri madenciliği yönteminin uygulanması için veri seti hazır hale getirilmiştir. Bunun yanı sıra başarı durumu yeniden kodlanmıştır. Başarı durumu “withdraw” ve “fail” olanlar sıfır (0) yani başarısız; başarı durumu “pass” ve “distinction” olanlar bir (1) yani başarılı olarak yeniden kodlanmıştır. Başarı durumu böylelikle başarılı ve başarısız olmak üzere binary (ikili) şeklinde kodlanmıştır.

Veri Madenciliği Yönteminin Uygulanması

Sınıflama doğruluğu ve anlaşılır özet bir sonuç gösterimine sahip olması gibi üstünlüklerden dolayı karar ağaçları yaygın bir kullanım alanına sahiptir (Keskin ve Yurdugül, 2020). Karar ağacı, bir veri kümesinin sınıflandırılması için kuralları içeren ve karar kümelerini temsil eden ağaç şeklindeki yapılardır (Baradwaj ve Pal, 2011). Yordayan değişkenlerin kategorik ve bağımlı değişkenlerin sürekli olduğu durumlarda karar ağacı algoritmaları arasından Classification and Regression Trees (CART) algoritmasının regresyon modellerine kıyasla daha iyi tahmin sağladığı görülmektedir (Razi ve Athappilly, 2005). Bu nedenle araştırmada öğrencilerin öğrenme materyalleriyle etkileşimlerinin başarı durumuna etkisinin incelenmesinde karar ağaçları algoritmalarından CART algoritması tercih edilmiştir. Bu analizin yürütülmesinde IBM SPSS 26 programı işe koşulmuştur.

Değerlendirme ve Yorumlama ve Bilginin Keşfi

VBK aşamalarının değerlendirme ve yorumuna ve bilginin keşfine ilişkin açıklamalara araştırmanın bulgular ve tartışma kısmında yer verilmiştir.

Bulgular

Araştırmada dataplus, dualpane, externalquiz, folder, forumng, glossary, homepage, htmlactivity, oucollaborate, oucontent, ouilluminate, ouwiki, page, questionnaire, quiz, repeatactivity, resource, sharedsubpage, subpage ve url öğretim materyallerine tıklama sayıları bağımsız değişken olarak CART analizine dahil edilmiştir. Çapraz doğrulama yöntemiyle (Cross Validation) en fazla beş ağaç derinliğinde yapılan analiz bulgularında homepage, forumng, subpage, resource, url, oucontent, quiz, ouwiki, oucollaborate, page, questionnaire, glossary, dualpane, folder, sharedsubpage, dataplus, externalquiz, ouilluminate, htmlactivity ve repeatactivity öğrenme materyallerine tıklama sayıları bağımsız değişken olarak modelde ortaya çıkmıştır. Tablo 1’de CART analizinin doğruluk bulgularına yer verilmiştir.

Tablo 1
CART Analizinin Doğruluk Bulguları

		Tahmin Edilen		
		Başarısız	Başarılı	Doğruluk Yüzdesi
Gözlenen	Başarısız	13691	3517	0.796
	Başarılı	2127	13258	0.862
	Genel Yüzde	0.485	0.515	0.827

Tablo 1'e göre CART analizinin %82,7 oranla doğru bir şekilde tahmin ettiği belirlenmiştir. Bu modelin tahminine ilişkin başarısız öğrencilerin %79,6'sı doğru tahmin edilirken başarıların %86,2'si doğru tahmin edilmiştir. Modelde beş ağaç derinliğinde 49 düğüm (node) elde edilmiştir. Öğrencilerin başarı durumlarına ilişkin karar ağacı modelinin görüntüsü okuyucu dostu olması açısından bazı düğümler (node) gizlenerek Şekil 2'de sunulmuştur.

Öğrencilerin sanal öğrenme ortamındaki öğretim materyalleri ile etkileşimlerinin incelendiği bu çalışmada OULAD veri setine CHAID analizi uygulanmıştır. Şekil 2’de yer alan analiz bulgusunun düğümlerindeki öğretim materyallerine ilişkin açıklamalar aşağıda sunulmuştur (Qiu ve diğerleri, 2022):

- Homepage - Öğrenme platformunun ana arayüzüne erişim
- Page- Dersin arayüzüne erişim
- Subpage- Dersin alt arayüzüne erişim
- Glossary- Sözlüğe erişim
- Ouwiki- Wikipedia ile sorgulama
- Resource- Arama platformu kaynakları
- URL- Dersin URL bağlantısına erişim
- Oucontent- Öğrenme platformundan öğretim kaynaklarını indirme
- Forumng- Dersin forumuna katılım
- Oucollaborate- İşbirliğine dayalı faaliyetlere katılım
- Ouilluminate- Simülasyon dersinin seminerlerine katılım
- Externalquiz- Müfredat dışı sınavları tamamlama

Araştırmanın analiz bulguları incelendiğinde öğrencilerin başarı durumlarında etkili olan en başat öğrenci etkinliğinin, öğrencilerin sanal öğrenme ortamının anasayfasına erişimleri olduğu görülmektedir. Bu anasayfa ile düşük ve yüksek etkileşime göre öğrencilerin başarı durumları değişmektedir. Anasayfaya 100’den daha az tıklama yapan 14640 öğrencinin %85,7’sinin başarısız olduğu görülmektedir. Bu öğrenciler arasında kısa sınav (quiz) etkinliği ile etkileşimi düşük olan öğrencilerin çoğunluğunun (%94,8) başarısız olduğu görülmüştür. Ancak kısa sınav etkinliğiyle etkileşim düzeyi yüksek olan öğrencilerin oranının başarısızlara göre düşük olmasına rağmen başarılı öğrenci oranının (%40,9) arttığı görülmektedir. Bu durum daha detaylı bir şekilde karar ağacı üzerinden incelendiğinde (Node 9) sanal öğrenme ortamından öğretim kaynaklarını indirme (oucontent) davranışını daha fazla gösteren öğrencilerden başarılı olanların oranı (%74,3) daha yüksektir. Ayrıca Node 21’de görüleceği üzere sanal öğrenme ortamından öğretim kaynaklarını indirme (oucontent) davranışını daha az gösteren öğrencilerin başarılı olmak için forum etkinliğini tercih ettikleri görülmektedir. Aslında bu durum, birçok öğretim materyali ile etkileşimi düşük olmasına rağmen forum etkinliğinde aktif davranış sergileyen öğrenciler arasında başarılı (%61,2) olan öğrencilerin daha fazla olduğu görülmektedir.

Öğrencilerden ilk olarak anasayfa (homepage) ile etkileşimi yüksek olup daha sonra dataplus etkinliği ile etkileşim kurup quiz etkinliğiyle de yoğun bir etkileşimde bulunanların çoğunluğunun (%97,5) başarılı olduğu keşfedilmiştir. Bu öğrencilerin doğrudan başarı odaklı olduğu söylenebilir. Bu durumu destekler nitelikte öğrenciler dataplus etkinliği ile düşük etkileşime geçen, page etkinliğiyle de yüksek etkileşim kuran ve kısa sınav (quiz) etkinliğiyle düşük etkileşime geçenlerin başarısız (%88,1), yüksek etkileşime geçenlerin başarılı olanlarının (%73,3) daha fazla olduğu görülmektedir. Node 5’te görüleceği üzere sanal öğrenme ortamının ana arayüzüne erişim düzeyi düşük öğrencilerden kısa sınav (quiz) etkinliğiyle yüksek düzeyde etkileşime geçenlerin (%70,7) daha başarılı olduğu görülmektedir. Ancak Node 11’de görüleceği üzere kısa sınav (quiz) etkinliğiyle düşük etkileşime geçenlerin başarısız olanların daha fazla (%55,9) olmasına rağmen dersin arayüzü (page) ile etkileşimi düşük ve işbirliğine katılım faaliyetleri (oucollaborate) ile etkileşimi yüksek olan öğrencilerden (Node 38) başarılı öğrencilerin daha fazla olduğu dikkat çekmektedir. Bu doğrultuda etkileşim oranı düşük olan öğrencilerin işbirliğine dayalı çalışmalarla etkileşim kurarak başarılı olma eğiliminde olduğu söylenebilir.

Tartışma

Sanal öğrenme ortamlarında öğrenciler sadece akranları veya öğretmenleri ile etkileşimde bulunmaz, öğrenme içerikleri ile etkileşimde bulunurlar. Moore (1989) tarafından ifade edilen öğrenci-içerik etkileşimi öğrencinin öğrenme içeriğiyle etkileşimde bulunmasını temsil etmektedir. Bu araştırma kapsamında veritabanlarından bilgi keşfi yöntemi kullanılarak büyük veriden öğrencilerin öğretim içerik tercihlerinin başarıya etkisinin belirlenmesi amaçlanmıştır. Veritabanlarından bilgi keşfi yöntemi öğrencilerin sanal öğrenme ortamlarında bıraktıkları ham verilerin bilgiye dönüştürülmesinde kullanılacak bir yöntemdir. Bu doğrultuda sanal öğrenme ortamlarından elde edilen verilere veri madenciliği algoritmaları uygulanarak keşfedilen bilgiler, uzaktan öğretim ortamlarının geliştirilmesinde alınacak kararlarda etkili olabilmektedir. Bu doğrultuda araştırmada veri tabanlarından bilgi keşfi yöntemiyle öğrencilerin öğrenme materyalleri ile etkileşimlerinin belirlenmesine yönelik karar ağaçları algoritmalarından CART algoritması kullanılmıştır.

Araştırmada öğrencilerin sanal öğrenme ortamları ile etkileşimlerinin başarı durumlarında etkili olduğu görülmüştür. Öğrenme materyalleri ile daha fazla etkileşimde bulunan öğrencilerin başarılı olma eğiliminde oldukları söylenebilir. Araştırmanın bu bulgusuyla tutarlı olarak Lee ve Recker (2021) tarafından yapılan çalışmada daha düşük final notlarına sahip öğrencilerin çevrimiçi konuşma/dinleme etkinliklerine daha az katıldıkları ve dersin içeriğine daha az katkıda bulunduğu görülmüştür. Riestra-González ve diğerleri (2021) tarafından yapılan çalışmada ise öğrencilerin öğrenme yönetim sisteminde gerçekleştirdiği eylemlerle ilgili günlük kayıtları (log), performansları üzerinde güçlü bir etki göstermektedir. Araştırmanın bulgusuna göre öğrencilerin başarılı olmalarında anasayfa (homepage) etkileşimlerinin önemli olduğu görülmektedir. Bu bulgu sanal öğrenme ortamı ile etkileşimde bulunanların başarılı olma eğiliminde olduğu düşüncesini ortaya çıkarmaktadır. Bu durum dersin başlangıcındaki dersleri görüntüleme eylemlerinin başarılı öğrencilerin tespit edilmesinde güçlü bir gösterge (Riestra-González ve diğerleri, 2021) olarak ele alınabilir. Dolayısıyla öğrencilerin sanal öğrenme ortamlarındaki etkileşimlerinin başarılı olmalarında etkili olduğu söylenebilir.

Uzaktan eğitimde öğrencilerin öğrenmesini kolaylaştırmak için öğretim elemanları çeşitli öğretim materyalleri sunmaktadırlar. Öğrenciler bu öğretim materyalleriyle etkileşimde bulunarak öğrenmelerini yapılandırır. Bunun sonucunda da dersten başarılı ya da başarısız olma eğilimleri ortaya çıkar. Riestra-González ve diğerleri (2021) tarafından yapılan çalışmada öğretim görevlileri tarafından yüklenen kaynakları az görüntüleme davranışı dersi bırakma riski altındaki ve başarısız öğrenciler üzerinde önemli bir etkiye sahiptir. Dolayısıyla öğrenme yönetim sisteminde bazı öğretim materyalleriyle düşük etkileşim gösteren öğrencilerin başarılı olmak için farklı öğretim materyalleri ile etkileşimde buldukları düşüncesi ortaya çıkmaktadır. Araştırmanın bulgusunda görüleceği üzere anasayfa ile etkileşimleri düşük olan öğrencilerin kısa sınavla etkileşimlerini yüksek tutarak başarılı olmaya çalışmışlardır. Bu durumda öğrencilerin doğrudan başarı elde etmelerinde etkili olan kısa sınav gibi öğretim materyallerine yöneldikleri söylenebilir.

Öğrencilerin akranlarıyla bilgi ve duygularını paylaşmaya daha istekli oldukları (He, 2013) göz önüne alınırsa öğrenciler arasında bilgi alışverişinde forumlar etkili bir araç olarak değerlendirilebilir. Araştırmada elde edilen bulgularda forum ile daha fazla etkileşimde bulunan öğrencilerin başarılı oldukları görülmektedir. Çünkü forum, içeriğin en fazla okunduğu ve dönem boyunca öğrencilerin aktif kaldığı bir bileşendir (Serradell-Lopez ve diğerleri, 2023). Bu doğrultuda öğrencilerin bilgi alışverişini gerçekleştirmesi ve öğrenci-öğretmen ile öğrenci-öğrenci etkileşiminin sağlanması için forumların uzaktan eğitimde etkili bir öğrenme materyali olduğu söylenebilir.

Sonuç olarak araştırmanın bulgularında da görüleceği üzere The Open University tarafından öğrencilere 20 farklı öğrenme içeriğinin sunulması, öğrencilerin başarılı olmalarında kendi öğrenme yollarını seçmelerine olanak sağlamaktadır. Bu doğrultuda öğrenciler dersten başarılı olmak için etkileşim kurdukları öğrenme materyalleri çeşitlenmektedir. Dolayısıyla öğrencilerin kendi öğrenme süreçlerini yapılandırmaları için çeşitli öğrenme materyallerinin sunulması başarılı olmalarında kendi öğrenme yollarını tercih etmelerini sağlayacaktır.

Sınırlılıklar ve Öneriler

Araştırma veritabanlarından bilgi keşfi yöntemi temelinde yürütülmüştür. Ancak SEMMA, CRISP-DM gibi diğer yöntemler ele alınmamıştır. Bu nedenle veri madenciliği uygulamalarında diğer yöntemlerin de araştırmacılar tarafından incelenmesi ve elde edilen ham verilerin doğasına uygun metodolojik sürecin tercih edilmesi önerilmektedir.

Bu araştırma, 2013-2014 eğitim-öğretim yıllarında The Open University derslerinden elde edilen OULAD veri seti ile sınırlıdır. Değişen zaman içerisinde öğrencilerin davranışlarının değişebileceği ve farklı kültürlerde ve kurumlarda öğrenci davranışlarının farklılık gösterebileceği göz önüne alınarak öğrencilerin sanal öğrenme ortamlarındaki davranışlarını inceleyen çalışmaların yürütülmesi önerilmektedir. Bunun yanı sıra araştırmada CART algoritması işe koşulmuştur. Yapılacak araştırmalarda farklı algoritmaların tercih edilmesi bu araştırmanın bir diğer önerisidir.

Araştırmada öğretim materyalleriyle öğrencilerin etkileşimlerinin başarı durumlarının etkisi incelendiği için OULAD veri setinde yer alan öğrencilerin 20 farklı öğrenme içerikleriyle etkileşimleri boyut indirgemesi (dimension reduction) yapılmadan analiz edilmiştir. Boyut indirgemenin araştırmanın bir sınırlılığı olarak ele alınırsa yapılacak araştırmalarda öğrencilerin öğrenme içerikleriyle etkileşimlerinde boyut indirgemesinin yapılması ve analizin bu şekilde uygulanması önerilmektedir.

Bir dersin verilerini öğrenme analitiği ile analiz etmek, öğretim tasarımı kararlarını etkileyebilecek öngörü sağlayabilir (Davies ve diğerleri, 2021). Bu açıdan uzaktan öğretimin planlanmasında ve tercih edilecek öğretim materyallerinin belirlenmesinde araştırmanın bulgularının temel alınmasının faydalı olacağı düşünülmektedir. Bu doğrultuda araştırmanın bulgularında öğrenci başarısının artırılmasına yönelik tercih edilebilecek öğrenme materyallerinin belirlenmesi ve buna yönelik öğrenme içeriklerinin düzenlenmesi uzaktan eğitim uygulayıcılarının karar almalarına katkı sağlayacak bir öneridir. Böylelikle uzaktan öğretim uygulayıcılarının kullandıkları öğrenme yönetim sistemi gibi sanal öğrenme ortamlarında toplanan verilerin analiz edilmesi ve analiz bulguları doğrultusunda alınan kararların gözden geçirilmesi bu araştırmanın bir diğer önerisidir.

Kaynakça

- Adak, M. F. ve Yurtay, N. (2013). Gini algoritmasını kullanarak karar ağacı oluşturmayı sağlayan bir yazılımın geliştirilmesi. *Bilişim Teknolojileri Dergisi*, 6(3), 1-6. Erişim adresi: <https://dergipark.org.tr/tr/pub/gazibtd/issue/6629/87997>
- Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á. ve Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, 31, 542-550. doi: 10.1016/j.chb.2013.05.031
- Akçapınar, G. ve Hasnine, M. N. (2022). Discovering the effects of learning analytics dashboard on students' behavioral patterns using differential sequence mining. *Procedia Computer Science*, 207, 3818-3825. doi: 10.1016/j.procs.2022.09.443
- Aljohani, N. R., Fayoumi, A. ve Hassan, S.-U. (2019). Predicting at-risk students using clickstream data in the virtual learning environment. *Sustainability*, 11(24), 7238. doi: 10.3390/su11247238
- Azevedo, A. ve Azevedo, J. M. (2021). Learning analytics: A bibliometric analysis of the literature over the last decade. *International Journal of Educational Research Open*, 2, 100084, 1-12. doi: 10.1016/j.ijedro.2021.100084
- Azevedo, A. ve Santos, M. F. (2008). *KDD, SEMMA and CRISP-DM: A parallel overview*. IADIS European Conference Data Mining, 182-185, MCCSIS 2008. Erişim adresi: <http://recipp.ipp.pt/handle/10400.22/136>.
- Baradwaj, B. K. ve Pal, S. (2011). Mining educational data to analyze students' performance. *International Journal of Advanced Computer Science and Applications*, 2(6). doi: 10.14569/IJACSA.2011.020609
- Calders, T. ve Pechenizkiy, M. (2012). Introduction to the special section on educational data mining. *ACM SIGKDD Explorations Newsletter*, 13(2), 3-6. doi: 10.1145/2207243.2207245
- Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F. ve Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73, 247-256. doi: 10.1016/j.chb.2017.01.047
- Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative* (4th Edition). NJ: Prentice Hall Upper Saddle River.
- Davies, R., Allen, G., Albrecht, C., Bakir, N. ve Ball, N. (2021). Using educational data mining to identify and analyze student learning strategies in an online flipped classroom. *Education Sciences*, 11(11), 668. doi: 10.3390/educsci11110668
- de Andrade, T. L., Rigo, S. J. ve Barbosa, J. L. V. (2021). Active methodology, educational data mining and learning analytics: A systematic mapping study. *Informatics in Education*, 20(2), 171-204. doi: 10.15388/infedu.2021.09

- dos Santos, R.A., Paulista, C.R. ve da Hora, H.R.M. (2021). Education data mining on PISA 2015 best ranked countries: What makes the students go well. *Tech Know Learn*. Çevrimiçi ön yayın. doi: 10.1007/s10758-021-09572-9
- Eygü, H. ve Bayhan, Y. C. (2020). *Karar ağaçları ile sınıflandırma CHAID algoritması teori ve bir uygulama*. Bursa: Ekin Yayınevi.
- Fayyad, U. (2001). Knowledge Discovery in Databases: An Overview. In: Džeroski, S., Lavrač, N. (eds) *Relational Data Mining*. Springer, Berlin, Heidelberg. doi: 10.1007/978-3-662-04599-2_2
- Fayyad, U., Piatetsky-Shapiro, G. ve Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazin*, 17(3), 37-54. doi: 10.1609/aimag.v17i3.1230
- Feldman-Maggor, Y., Barhoom, S., Blonder, R. ve Tuvi-Arad, I. (2021). Behind the scenes of educational data mining. *Education and Information Technologies*, 26(2), 1455-1470. doi: 10.1007/s10639-020-10309-x
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. (2012). *How to design and evaluate research in education*. McGraw-Hill.
- Greller, W. ve Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, 15(3), 42-57. Erişim adresi: <http://www.jstor.org/stable/jeductechsoci.15.3.42>
- Han, J. ve Kamber, M., (2012), *Data Mining Concepts and Techniques* (3rd edition), Waltham: Morgan Kauffmann Publishers Inc.
- He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 29(1), 90-102. doi: 10.1016/j.chb.2012.07.020
- Keskin, S. ve Yurdugül, H. (2020). *Kestirim yöntemleri*. Güyer, T., Yurdugül, H., Yıldırım, S. (Ed.) Eğitsel veri madenciliği ve öğrenme analitikleri (s. 66-88) içinde. Ankara: Anı Yayıncılık.
- Kuzilek, J., Hlosta, M. ve Zdrahal, Z. (2017). Open University learning analytics dataset. *Scientific Data*, 4(1), 170171. doi: 10.1038/sdata.2017.171
- Lee, J.-E. ve Recker, M. (2021). The effects of instructors' use of online discussions strategies on student participation and performance in university online introductory mathematics courses. *Computers & Education*, 162, 104084. doi: 10.1016/j.compedu.2020.104084
- Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education*, 3(2), 1-7. doi: 10.1080/08923648909526659
- Nkomo, L. M. ve Nat, M. (2021). Student engagement patterns in a blended learning environment: An educational data mining approach. *TechTrends*, 65(5), 808-817. doi: 10.1007/s11528-021-00638-0
- Özdemir, A., Yalçın Aslay, F. ve Çam, H. (2010). Veri tabanında bilgi keşfi süreci: Gümüşhane Devlet Hastanesi uygulaması. *Sosyal Ekonomik Araştırmalar Dergisi* 10(20), 347-366. Erişim adresi: <https://dergipark.org.tr/tr/pub/susead/issue/28415/302475>

- Özkan, Y. (2013). *Veri madenciliği yöntemleri*. İstanbul: Papatya Yayıncılık Eğitim.
- Papamitsiou, Z. ve Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49-64. Erişim adresi: <http://www.jstor.org/stable/jeductechsoci.17.4.49>
- Qiu, F., Zhang, G., Sheng, X., Jiang, L., Zhu, L., Xiang, Q., Jiang, B. ve Chen, P. (2022). Predicting students' performance in e-learning using learning process and behaviour data. *Scientific Reports*, 12(1), 453. doi: 10.1038/s41598-021-03867-8
- Razi, M. ve Athappilly, K. (2005). A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. *Expert Systems with Applications*, 29(1), 65-74. doi: 10.1016/j.eswa.2005.01.006
- Riestra-González, M., Paule-Ruíz, M. del P. ve Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*, 163, 104108. doi: 10.1016/j.compedu.2020.104108
- Rokach, L. ve Maimon O. (2014). Data mining with decision trees: Theory and applications. *Series in Machine Perception and Artificial Intelligence*, 81, 1-303. doi: 10.1142/9097
- Sachin, R. B. ve Vijay, M. S. (2012). A survey and future vision of data mining in educational field. *2012 Second International Conference on Advanced Computing & Communication Technologies*, 2012, 96-100. doi: 10.1109/ACCT.2012.14
- Sarıtaş, M. T., Börekci, C. ve Demirel, S. (2022). Quality assurance in distance education through data mining. *International Journal of Technology in Education and Science*, 6(3), 443-457. doi: 10.46328/ijtes.396
- Serradell-Lopez, E., Lara-Navarra, P. ve Martínez-Martínez, S. (2023). The pareto principle in virtual communities of learning. *Computers in Human Behavior*, 138, 107444. doi: 10.1016/j.chb.2022.107444
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57(10), 1380-1400. doi: 10.1177/0002764213498851
- Uluyardımcı, M. M. ve Zontul, M. (2019). Veri madenciliği yöntemleri ile uçuş biletleme analizi. *AURUM Mühendislik Sistemleri ve Mimarlık Dergisi*, 3(2). 153-168. Erişim adresi: <https://dergipark.org.tr/tr/pub/ajesa/issue/52409/529493>.
- Xing, W. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, 47, 168-181. doi: 10.1016/j.chb.2014.09.034
- Zacharis, N. Z. (2018). Classification and regression trees (CART) for predictive modeling in blended learning. *International Journal of Intelligent Systems and Applications*, 10(3), 1-9. doi: 10.5815/ijisa.2018.03.01

Extended Abstract

Purpose

Although most researchers use the concepts of data mining and knowledge discovery in databases, data mining is used at a stage of the knowledge discovery process from the database (dos Santos et al., 2021; Özdemir et al., 2010; Rokach and Maimon, 2015). It is essential to follow the methodological process according to the knowledge discovery process from the database in studies to be conducted on data mining. In this direction, a specific systematic procedure is followed in the knowledge discovery in databases methodology. This study aimed to determine the effect of students' teaching content preferences on success from big data by using the knowledge discovery in databases method.

Design and Methodology

In this study, which was carried out to determine the effect of students' teaching content preferences from big data using the knowledge discovery in databases method, the knowledge discovery in databases (KDD) method was employed. Developing an understanding of the application domain, which constitutes the first stage of this method, is explained in the introduction part of the research and expressed with research questions. In selecting and adding data phase, The Open University Learning Analytics Dataset - OULAD, one of the open data, constitutes the research data. In this dataset, students' demographic information is kept in the studentInfo.csv file, their interactions in the virtual learning environment are stored in the studentVle.csv file, and information on the interactions in the virtual learning environment is also stored in the vle.csv file. Therefore, using the RapidMiner program, the data were matched with the join method. In the data transformation phase of the research, the students' data were converted by getting the sum of the number of clicks on the teaching materials with the pivot operator. It is seen that among the decision tree algorithms, the Classification and Regression Trees (CART) algorithm provides better estimation compared to regression models in cases where the predictor variables are categorical and the dependent variables are continuous (Razi and Athappilly, 2005). For this reason, the CART algorithm, one of the decision tree algorithms, was preferred in the study to examine the effect of students' interactions with learning materials on their success. The research findings and discussion section explained the evaluation and using the discovered knowledge of the last phase of KDD.

Findings

Homepage, forumng, subpage, resource, url, oucontent, quiz, ouwiki, oucollaborate, page, questionnaire, glossary, dualpane, folder, sharedsubpage, dataplus, externalquiz, ouilluminate, the number of clicks on the htmlactivity and repeatactivity learning materials appeared in the model as an independent variable. It was determined that the CART analysis predicted correctly with a rate of 82.7%.

It is seen that the most crucial student activity that is effective in the success of the students is the access of the students to the homepage of the virtual learning environment. According to low and high interaction with this homepage, students' success status changes. It is seen that 85.7% of 14640 students whose interaction with the homepage were less than 100 clicks failed. Among these students, it was seen that many students (94.8%) who had low interaction with the quiz activity were unsuccessful. However, it is seen that the rate of successful students (40.9%) increased, although the level of interaction with the quiz activity was lower than unsuccessful students.

It was discovered that the majority (97.5%) of the students who first interacted with the homepage and then interacted with the dataplus activity and interacted intensively with the quiz activity were successful. It can be said that these students are directly success oriented. In support of this situation, it is seen that the students who had low interaction with the dataplus activity and high interaction with the page activity were unsuccessful (88.1%) with the quiz activity, and those who were highly engaged were more successful (73.3%). As can be seen in Node 5, it is seen that among the students who have low interaction with the access level of the main interface of the learning platform, those who interact with the quiz activity at a high level (70.7%) are more successful. However, as can be seen in Node 11, although those who had low interaction with the quiz activity were more unsuccessful (55.9%), students who had low interaction with the course interface (page) and high interaction with cooperation participation activities (oucollaborate) (Node 38) draw attention to the fact that there are more successful students than others. In this direction, it can be said that students with a low interaction rate tend to be successful by interacting with collaborative studies.

Research Limitations

Within the research scope, the knowledge discovery method in databases was preferred. However, other methods, such as SEMMA, CRISP-DM have not been used. For this reason, it is recommended to examine different methods in data mining applications by researchers and choose the methodological process suitable for the nature of the raw data obtained.

Implications

This research is limited to the OULAD dataset obtained from The Open University courses in the 2013-2014 academic years. Considering that students' behaviors may change over time and differ in different cultures and institutions, conducting studies examining students' behaviors in virtual learning environments is recommended. In addition, the CART algorithm was used in the research. Another suggestion of this research is that different algorithms should be preferred in future studies.

Originality/Value

Analyzing the data of a course with learning analytics provides insights that can affect instructional design decisions (Davies et al., 2021). In this respect, it is thought that it will be helpful to base the research findings on the planning of distance education and determining the preferred teaching materials. In this direction, selecting the learning materials that can be preferred to increase student success in the research findings and arranging the learning content for this is a suggestion that will contribute to the decision-making of distance education practitioners. Thus, another suggestion of this research is to analyze the data collected in virtual learning environments, such as the learning management system used by distance education practitioners, and review the decisions taken in line with the analysis findings.

Purpose

Researchers worldwide are working on educational data mining or learning analytics to understand students' dropping out (de Andrade et al., 2021). Some studies predetermine the failure of students from the course (Aljohani et al., 2019; Costa et al., 2017). In addition, there are also studies examining student behavior in virtual learning environments. Saritas et al. (2022), the system reports of students and instructors at a state university, and the behaviors and habits of users in distance education are presented in a visual and table. At this point, studies examine students' behavior toward interacting with the learning content delivered through the learning management system (Akçapınar and Hasnine, 2022; Xing, 2015). Therefore, it is considered beneficial to explore the interaction of students with virtual learning environments such as the learning

management system (Agudo-Peregrina et al., 2014; Riestra-González et al., 2021). Therefore, it can be said that studies on predicting student achievement and examining student behavior are concentrated in the literature. Looking at the reflection of students' interactions in virtual learning environments on students' success is considered necessary. In this direction, there is a need to examine how students' preferences for learning materials affect their success based on students' interactions with learning materials in virtual learning environments.

Arařtırmacı Katkısı: Can MEŐE (%100).