



T.C.
HATAY MUSTAFA KEMAL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**EĞİTSEL VERİ MADENCİLİĞİNDEKİ EĞİLİMLERİN ÇOK YÖNLÜ
İNCELENMESİ**

[NUR GİZEM AÇIL]

ENFORMATİK ANABİLİM DALI

YÜKSEK LİSANS TEZİ

HATAY
[ARALIK-2019]



T.C.

HATAY MUSTAFA KEMAL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

EĞİTSEL VERİ MADENCİLİĞİNDEKİ EĞİLİMLERİN ÇOK YÖNLÜ
İNCELENMESİ

NUR GİZEM AÇIL

ENFORMATİK ANABİLİM DALI

YÜKSEK LİSANS TEZİ

HATAY
ARALIK-2019

06/12/2019

TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını ve tez üzerinde Yükseköğretim Kurulu tarafından hiçbir değişiklik yapılamayacağı için tezin bilgisayar ekranında görüntülendiğinde asıl nüsha ile aynı olması sorumluluğunun tarafıma ait olduğunu beyan ederim.

Nur Gizem AÇIL

TEŞEKKÜR

Yüksek lisans öğrenimim boyunca ve tez çalışmamın her aşamasında bilgilerini ve deneyimlerini benimle paylaşan, önerileri ile çalışmamı destekleyen saygıdeğer danışman hocam Dr. Öğr. Üyesi Hamide ŞAHİNKAYASI'na ve Arş. Gör. Muhammet Mustafa TAT'a teşekkürü borç bilirim. Çalışma esnasında desteklerini esirgemeyen eşim sayın İlker AÇIL'a en içten dileklerimle teşekkür ederim.

Bu tez çalışmasını her zaman yanımda olan değerli annem Meral AK ve babam Bünyamin AK'a ithaf ediyorum.



ÖZET

EĞİTSEL VERİ MADENCİLİĞİNDEKİ EĞİLİMLERİN ÇOK YÖNLÜ İNCELENMESİ

Bu tez çalışmasının amacı, eğitsel veri madenciliği kullanan bilimsel araştırma makalelerinin belirli değişkenler açısından incelenerek eğitsel veri madenciliği alan eğilimlerini ortaya çıkarmaktır. Eğitsel veri madenciliği gelişmekte olan bir alan olduğu için bu alanla ilgili çalışmaların tarandığı alanyazın incelemeleri, alan eğilimlerinin ortaya çıkarılmasına katkı sağlamaktadır.

Çalışma nitel araştırma yöntemi ile tasarlanmıştır. Veri toplama yöntemi olarak doküman analizi kullanılmıştır. Bu çalışma kapsamında 1 Ocak 2013 ile 31 Aralık 2018 tarihleri arasında belirli veri tabanlarında Türkçe ve İngilizce dillerinde eğitsel veri madenciliği konusunda indekslenen alan incelemesi veya nitel olamayan 122 makale incelenmiştir. Veri toplama aracı olarak araştırmacı tarafından geliştirilen makale inceleme ölçütleri ile oluşturulan bir çizelge kullanılmıştır. Yayınlar ait demografik özellikler, araştırma konusu, örneklem özellikleri, veri toplama süreci, veri madenciliğine ilişkin kullanılan yöntemler ve araştırma sonuçları bilgileri düzenli ve sistematik bir şekilde kaydedilmiştir. Elde edilen veriler betimsel istatistik yöntemleriyle Microsoft Excel 2016 programında analiz edilmiştir. Bulgular şekil ve çizelge kullanılarak sunulmuştur.

Araştırma bulgularına göre çalışma sayısının son yıllarda artış eğiliminde olduğu ve Scimedirect ve Scopus veri tabanlarında erişim sağlanan yayın sayısının fazla olduğu görülmüştür. Çalışmaların yazar dağılımı analizinde 358 farklı yazarın eğitsel veri madenciliği konusunda yayın yaptığı sonucuna ulaşılmıştır. Çalışmaların en fazla öğrenci performansını tahmin etme ve geri bildirim sağlama konularında olduğu görülmüştür. Yapılan çalışmalarda en fazla tercih edilen çalışma grubu ise üniversite öğrencileri olmuştur. Geleneksel eğitim ortamları, veri toplamak için dijital öğrenme ortamlarına göre daha az tercih edilmiştir. Çalışmalarda veri toplama ortamı olarak en fazla e-öğrenme ve öğrenme yönetim sistemleri tercih edilmiş olup Moodle en fazla veri toplanan öğrenme yönetim sistemi olmuştur. Veri madenciliği modelleme yöntemi olarak daha çok tahmin edici modeller tercih edilmiştir. Sınıflandırma analizinin en fazla kullanılan tahmin edici model; kümeleme analizinin ise en fazla kullanılan tanımlayıcı model olduğu sonucuna ulaşılmıştır.

2019, 83 sayfa

Anahtar Kelimeler: Eğitsel veri madenciliği, doküman analizi, betimsel analiz.

ABSTRACT

MULTI-WAY INVESTIGATION OF TRENDS IN EDUCATIONAL DATA MINING

The aim of this thesis is to investigate the tendencies of the field of educational data mining by examining the studies using educational data mining in terms of certain variables. Since educational data mining is an emerging field, conducting literature reviews examining educational data mining studies contributes to the observation of field trends.

The study was designed with qualitative research method. Document analysis was used as data collection method. Within the scope of this study, 122 articles indexed on educational data mining in Turkish and English languages in certain databases between 1 January 2013 and 31 December 2018 were examined. Purely qualitative and literature review studies were excluded. A table including columns from a set of article review criteria developed by the researcher was used as data collection tool. The demographic characteristics of the publications, research topic, sample characteristics, data collection process, data mining methods and research results are recorded regularly and systematically. The obtained data were analyzed with descriptive statistical methods in Microsoft Excel 2016 program. The findings are presented using figures and tables.

According to the findings, it was seen that the number of studies has been increasing in recent years and the number of publications provided in Scimedirect and Scopus databases was high. In the analysis of author distribution of the studies, it was concluded that 358 different authors published educational data mining. It was seen that the most studies were about estimating student performance and providing feedback. The most preferred study group in the studies was university students. Traditional educational environments were less preferred than digital learning environments for data collection. In the studies, e-learning and learning management systems were mostly preferred to collect data, and Moodle was the most used learning management system. Prediction models were mostly preferred as data mining modeling method. It is concluded that classification analysis is the most used predictive model and cluster analysis is the most used descriptive model.

2019, 83 pages

Key Words: Educational data mining, document analysis, descriptive analysis.

İÇİNDEKİLER

TEŞEKKÜR.....	I
ÖZET	II
ABSTRACT.....	III
İÇİNDEKİLER	IV
ŞEKİLLER DİZİNİ.....	V
ÇİZELGELER DİZİNİ	VI
SİMGELER VE KISALTMALAR DİZİNİ	VII
1. GİRİŞ	1
1.1. Araştırmanın Problem Durumu	2
1.2. Araştırmanın Amacı.....	2
1.3. Araştırma Soruları	3
1.4. Araştırmanın Önemi	4
1.5. Sayıtlılar.....	4
1.6. Sınırlılıklar.....	5
1.7. Veri Madenciliği.....	5
1.7.1. Veri Madenciliği Modelleme Yöntemi	6
1.7.2. Veri Madenciliği Yöntemleri	7
1.8. Eğitsel Veri Madenciliği.....	10
1.8.1. Eğitsel Veri Madenciliği Süreci	12
1.8.2. Eğitsel Veri Madenciliğinin Görevleri ve Uygulamaları	13
1.8.3. Eğitsel Ortam Türleri	15
1.8.4. Eğitsel Veri Madenciliği Yöntemleri	16
2. ÖNCEKİ ÇALIŞMALAR.....	19
3. MATERYAL VE YÖNTEM	26
3.1. Araştırma Modeli.....	26
3.2. Veri Toplama	27
3.2.1. Makale İnceleme Ölçütleri ve Değerleri	27
3.3. Veri Analizi	29
4. ARAŞTIRMA BULGULARI VE TARTIŞMA	32
4.1. Birinci Araştırma Sorusu: Yayınların Demografik Özellikleri	32
4.2. İkinci Araştırma Sorusu: Çalışmaların Örneklem Özellikleri	52
4.3. Üçüncü Araştırma Sorusu: Veri Toplama Süreci Özellikleri.....	53
4.4. Dördüncü Araştırma Sorusu: Veri Madenciliğine İlişkin Kullanılan Yöntemlerin Özellikleri.....	57
4.5. Beşinci Araştırma Sorusu: Çalışmaların Sonuç Dağılımı	61
5. SONUÇ VE ÖNERİLER	63
5.1. Sonuç	63
5.2. Öneriler.....	65
KAYNAKLAR	67
ÖZGEÇMİŞ	73
EK 1: Çalışma Kapsamında İncelenen Yayınlar.....	74

ŞEKİLLER DİZİNİ

Şekil 1.1. Veri madenciliği modelleri ve işlevleri (Dunham, 2003)	6
Şekil 1.2. Veri madenciliği modelleme yöntemleri (Akpınar, 2000).....	7
Şekil 1.3. Disipliner alan EVM (Romero ve Ventura, 2013).....	10
Şekil 1.4. Eğitsel bilgi keşfi ve veri madenciliği süreci (Romero ve Ventura, 2013).....	12
Şekil 1.5. Eğitim ortamları (Romero ve Ventura, 2013).....	15
Şekil 4.1. Çalışmaların yayın yıl dağılımı.....	32
Şekil 4.2. Çalışmaların veri tabanı dağılımı.....	46
Şekil 4.3. Çalışmaların araştırma konusu dağılımı	51
Şekil 4.4. Çalışmaların örneklem grubu dağılımı	52
Şekil 4.5. Çalışmaların çalışma grubu boyutu	53
Şekil 4.6. Çalışmaların veri toplama ortam dağılımı	54
Şekil 4.7. Çalışmaların bilgisayar tabanlı ortam türü dağılımı	55
Şekil 4.8. Çalışmaların veri toplama aracı dağılımı.....	56
Şekil 4.9. Çalışmaların veri madenciliği modelleme yöntemi dağılımı.....	57
Şekil 4.10. Çalışmaların veri madenciliği veri analiz metodu dağılımı.....	60
Şekil 4.11. Çalışmalarda kullanılan yazılım dağılımı	60

ÇİZELGELER DİZİNİ

Çizelge 1.1. Onur ve yüksek onur alan öğrencilerin listesi (Silahtaroglu, 2016)	8
Çizelge 1.2. Eğitsel veri madenciliği paydaşları ve hedefleri (Romero ve Ventura, 2013)	11
Çizelge 1.3. Eğitsel veri madenciliği görev ve uygulamaları (Romero ve Ventura, 2010)	14
Çizelge 1.4. Bilgisayar tabanlı eğitim ortamları (Romero ve Ventura, 2013)	16
Çizelge 1.5. Eğitsel veri madenciliği popüler yöntemleri (Romero ve Ventura, 2013)..	18
Çizelge 2.1. EVM konulu alanyazın inceleme çalışmaları	24
Çizelge 3.1. Makale inceleme ölçütleri ve değerleri.....	28
Çizelge 4.1. Kategoriler ve anahtar kelime sayısı dağılımı	35
Çizelge 4.2. Örneklem, teknoloji, merkezi öğrenci değerlendirme sınavı ve konu/alan/disiplin kategorileri ve kodları	37
Çizelge 4.3. E-öğrenme ortamı, öğrenme kuram ve modelleri ve veri ile ilgili kavramlar kategorileri ve kodları	39
Çizelge 4.4. Veri türleri, veri analiz yazılımı ve veri analizi kategorileri ve kodları.....	41
Çizelge 4.5. Teori/model/süreç/teknik, veri madenciliği görevi ve öğrenci performansı kategorileri ve kodları	43
Çizelge 4.6. Öğrenci atipik davranış, öğrenci stil/beceri/davranış ve eğitim-öğretim ile ilgili kavramlar	45
Çizelge 4.7. Çalışmaların yazar dağılımı	47
Çizelge 4.8. 1 Ekim 2019 tarihi itibarıyla Google Scholar ve Scopus veri tabanında ilk 10 makale	48

SİMGELER VE KISALTMALAR DİZİNİ

EVM: Eğitsel veri madenciliği

LYS: Liseye Geçiş Sınavı

MOOC: Massive Open Online Course (Kitlesel Açık Çevrim İçi Dersler)

ÖA: Öğrenme analitikleri

ÖYS: Öğrenme yönetim sistemi

TIMMS: Trends in International Mathematics and Science Study (Uluslararası Matematik ve Fen Eğilimleri Araştırması)



1. GİRİŞ

Bilgi ve iletişim teknolojilerinin gelişmesi veri depolama sistemlerinin gelişmesini de olumlu yönde etkilemiş, verinin düşük maliyette depolanmasına ve veri miktarlarının çok büyük boyutlara ulaşmasına neden olmuştur. Bu gelişme ile depolanan veri miktarı artış göstermiş, verilerin analizi için geleneksel veri analizi yöntemleri yetersiz kalmıştır. Büyük veri tabanlarında bulunan verilerin analizi için geleneksel yöntemlere alternatif olarak disiplinler bir alan olan veri madenciliği alanı ortaya çıkmıştır. Pazarlama, sağlık, ticaret, banka, eğitim gibi alanlarda elde edilen büyük verilerden, amaca uygun işe yarar bilgiler elde etmede veri madenciliğinden yararlanılmıştır. Veri madenciliği, veri kalıplarını bulmak, veri tabanındaki gizli ilişkileri ortaya çıkarmak, verilerin içindeki ilişkileri inceleyerek aralarındaki bağlantıyı bulmak, verilerin homojen kümelerini oluşturmak ve bilgisayar tabanlı bilgi sistemi tarafından kolay üretilmeyen bilgileri açığa çıkarmak için kullanılmaktadır (Pena-Ayala, 2014).

Büyük veriden anlamlı bilgi oluşturma işlevi gören veri madenciliği ile eğitim ortamlarından elde edilen veriler ile eğitimi iyileştirici bilgiler oluşturma süreci yeni bir veri madenciliği alanı olan eğitsel veri madenciliği (EVM) alanını ortaya çıkarmıştır. EVM öğrenen özelliklerinin açığa çıkarılmasına, öğrenme ortamlarının iyileştirilmesine, öğrenenlerin ilgi ve tutuma göre modellenmesine, benzer özelliklere sahip öğrenenleri gruplandırılmasına, normal olmayan öğrenci davranışlarının açığa çıkarılmasına, daha iyi bir eğitim ortamının oluşturulmasına, uyarlanabilir eğitim ortamı oluşturulmasına ve kavramlar arasında ilişkilerin açığa çıkarılmasına katkı sağlamaktadır.

Ülkemizde de eğitimi iyileştirmek adına Milli Eğitim Bakanlığı tarafından yayınlanan 2023 Vizyon Belgesi'nde veriye dayalı yönetimden bahsedilmiştir. Vizyon belgesinde bakanlığın tüm kararlarının veriye dayalı hale getirilmesi ve okul bazında veriye dayalı yönetime geçmek için iki farklı alanda eylem planları hazırlanmıştır. Bakanlığın kararlarını veriye dayalı hale getirmesi için bakanlığa bağlı mevcut bilişim sistemlerinden gelen veriler kolay erişilebilmesi ve birlik içinde olması için Eğisel Veri Ambarları geliştirilecek; öğrencilerin akademik, ilgi, yetenek ve mizacına yönelik verilerin bir arada tutulacağı "Öğrenme Analitiği (ÖA) Platformu" kurulacak; öğrenimi ve öğretimi daha iyi anlamak için öğrenme analitiği araçları geliştirilecek; verilerle ilgili işlemler için Veri Denetim Birimi kurulacak ve gerekli mevzuat değişiklikleri yapılacaktır;

üniversiteler ve sivil toplum kuruluşları tarafından yapılan bilimsel çalışmaları derlemek için birim oluşturulacaktır. Okul bazında veriye dayalı yönetime geçilmesi için okul gelişim planlarına erişim için çevrim içi bir platform oluşturulacak; kaynak planlaması için Coğrafi Bilgi Sistemi kurulacak; öğretmen-veli-okul etkileşimini artırmak için Veli Bilgilendirme Sistemi geliştirilecek ve desteğe ihtiyaç duyan öğrencilerin tespiti için gerekli eylemlere yer verilmesi sağlanacaktır (MEB,2018). Eğitsel veri madenciliği alanı eğitimi iyileştirmek için gerekli olan can suyunu sağlayan bir alan olarak görülmeli ve önemsenmelidir.

EVM gelişmekte olan bir alan olması sebebiyle bu alanla ilgili çalışmaların incelendiği alanyazın taraması çalışmalarının yapılmasının alan eğilimlerinin görülmesine katkı sağladığı düşünülmektedir. Romero ve Ventura (2007), Baker ve Yasef (2009), Romero ve Ventura (2010), Pena-Ayala (2014), Tekin ve Öztekin (2016) alan eğilimlerini ve gelişimini genel olarak ortaya çıkaran; Mohamad ve Tasir (2013) e-öğrenme ortamındaki çalışmaların; Dutt ve ark. (2017) kümeleme yöntemini kullanan çalışmalarının alan eğilimlerini daha özel olarak ortaya çıkaran çalışmalardır. Bu çalışmanın amacı 2013-2018 yılları arasında belirlenen veri tabanlarında yayınlanan eğitsel veri madenciliği ile ilgili çalışmaların belirli değişkenlere göre incelenmesidir.

1.1. Araştırmanın Problem Durumu

Çalışmanın problem durumu, alanyazın incelemeleri sonucunda 2013-2018 yılları arasında ulusal ve uluslararası yayınlanan eğitsel veri madenciliği ile ilgili makale çalışmalarının incelenmesini konu edinen tez ve makale çalışmasının olmamasıdır.

1.2. Araştırmanın Amacı

Bu araştırmanın amacı, 2013-2018 yılları arasında ‘eğitsel veri madenciliği’ kullanan bilimsel makalelerin çeşitli değişkenler açısından incelenmesidir.

1.3. Araştırma Soruları

Araştırma Sorusu 1: 2013-2018 yılları arasında yayımlanan; eğitsel veri madenciliği kullanan yayınların demografik özellikleri nelerdir?

1. Çalışmaların yıllara göre dağılımı nasıldır?
2. Çalışmaların erişim sağlanan veri tabanı dağılımı nasıldır?
3. Çalışmaların yazar dağılımı nasıldır?
4. Çalışmaların anahtar kelime dağılımı nasıldır?
5. Çalışmaların atıf sayısı dağılımı nasıldır?
6. Çalışmaların araştırma konusuna göre dağılımı nasıldır?

Araştırma Sorusu 2: 2013-2018 yılları arasında yayımlanan; eğitsel veri madenciliği kullanan yayınların örneklem özellikleri nelerdir?

1. Çalışmaların çalışma grubu dağılımı nasıldır?
2. Çalışmaların çalışma grubu boyutuna göre dağılımı nasıldır?

Araştırma Sorusu 3: 2013-2018 yılları arasında yayımlanan; eğitsel veri madenciliği kullanan yayınların veri toplama süreci nasıldır?

1. Çalışmaların veri toplama ortamına göre dağılımı nasıldır?
2. Bilgisayar tabanlı eğitim ortamı olan çalışmaların ortam türü dağılımı nasıldır?
3. Çalışmaların veri toplama aracına göre dağılımı nasıldır?

Araştırma Sorusu 4: 2013-2018 yılları arasında yayımlanan; eğitsel veri madenciliği kullanan yayınlarda veri madenciliğine ilişkin kullanılan yöntemler nelerdir?

1. Çalışmaların veri madenciliği modelleme yöntemine göre dağılımı nasıldır?
2. Çalışmalarda veri madenciliği veri analiz yöntemine göre dağılımı nasıldır?
3. Çalışmalarda veri madenciliği veri analizi için hangi metodlar kullanılmıştır?
4. Çalışmalarda veri analizi için hangi yazılımlar kullanılmıştır?

Araştırma Sorusu 5: 2013-2018 yılları arasında yayımlanan; eğitsel veri madenciliği kullanan yayınların öğrenci performansını tahmin etme, atipik davranış tespit etme ve geri bildirim sağlama araştırma konulu yayınların araştırma sonuçları nasıldır?

1.4. Araştırmanın Önemi

Romero ve Ventura (2007) yaptıkları çalışmada eğitsel veri madenciliği konusunda 1995-2005 yılları arasında alanın ilk çalışmaları olarak nitelendirilen çalışmaları analiz etmiştir. Baker ve Yasef (2009) yaptıkları çalışma ile eğitsel veri madenciliği alanında ilk yıllarda yayınlanan makaleleri tartışmış ve 2008-2009 yıllarında yapılan çalışmaların eğilimleri ile karşılaştırmıştır. Romero ve Ventura (2010) yaptıkları çalışmada o güne kadar eğitsel veri madenciliği ile en ilişkili makaleleri analiz etmişlerdir. Pena-Ayala (2014) ise 2010 yılı ile 2013 yılının ilk çeyreği arasında yayınlanan çalışmaları analiz etmiştir. Tekin ve Öztekin (2016) 2006 ile 2016 yılları arasında yayınlanan 140 yayını çeşitli değişkenler açısından analiz etmişlerdir. Dutt, Ismail ve Herawan (2017), 1983-2016 yılları arasında eğitsel veri madenciliği alanında kümeleme yöntemi ile ilgili yayınlanan 166 çalışmayı; Widyahastuti ve Tjhin (2018) veri madenciliği teknikleri kullanılarak çevrim içi tartışma forumlarında performans tahmini ile ilgili 2011-2016 yılları arasında yayınlanan 22 çalışmayı öğrenci performansını tahmin etme yöntemi ve kullanılan veri madenciliği yöntemleri açısından; Manjarres, Sandoval ve Suárez (2018), 1993-2015 yılları arasında yayınlanan toplamda 305 yazarın katıldığı ve 12 veri madenciliği tekniğinin tanımlandığı 127 çalışmada eğitsel veri madenciliği görevi ve kullanılan veri madenciliği tekniğini analiz etmişlerdir. Son yıllarda yapılan alanyazın inceleme çalışmalarının araştırma konusu, veri madenciliği yöntemi ve tekniği açısından belirli değişkenlere odaklanmıştır. Alanla ilgili yapılan alanyazın taramasında 2013-2018 yılları arasında Türkiye’de ve uluslararası bu alanda yapılmış eğitsel veri madenciliği çalışmalarını kapsayan ve anahtar kelime analizi ve sonuç analizine odaklanan bir literatür inceleme çalışmasına rastlanamamıştır. Bu literatür inceleme çalışması, eğitsel veri madenciliği alanında çalışmak isteyen araştırmacılara alanın güncel ve geçmişteki eğilimlerini görmesine olanak tanıyarak önemli bir boşluğu dolduracaktır.

1.5. Sayıtlar

Bu çalışmada kabul edilen varsayımlar şunlardır:

- Hatay Mustafa Kemal Üniversitesi kütüphanesinin tam metin erişim sağladığı çevrim içi veri tabanlarının eğitsel veri madenciliği konusunda 2013-2018 yılları

arasında yayınlanan tüm makalelerin önemli bir kısmını kapsadığı varsayılmıştır.

- İncelenen makalelerin doğru ve gerçek bilgiler sunduğu varsayılmıştır.

1.6. Sınırlılıklar

Araştırma Hatay Mustafa Kemal Üniversitesi Kütüphanesinin tam metin erişim sunduğu veri tabanlarında bulunan makaleler ile sınırlıdır.

1.7. Veri Madenciliği

Bilgi ve iletişim teknolojilerinin gelişmesi veri depolama sistemlerinin gelişmesini de olumlu yönde etkilemiş, verinin düşük maliyetle depolanmasına ve veri miktarlarının çok büyük boyutlara ulaşmasına neden olmuştur. Elde bulunan veri bir amaca hizmet etmediği sürece değersiz görülmüştür. Veri analizi için yeterli olmayan eski yöntemlere alternatif olarak istatistik, veri tabanı ve bilgisayar teknolojilerini bir araya getiren çok disiplinli bir alan olan veri madenciliği ortaya çıkmıştır. Bu ihtiyaçlar sonucunda makine öğrenmesi, istatistik, veri tabanı ve bilgisayar teknolojilerini bir araya getiren çok disiplinli bir alan olan veri madenciliği ortaya çıkmıştır (Aydın, 2007). Veri madenciliği büyük miktardaki veriyi yorumlama ve analiz etme için ortaya çıkan yeni nesil araç ve tekniklerdir.

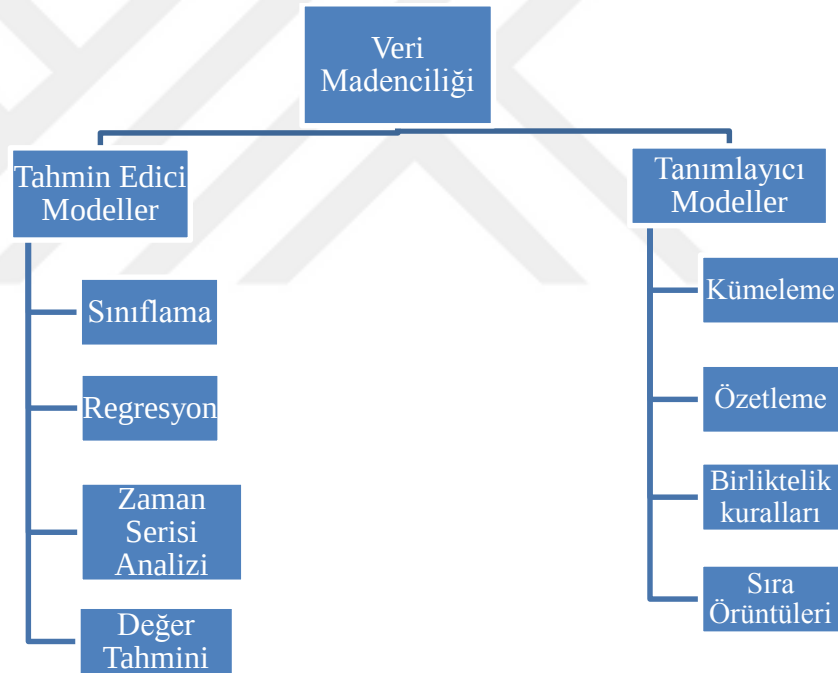
Emel ve Taşkın (2005), veri madenciliğini istatistiksel algoritmaları ve yapay zekâ yöntemlerini kullanarak veri tabanlarında bulunan büyük veri içindeki gizli örüntüleri ve ilişkileri bulmak olarak tanımlamıştır. Aydın (2007), veri madenciliğini hedeflere ulaşmak için hazırlanmış veriyi analiz etmek için algoritmaların uygulandığı bilgi keşif süreci olarak ele almıştır. Özbay (2015), veri madenciliğinin eldeki sistemleri geliştirmek ve eksikliklerini giderebilmek amacıyla elde olan verilerden anlamlı bilgiler elde etmeyi sağladığını belirtmiştir. Veri madenciliğini, kayıt altına alınan her veri ile analiz yapılması sonucunda her alanda kullanmak mümkündür. Veri madenciliği sağlık, endüstri, mühendislik, pazarlama ve eğitim alanlarında elde edilen verilere uygulanarak bu alanlarda sık kullanılmaktadır.

Veri madenciliği geleneksel sınıf ortamında ve dijital eğitim-öğretim ortamında elde edilen verileri kullanıp anlamlı bilgiler ortaya çıkararak eğitim alanına katkı

sağlamaktadır. Veri madenciliği, okuldan ayrılma davranışı sergileyebilecek öğrencileri belirleme, öğrenci motivasyonunu etkileyen değişkenleri belirleme, öğrenme yönetim sistemini (ÖYS) etkili kullanma, bireyselleştirilmiş öğrenme sistemlerinin gelişimi için veri analizi, kurs ve eğitsel yazılım oluşturma için gerekli olan geri bildirimi sağlama vb. durumlar için kullanılabilir.

1.7.1. Veri Madenciliği Modelleme Yöntemi

Dunham (2003) tanımlayıcı ve tahmin edici modeller olmak üzere veri madenciliği modelleme yöntemlerini iki kategori altında incelenmektedir. Dunham (2003)'e göre veri madenciliği modelleri ve işlevleri Şekil 1.1'de gösterilmiştir.

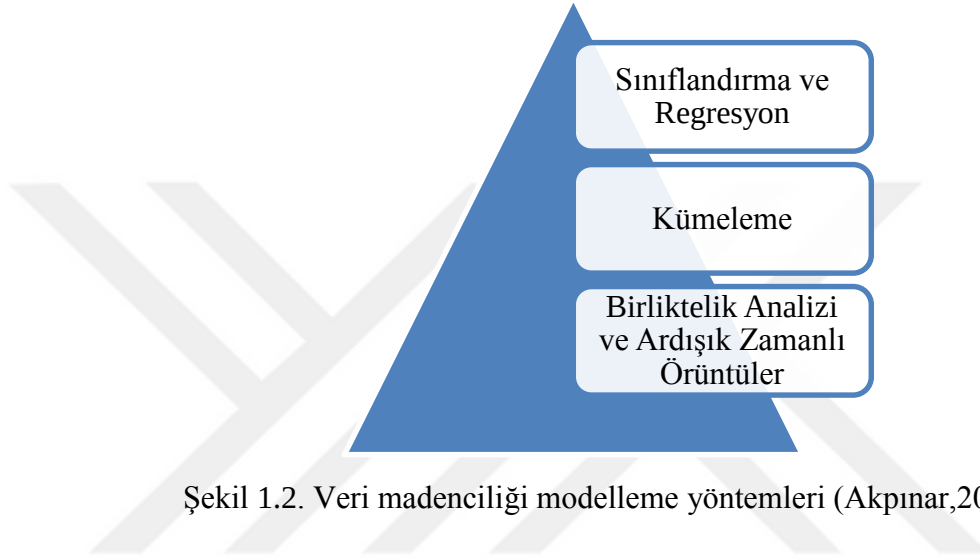


Şekil 1.1. Veri madenciliği modelleri ve işlevleri (Dunham, 2003)

Veri madenciliği modelleme yöntemlerini regresyon ve sınıflandırma, kümeleme ve birliktelik kuralları kategorilerine ayırabiliriz (Akpınar, 2000). Tanımlayıcı modeller, karar verme sürecine yardımcı olabilecek verilerdeki örüntüleri, kalıpları ve ilişkileri tanımlamayı sağlamaktadır. Tahmin edici modeller ise farklı verilerde bulunan bilinen sonuçları kullanarak model geliştirilmesini ve bundan yola çıkarak bilinmeyen tahmin etmeyi amaçlar.

1.7.2. Veri Madenciliği Yöntemleri

Veri madenciliği modelleme yöntemleri altında bulunan sınıflama, regresyon, zaman serisi analizi, değer tahmini, kümeleme, özetleme, birliktelik kuralları ve sıra örüntüleri işlevlerine göre üç kategoriye ayrılmaktadır (Akpınar, 2000). Akpınar (2000)'e göre veri madenciliği modelleme yöntemleri Şekil 1.2'de gösterilmiştir.



Şekil 1.2. Veri madenciliği modelleme yöntemleri (Akpınar,2000)

Sınıflandırma ve Regresyon

Eldeki verilerden yola çıkarak gelecekteki durumun tahmin edilmesinde en fazla faydalanan veri madenciliği yöntemi sınıflandırma ve regresyondur. Sınıflandırma yöntemi eldeki veriye eklenen yeni bir veriyi önceden ayrılmış kategorilere atmayı amaçlar. Sınıflama kategorik değerleri tahmin ederken, regresyon süreklilik gösteren değerlerin tahmin edilmesinde kullanılır (Kamber ve Morgan, 2000). Örneğin, sınıflandırma yöntemi bir bankanın kredi uygulamalarının güvenilir ve güvenilir değil şeklinde kategoriye ayırması amacıyla kullanılırken, regresyon modeli maaşı ve mesleği verilen müşterilerin bilgisayar parçaları alırken yapacakları ödeme tutarlarını tahmin etmek için kullanılabilir (Özekeş, 2003). Bir kutuda ne kadar mavi renk olduğunun veya ertesi gün havanın nasıl olacağının tahmin edilmesi aslında bir sınıflandırma işlemidir (Dunham 2003). Sınıflandırma ve regresyon yöntemlerinde kullanılan başlıca teknikler (Han ve Kamer, 2006):

- Karar ağaçları (Decision Trees),
- Yapay sinir ağları (Artificial Neural Networks),
- Genetik Algoritmalar (Genetic Algorithms),
- K-En Yakın Komşu (K-Nearest Neighbor),
- Naive Bayes
- Lojistik Regresyon (Logistic Regression).

Örneğin, Çizelge 1.1’de her bir öğrencinin onur, yüksek onur veya ödölsüz olduđu aşıktır. Aynı zamanda bu çizelge içinde bu öğrencilerin hangi kriterler dâhilinde onur ve yüksek onur ödöllerini aldıkları da gizlidir. İşte sınıflandırma algoritmaları, bu gizli kalmış kuralları ortaya çıkararak, veritabanında kaydı bulunmayan, ancak verileri sonradan girilen yeni bir öğrencinin onur, yüksek onur öğrencisi olup olmadığına karar verir (Silahtaröglü, 2016).

Çizelge 1.1. Onur ve yüksek onur alan öğrencilerin listesi (Silahtaröglü, 2016)

Öğrenci adı	Yıl ort.	Yıl ort	Yıl ort	Spor durumu	Disiplin durumu	Ödöl
Aysel	2.5	3.5	2.8	***	*****	ONUR
Sevim	2.8	3.9	2.9	****	*****	Yüksek ONUR
Ayhan	3.2	2.4	3.0	**	*****	ONUR
Can	4.0	3.5	2.9	**	****	Ödölsüz

Eğitim verilerinde sınıflandırma teknikleri sistemden elde edilen öğrenci sonucunu tahmin etme, zayıf öğrencileri belirleme ve elde edilen sonuçları iyileştirme konusunda yardımcı olmak için kullanılabilir (Kumar ve Vijiyalakshimi, 2011).

Kümeleme

Kümeleme analizi, otomatik sınıflandırma olarak adlandırılan ve sınıflandırmada olduđu gibi sahip olunan verileri gruplayarak ayırma işlemidir (Silahtaröglü, 2016; Han ve Kamber, 2006). Kümeleme yöntemi gruplar arasındaki benzerliğı ve farklılığı maksimuma çıkarma ilkesine dayanır (Han ve Kamber, 2006). Kümelemede oluşturulan grupların benzerlik ve farklılıklarına göre otomatik oluşturulması ve veri etiketinin kullanılmaması kümeleme ve sınıflandırma arasındaki en önemli farktır (Han ve Kamber, 2006). Kümeleme analizi, heterojen yapıdaki büyük veri

kümelerinde bulunan verileri benzer veya birbirleriyle ilişkili olma durumlarına göre homojen veri kümeleri oluşturmaktır (Özbay, 2015). Bu sebeple birbirleriyle ilişkili olan nesneler aynı küme, ilişkili olmayan nesneler farklı küme oluşturur. Nesnelerin birbirleriyle ilişkili olmasından kasıt değerlerin sayısal olarak birbirlerine yakın olmasıdır.

Örneğin, kitlesel açık ders sisteminde bir kursa katılan öğrencilerin isim ve yaş bilgileri veritabanında bulunmaktadır. Kurs katılımcılarını kümelemek istersek, yaşlara göre kümelemek daha doğru olacaktır. Kümeleme analizi yapılırken yaşları birbirine yakın olanlar aynı kümede olacaktır. Yaşları 20, 21, 23, 25, 32, 33, 35 ve 35 olan 8 katılımcı varsa, 20-25 yaş aralığındakiler ve 32-35 yaş aralığındakiler iki ayrı küme olarak toplanacaktır. Çünkü 20-25 yaş aralığında bulunan kişiler birbirlerine daha yakın olduklarından bir kümede; 32-35 yaş aralığında bulunan kişiler birbirlerine daha yakın olduklarından bir kümede toplanacaktır. Eğer veritabanındaki yaşlar 19, 20, 22, 22, 25, 26, 27 ve 27 olsaydı 19-22 yaş aralığı bir kümede, 25-27 yaş aralığı bir kümede olacaktı. İlk örnekte 20 yaşında biri ile 25 yaşında birisi aynı kümede iken, ikinci örnekte ayrı kümelerde bulunacaklardır. Kümeleme analizinde görüldüğü gibi veriler sadece değerlerine göre değil, diğer verilere olan benzerliği ve diğer verilerin durumuna göre kümelere ayrılır.

Birliktelik Analizi ve Ardışık Zamanlı Örüntüler

Birliktelik analizi, veri tabanında bulunan bir grup verinin diğer verilerle olan ilişkilerini eş zamanlı olarak açığa çıkarır. Özellikle satış-pazarlama alanında herhangi bir ürünü satın alırken, aynı anda bu ürünün yanında başka bir ürün satın alınması ile ürünler arasındaki bağlantıyı ortaya koymasından dolayı “Pazar Sepeti Analizi” olarak sık kullanılmaktadır. Üniversiteye yerleşen bir öğrencinin lise mezuniyet notu ile üniversite mezuniyet notu arasındaki ilişkiyi ortaya çıkarmak için birliktelik analizi yapılır.

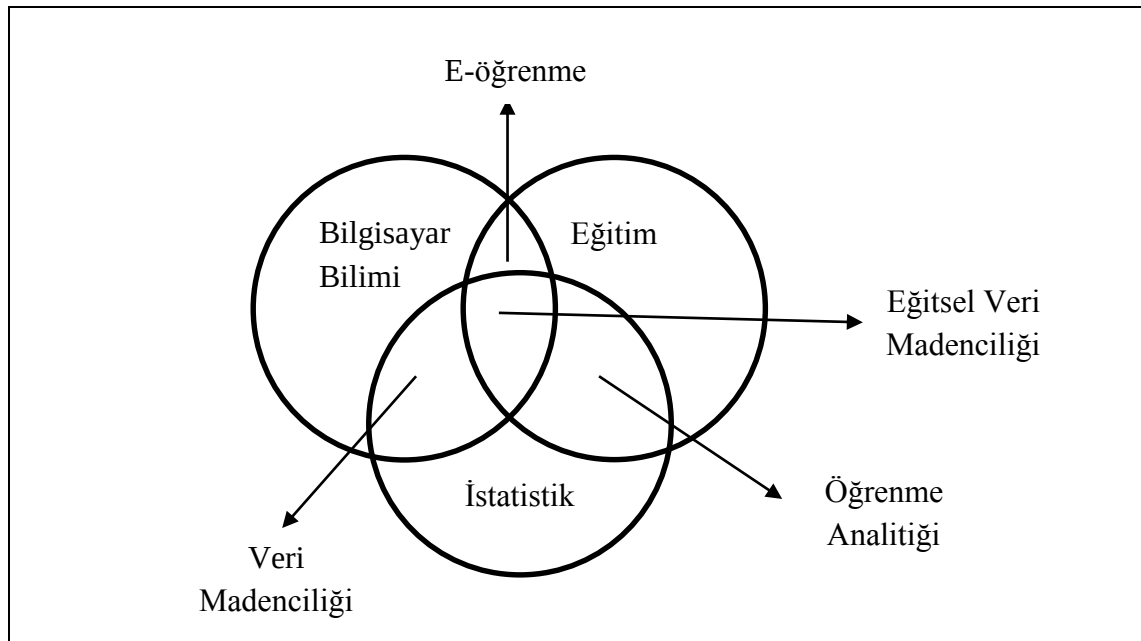
Ardışık zaman örüntüleri de birliktelik analizindeki gibi veriler arasındaki ilişkiyi ortaya çıkarır fakat birliktelik analizi eş zamanlı ortaya çıkarırken ardışık zaman örüntüleri faaliyetlerin zaman sırası üzerinden ilişkileri ortaya çıkarır. Örneğin; birliktelik analizi marketten aynı anda aldığınız iki ürün arasındaki ilişkiye bakarken, ardışık

zamanlı örüntüler marketten alınan bir ürün ile o hafta içinde alınan başka bir ürün arasındaki ilişkiyi incelemektedir.

Birlikte analiz madenciliği ile öğrenci verileri içindeki öğrencinin bir öğrenme ortamındaki davranışı, öğrenme stilleri, sınav modelini ve değerlendirmesi gibi gizli ilişkileri ortaya çıkarmak için kullanılabilir (Kumar ve Vijiyalakshimi, 2013).

1.8.Eğitsel Veri Madenciliği

Teknoloji gelişimi eğitim ve öğretim ortamlarının değişimine de sebep olmuştur. Dijital dünyada eğitim ortamı olan e-öğrenme, öğrenme yönetim sistemleri ve kitlesel açık çevrim içi ders ortamlarına günden güne talep artmıştır. Artan öğrenen sayısına bağlı olarak elde edilen veri miktarında da artış olmaktadır. Araştırmacılar eğitim ortamlarında artan veri miktarını öğrenenin, öğretmenin, yöneticinin ve sistem geliştiricileri için anlamlı hale getirmek için veri madenciliğine başvurmuşlardır. Büyük veriden anlamlı bilgi oluşturma işlevi gören veri madenciliği ile eğitim ortamlarından elde edilen veriler ile eğitimi iyileştirici bilgiler oluşturma süreci yeni bir veri madenciliği alanı olan eğitsel veri madenciliği alanını ortaya çıkarmıştır. Romero ve Ventura (2013) eğitsel veri madenciliği alanını bilgisayar bilimi, eğitim ve istatistik alanları ile ilişkili olan disiplinler bir alan olarak tanımlamıştır (Şekil 1.3).



Şekil 1.3. Disipliner alan EVM (Romero ve Ventura, 2013)

Eğitsel veri madenciliği topluluğu web sitesinde, www.educationaldatamining.org, eğitsel veri madenciliği şu şekilde tanımlanmaktadır; “Eğitim ortamlarından gelen ve artan benzersiz veriyi araştırmak için yöntemler geliştirmek ve öğrencileri daha iyi anlamak için bu yöntemleri kullanmakla ilgilenen ve gelişmekte olan bir disiplindir.” Eğitim ortamlarından elde edilen verilerin veri madenciliği süreci işlemlerinden geçerek elde edilen yararlı bilgilerin öğrencilerin anlaşılmasına ve eğitim ortamlarının iyileştirilmesine katkı sağlaması ile ilgilenen disipline eğitsel veri madenciliği denilmektedir (Siemens ve Baker, 2012). Eğitsel veri madenciliği, öğrencilere fayda sağlamak için, veri madenlerini kullanarak öğrenenleri ve öğrenmeyi daha iyi anlamaya çalışarak veri ve teoriyi birleştiren hesaplamalı yaklaşımlar geliştirmeyi amaçlamaktadır (Romero ve Ventura, 2010).

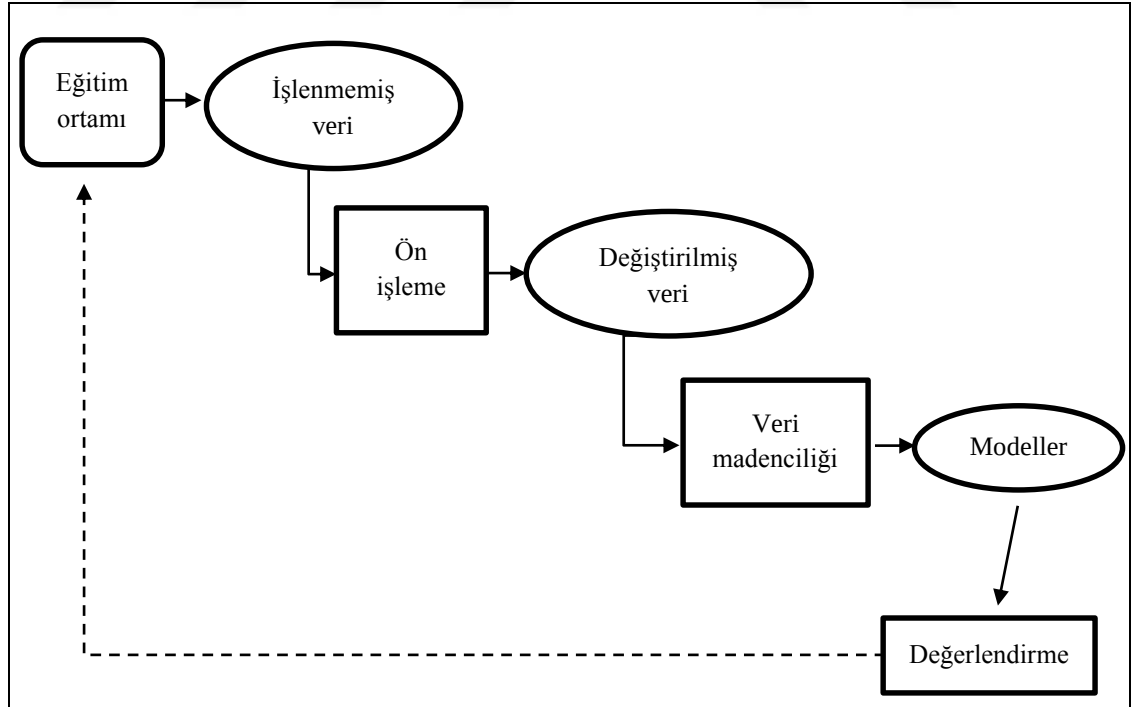
Eğitsel veri madenciliğinin amacı bir çok kaynakta eğitimi iyileştirmek olarak belirtilmiş olsa da; eğitim, çok yönlü bir sistem olduğu için veri madenciliğini daha önce keşfedilmemiş bilgileri ortaya çıkararak eğitim ortamlarının genel bakış açısı ve çözülmesi gereken durumlarla ilgili paydaşlara göre özel hedefleri de vardır. Romero ve Ventura (2013)’ya göre eğitsel veri madenciliğinde paydaşlar ve paydaş hedefleri açısından özel hedefler Çizelge 1.2’de verilmiştir.

Çizelge 1.2. Eğitsel veri madenciliği paydaşları ve hedefleri (Romero ve Ventura, 2013)

Paydaşlar	Hedefler
Öğrenciler	Öğrencilere bireyselleştirilmiş geri bildirim sağlamak, öğrenen performansını arttırmak
Eğitimciler	Öğrencilerin öğrenme süreçlerini anlamak, öğretim performansını artırmak, sosyal, bilişsel ve davranışsal yönleri anlamak
Araştırmacılar	Özel öğretim görevi veya problemi çözmek için en faydalı çözüm yöntemini bulmak, öğrenme etkinliğini değerlendirebilecek veri madenciliği teknikleri geliştirmek
Yöneticiler	Kurumsal kaynaklardan en iyi yararlanma yöntemini bulmak

1.8.1. Eğitsel Veri Madenciliği Süreci

Eğitsel veri madenciliğinde amaç sadece veriyi bilgiye dönüştürmek değildir, amaç aynı zamanda öğrencinin öğrenmesini geliştirmek ve eğitim ortamının nasıl iyileştirileceğine karar vermektir. Bu sebeple veri madenciliğinin eğitim verilerine uygulanması sürecini diğer alanlara uygulanmasından farklılaştıran üç durum vardır. Birincisi amaçtır, iş dünyasında müşteri sayısı, para miktarı ve kazancı arttırmaya yönelik konular için veri madenciliği uygulanırken, eğitsel veri madenciliği eğitim olaylarını derinlemesine inceleyerek öğrenme sürecini iyileştirmek ve öğrencilerin öğrenmesine rehberlik etmek gibi araştırma hedeflerine uygulanmıştır. İkincisi veridir, çeşitli eğitim ortamlarında kendi alanına özgü, kendi anlamları yüklü ve kendi hiyerarşisi olan çok sayıda veri olması ve eğitim verilerine veri madenciliği uygulanırken verilerin pedagojik yönlerinin de ele alınması gerekmesidir. Farklılaştıran üçüncü durum ise tekniktir. Eğitimsel veriler ve durumlar için madencilik konusunun farklı şekilde ele alınmasını gerektiren özelliklere sahiptir (Romero ve Ventura, 2007). Romero ve Ventura (2013) Eğitsel bilgi keşfi ve VM süreci Şekil 1.4’de gösterilmiştir.



Şekil 1.4. Eğitsel bilgi keşfi ve veri madenciliği süreci (Romero ve Ventura, 2013)

Herhangi bir eğitim ortamından yani geleneksel eğitim, e-öğrenme, öğrenme yönetim sistemi gibi ortamlardan alınan veri ön işlemden geçirilerek analize hazır hale getirilir. Analize hazır hale gelen veri, amaca göre uygulanacak veri madenciliği yöntemi ile çözümlenir. Sonuçlar yorumlanarak eğitim ortamını iyileştirecek önerilerde bulunulur.

1.8.2. Eğitsel Veri Madenciliğinin Görevleri ve Uygulamaları

Veri madenciliği sadece eğitim ortamını iyileştirmek ve öğrenme sürecini anlamak amacıyla yapılmış gibi görülsede, eğitsel veri madenciliğinde öğrencilere, öğretmenlere, yöneticilere ve araştırmacılara özel amaçları vardır. Veri madenciliği eğitim ortamında çok farklı sorunları çözmek için kullanılmıştır. Baker (2009), Romero ve Ventura (2010), Castro ve ark. (2007) çalışmalarında eğitsel veri madenciliği görevlerini ve uygulamalarını öğrenci performansını tahmin etme, öğrencilere geri bildirimde bulunma, öğretmen, öğrenci ve idarecilere tavsiyelerde bulunma, öğrenci gruplandırma ve modellendirme vb. sınıflandırmışlardır.

Baker (2009) eğitsel veri madenciliğinin dört uygulama alanı olduğunu bildirmiştir. Bunlar öğrenci modelleri geliştirme, öğrenme ve öğrenene yönelik bilimsel araştırmalar, ilgi alanı modelleri geliştirme ve öğrenme yazılımı tarafından sağlanan pedagojik desteğin incelenmesidir. Romero ve Ventura (2010) ise yaptıkları alanyazın incelemesi çalışmasında eğitsel veri madenciliğinin görevlerini ve uygulamalarını 11 kategoriye ayırdıklarını belirtmişlerdir (Çizelge 1.3). Verilerin analizi ve görselleştirilmesi ile geri bildirim sağlama kategorileri eğitimcilere, bireyselleştirilmiş öğrenme için öneri sistemleri öğrencilere bilgi sağlar; öğrenci performansını tahmin etme, öğrenci modelleme, öğrenci gruplandırma ve istenmeyen öğrenci davranışını tespit etme kategorileri öğrenci özelliklerini ortaya çıkarır; sosyal ağ analizi ve kavram haritaları geliştirme kategorileri sırasıyla öğrenciler ve kavramlar arasındaki ilişkiyi ortaya çıkarır; planlama ve zamanlama ile kurs içeriği ve eğitim yazılımı kategorileri ise sırasıyla planlamaya ve kurs içeriği ve yazılım oluştururken yardımcı olur (Romero ve Ventura, 2010). Bu çalışmada seçilen makaleler Romero ve Ventura (2010)'nın ortaya koyduğu eğitsel veri madenciliği görev ve uygulamalarına göre sınıflandırılmıştır.

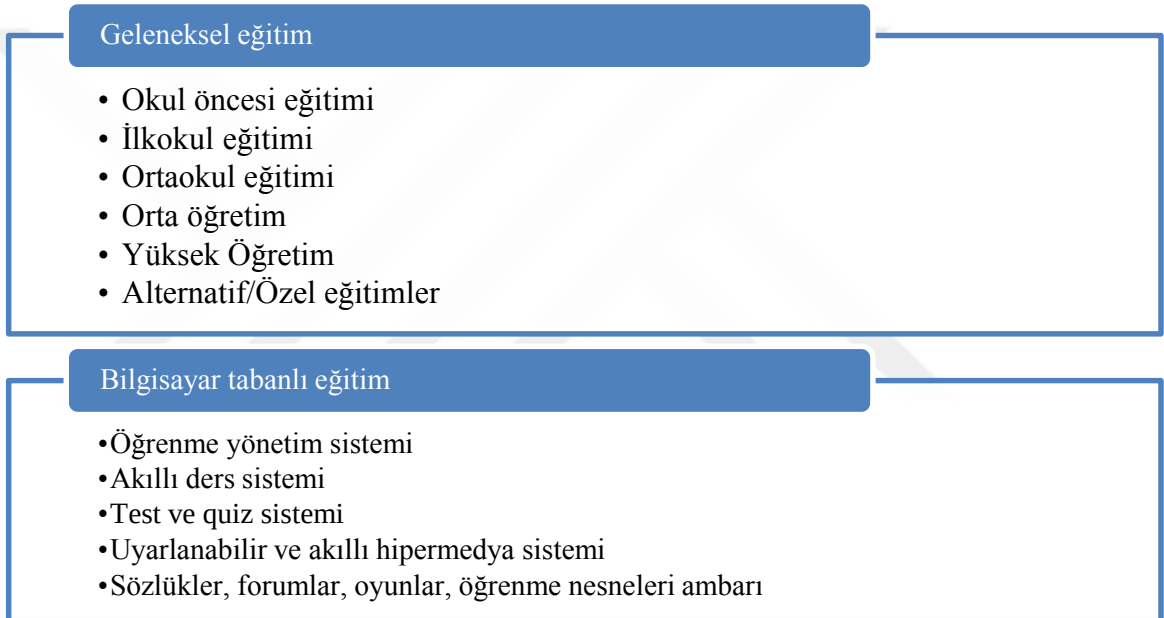
Çizelge 1.3. Eğitsel veri madenciliği görev ve uygulamaları (Romero ve Ventura, 2010)

Görevleri	Amaç	Örnek Çalışma
Verilerin analizi ve görselleştirilmesi	Elde edilen bilgiler ile eğitim ortamını iyileştirmek için karar vermeyi desteklemesi	Lavigne ve ark. (2015) Volk ve ark. (2015) Goldsmith ve ark. (2014) Yahya (2018)
Öğrenci performansını tahmin etme	Öğrenciyi tanımlayan değişkenlerin bilinmeyen değerlerinin tahmin edilmesi	Romero ve ark. (2013) Kabakchieva (2013) Tekin (2014) Migueis ve ark.(2018)
Geri bildirim sağlama	Öğretmenler, öğrenciler ve yöneticileri için öğrenmenin geliştirilmesi ve eğitim kaynaklarının daha verimli organize edilmesi	Chalaris ve ark. (2014) Joorabchi ve ark. (2016) Poitras ve ark. (2016) Verma ve ark. (2017)
Öğrenci gruplandırma	Kişisel özelliklerine göre öğrenci gruplandırma ve gruplara uyarlanabilir içerik sağlanması	Hamoud (2018) İsmail ve Abdullah (2015)
İstenmeyen öğrenci davranışlarını belirleme	Düşük motivasyon, hile yapma, akademik başarısızlık ve okulu bırakma gibi istenmeyen davranışın tespit edilmesi	Marquez-Vera ve ark. (2013) Kaur ve ark. (2015) Marquez-Vera ve ark. (2016) Guarin ve ark. (2015)
Bireyselleştirilmiş öğrenme için öneri sistemleri	Öğrencilerin etkinlik, ziyaret ve bağlantılarının analiz edilmesi ile öğrenciye bireyselleştirilmiş öneride bulunması	Jugo ve ark. (2015) Kausa ve ark. (2018) Luzcando ve ark. (2017)
Sosyal ağ analizi	Sosyal gruplardaki bireyler arasındaki ilişki analizi ile kişiselleştirilmiş tavsiyelerde bulunması	He (2013) Yoo ve Kim (2014)
Öğrenci modellemesi	Öğrencinin tutum ve becerisine göre modellemesi ve öğrenme davranışının modele göre otomatik değerlendirilmesinin yapılması	Lara ve ark. (2014) Akçapınar ve ark. (2015) Aksoy ve Narlı (2015)
Planlama ve zamanlama	Kaynak tahsisi, kurs planlama, ders programı hazırlama ve müfredat geliştirme sürecinin iyileştirilmesi	Soffer ve ark. (2017) Priyambada ve ark. (2017)
Kavram haritası geliştirme	Kavram haritası geliştirme sürecinde eğitimcilere yardımcı olması	Mühling (2016) Shao ve ark. (2018)
Kurs içeriği ve eğitim yazılımı yapılandırma	Kurs içeriği ve eğitim yazılımı oluşturma sürecinde eğitimci ve geliştiricilere geri bildirim sağlayarak yardımcı olması	Northcutt ve ark. (2015) Alexandron ve ark. (2017)

1.8.3. Eğitsel Ortam Türleri

Bilgisayarların eğitim ortamlarına girmesi ile eğitim geleneksel öğrenme ortamları ile sınırlı kalmamış, bilgisayar tabanlı öğrenme ortamları ortaya çıkmıştır. Öğrenme yönetim sistemleri, kitlesel çevrim içi açık dersler, uzaktan eğitim sistemleri, akıllı ders sistemleri ve uyarlanabilir öğrenme ortamları geleneksel eğitime alternatif olarak ortaya çıkan bilgisayar tabanlı eğitim ortamlarıdır (Romero ve Ventura,2013).

Romero ve Ventura (2013) eğitsel veri madenciliğinde elde edilen veri ortamlarını geleneksel eğitim ve bilgisayar tabanlı eğitim ortamları olmak üzere iki kategoriye ayırmıştır (Şekil 1.5).



Şekil 1.5. Eğitim ortamları (Romero ve Ventura, 2013)

Romero ve Ventura (2013), bilgisayar tabanlı eğitimi öğrenme yönetim sistemi, akıllı öğretim sistemi, uyarlamalı eğitim ortamı, test ve sınav sistemleri ve diğerleri olmak üzere beş kategoriye ayırmıştır. Kategoriler ve açıklamaları Çizelge 1.4’te verilmiştir.

Geleneksel eğitim ortamlarında öğretmenler sınıf içi tartışma, öğrenci katılımı, öğrenci notları, müfredat hedefleri vb bilgileri, eğitim kurumları ise geleneksel veri tabanlarındaki öğrenci bilgileri, eğitimcilerin bilgileri, sınıf ve program bilgileri vb veri madenciliği için kullanabilir (Romero ve Ventura, 2013). Bilgisayar tabanlı eğitim ortamlarında ise gelişen veri tabanı teknolojileri ile forum yazışmaları, öğrencinin fare ve klavye hareketleri, öğrencinin sistem ile etkileşimi, öğrenci cevapları, öğrenci puanları

vb özellikleri büyük veri havuzlarında saklanarak, eğitimi iyileştirmek amacıyla yararlı bilgiler elde etmek için kullanılabilir.

Çizelge 1.4. Bilgisayar tabanlı eğitim ortamları (Romero ve Ventura, 2013)

Eğitim ortamları	Açıklama
Öğrenme Yönetim Sistemi	Sınıf içi ve çevrim içi etkinliklerle, eğitim programlarının yönetimi, dokümantasyonu, izlenmesi ve raporlanmasını sağlayan yazılımlardır. Bir kurstaki katılımcılar arasındaki paylaşımı, iletişimi kolaylaştırmak için araçlar sunmaktadır. Okuma, yazma, görevleri yapma vb. etkinlikler öğrenme yönetim sistemi tarafından kaydedilmektedir.
Akıllı Öğretim Sistemleri	Öğrenci davranışlarını modelleyerek öğrencilerin bireysel öğrenme modelini tespit eden ve öğrenciye özelleştirilmiş öğretim sağlayan sistemlerdir. Fare tıklamaları, yazma, konuşma vb. öğretmen öğrenci etkileşimlerini kaydeder.
Uyarlamalı ve Akıllı Hipermedya Sistemleri	Öğrenci ihtiyaçlarına uyum sağlamak için öğrencinin amacı, tercihi ve bilgisini analiz ederek model oluşturur ve etkileşimde bulunan sistemlerdir. Fare tıklamaları, yazma, konuşma vb etkileşimlerini kaydeder.
Test ve Sınav Sistemleri	Katılımcılardan bilgi toplamak için bir dizi soru sorarak katılımcının kavram veya konu hakkındaki bilgi düzeyini ölçen sistemlerdir. Öğrenci cevapları, hesaplanan puanları ve istatistikleri hakkında çok fazla bilgi depolar.
Diğerleri	Kavram haritaları, sosyal ağlar, forumlar, eğitim oyunları, sanal gerçeklik ortamları, öğrenme nesneleri veri havuzları

1.8.4. Eğitsel Veri Madenciliği Yöntemleri

Eğitim ortamında elde edilen verilerden yararlı bilgiler elde etmek için uygulanan çeşitli veri madenciliği yöntemleri vardır. Sınıflandırma, kümeleme ve ilişki madenciliği en popüler yöntemlerdir (Linan ve Perez, 2015). Romero ve Ventura (2007) eğitsel veri madenciliği yöntemlerini;

- İstatistik ve görselleştirme
- Web madenciliği
 - Sınıflandırma, kümeleme ve aykırı değer tespiti
 - Birlikte kuralı madenciliği ve sıralı örüntü madenciliği
 - Metin madenciliği

olarak iki ana kategori ve alt kategorilere ayırmıştır.

Baker (2009), Moore (2006)'nın veri madenciliği taksonomisinden ve Romero ve Ventura (2007)'nin ilk tarama çalışmalarındaki kategorilendirmesinden yola çıkarak

yöntemleri yeniden kategorilendirmiştir. Eğitsel veri madenciliği yöntemlerine farklı bakış açısı getirmiş, çalışmaları şu şekilde kategorilendirmiştir:

- Tahmin
 - Sınıflandırma
 - Regresyon
 - Yoğunluk tahmini
- Kümeleme
- İlişki Madenciliği
 - Birlikte kural çıkarımı
 - Korelasyon madenciliği
 - Sıralı örüntü madenciliği
 - Nedensel veri madenciliği
- İnsan karar alma süreçleri için verinin işlenmesi
- Modellerle keşif (Baker,2009)

Romero ve Ventura (2013) çalışmalarında eğitsel veri madenciliği popüler yöntemlerini 2007 yılında yaptıkları sınıflandırmaya eklemeler yaparak yeniden belirtmişlerdir (Çizelge 1.5).

Çizelge 1.5. Eğitsel veri madenciliği popüler yöntemleri (Romero ve Ventura, 2013)

Yöntem	Genel tanımı	Eğitim alanında işlevi
Tahminleme	Bir verinin bazı özelliklerinden diğer özelliklerini bazı kombinasyonları kullanarak çıkarım yapmaktır.	Öğrenci performansını ve davranışını tahmin etmek.
Kümeleme	Verinin özelliklerini mesafe ölçüsü kullanarak benzer verileri gruplandırmak.	Öğrencileri benzer özelliklerine göre gruplandırmak.
Aykırı Algılama	Verilerin içinden önemli ölçüde farklı olan veriyi keşfetmektir.	Öğrenme güçlüğü olan öğrenciyi tespit etmek.
İlişki Madenciliği	Veri özellikleri arasındaki ilişkiyi tanımlamak ve sonra kullanmak için kodlamaktır.	Öğrencilerin davranış kalıplarındaki ilişkileri tanımlamak.
Sosyal Ağ Analizi	Sosyal bir ortamdaki varlıklar arasındaki ilişkileri analiz etmektir.	İletişim araçları ile iş birlikçi görev sürecindeki etkileşimi ortaya çıkarmak.
Süreç Madenciliği	Bilgi sistemi tarafından kayıt altına alınan olay günlüklerinden süreç ile ilgili bilgi elde etmektir.	Bireyin sisten tarafından kayıt altına alınan günlükleri ile davranışlarını ortaya çıkarmak.
Metin Madenciliği	Eldeki metinden yüksek kalitede bilgi edinmektir	Çeşitli iletişim ortamlarında bulunan metinlerin içeriğini analiz etmek.
İnsan Kararına İlişkin Kararların Damıtılması	Karar verme süreçleri için yararlı olacak bilgiyi özetleme, görselleştirme ve anlaşılır şekilde sunmaktır.	Eğitmenlere yardımcı olması için öğrenci faaliyetlerini görselleştirmek ve analiz etmek.
Modellerle Keşif	Önceden kullanılmış bir yöntemi başka bir veri analizinde kullanmak.	Öğrenci davranışları ile özelliklerinin ilişkilerin tanımlanması psikometrik modelleme çerçevelerinin makine öğrenmesi modellerine entegrasyonunu desteklemek.
Bilgi İzleme	Öğrenci becerilerini tahmin etmek için öğrenci bilgilerini izleyerek ve analiz ederek model oluşturmak.	Öğrenci olay günlüklerini takip etmek.
Negatif olmayan matris ayrışımı	Q Matris açısından doğrudan yorum yapmaya imkân sağlayan tekniktir.	Sınava giren bir kişinin test sonucu verilerini bir matristen iki matrise ayrıştırmak.

2. ÖNCEKİ ÇALIŞMALAR

Bu bölümde yer verilen çalışmalar ulusal ve uluslararası akademik bilimsel veri tabanlarında yer alan dergiler, EVM konferanslarında yayınlanan bildiriler araştırma bildirileri kapsamında taranarak seçilmiştir.

Veri madenciliği çalışmalarının artması ile eğitim ortamlarında yapılan veri madenciliği çalışmalarında da artış görülmektedir. Eğitsel veri madenciliği alanı günümüzde popüler bir alan haline gelmiş ve gün geçtikçe yapılan çalışma sayıları artmaktadır. Alanla ilgili genel durumu ortaya çıkarmak için alanyazın tarama çalışmaları da yapılmıştır.

Eğitsel veri madenciliği alanını ilk tarama çalışmasında Romero ve Ventura (2007), 1995-2005 yılları arasında yayınlanan eğitsel veri madenciliği alanının ilk makaleleri olan 81 makaleyi, elde edilen veri ortamı, eğitsel veri madenciliği görevleri ve veri madenciliği modelleri açısından incelemişlerdir. Veri madenciliği modellerini istatistik-görselleştirme ve web madenciliği olmak üzere iki alanda sınıflandırmıştır. Web madenciliğini de (1) kümeleme, sınıflama ve aykırı değer analizi; (2) birliktelik kuralları ve sıralı örüntü keşfi; (3) metin madenciliği olarak alt kategorilere ayırmışlardır. Elde edilen veri ortamını (1) geleneksel eğitim, (2) e-öğrenme, (3) öğrenme yönetim sistemi, (4) akıllı ders sistemi, (5) uyarlanabilir eğitim ortamı, (6) test-anket, (7) metin-içerik ve (8) diğerleri olarak alt kategorilere; eğitsel veri madenciliği görevlerini (1) veri analizi ve görselleştirilmesi, (2) öğrenci performansını tahmin etme, (3) öğretmenleri desteklemek için geri bildirimde bulunma, (4) öğrenci gruplama, (5) istenmeyen öğrenci davranışlarını tespit etme, (6) öğrenciler için öneriler, (7) sosyal ağ analizi, (8) öğrenci modelleme, (9) planlama ve programlama, (10) kavram haritası geliştirme ve (11) kurs içeriği oluşturma olarak alt kategorilere ayırmışlardır. Sonuç olarak Romero ve Ventura (2007) eğitsel veri madenciliği alanının ilk günlerindeki gibi olmadığını, hızlı bir şekilde geliştiğini, artık olgun bir alan olduğu belirtmişlerdir.

Baker ve Yasef (2009), alanın mevcut eğilimlerini ortaya çıkarmak için yaptığı tarama çalışmasında alanla ilgili 45 çalışmayı eğitsel veri madenciliği modelleri ve kullanım amaçlarına göre incelemişlerdir. Eğitsel veri madenciliği modelleri, Moore (2006)'nın veri madenciliği taksonomisinden ve Romero ve Ventura'nın (2007) ilk tarama çalışmalarındaki kategorilendirmesinden yola çıkarak Baker (2009)'ın kendi

geliştirdiği eğitsel veri madenciliği sınıflandırmasına göre kategorilendirilmiştir. Baker (2009) tarafından geliştirilen eğitsel veri madenciliği model taksonomisi şu şekildedir:

- Tahmin
 - Sınıflandırma
 - Regresyon
 - Yoğunluk tahmini
- Kümeleme
- İlişki Madenciliği
 - Birliktelik kural çıkarımı
 - Korelasyon madenciliği
 - Sıralı örüntü madenciliği
 - Nedensel veri madenciliği
- İnsan karar alma süreçleri için verinin işlenmesi
- Modellerle keşif

Baker ve Yasef (2009), eğitimde veri madenciliğinin kullanım amaçlarını, (1) eğitim yazılımından bireysel öğrenme ve bilgisayar destekli işbirlikli öğrenmede öğrenci modellerinin geliştirilmesi, (2) bir alanın bilgi modellerinin keşfedilmesi ve geliştirilmesi, (3) öğrencilere pedagojik destek sağlanması ve (4) daha iyi bir öğrenme sistemi tasarlamak için öğrenmeyi etkileyen faktörleri derinlemesine inceleme olarak kategorilendirmişlerdir. Halka açık eğitim veri tabanlarının ve çevrim içi kurs araçlarının artması eğitim verilerine daha geniş bir kitlenin ulaşmasını sağladığı böylelikle araştırmacıların önündeki engelin azaldığını, bu nedenle alanla ilgili çalışmaların artmasını beklediklerini ve kullanılan veri madenciliği modellerinin de yıllara göre değişime uğradığını belirtmişlerdir (Baker ve Yasef, 2009).

Romero ve Ventura (2010) 2007 yılında yaptıkları çalışmaya 2010 yılına kadar yapılan 225 çalışmayı daha ekleyerek yeni bir analiz yapmışlardır. Seçtikleri çalışmaları elde edilen veri ortamı ve EVM görevleri açısından analiz etmişlerdir. Yapılan analizler sonucunda EVM araçlarının karmaşık olduğunu, araçların basitleştirilmesi ile eğitimcilerin bu araçları kullanımının yaygınlaşacağını; daha fazla çalışma yapılacağını ve yapılan her çalışma ile alanın gelişeceğini belirtmişlerdir.

Pena-Ayala (2014) 2010'dan 2013 yılının 1. çeyreğine kadar yayınlanan 240 çalışmanın veri madenciliği yaklaşımları ile çıktılarını analiz etmiştir. Çalışmaları model,

yöntem ve teknik; eğitim sistemi çeşitliliği, EVM araştırma konusuna göre analiz etmiştir. EVM araştırma konularını (1) öğrenci modellemesi, (2) öğrenci davranışı modellemesi, (3) öğrenci performansı modellemesi, (4) değerlendirme, (5) destek ve geri bildirim sağlama ve (6) öğretmene yardımcı olma olarak kategorilendirmiştir.

Papamitsiou ve Economides (2014) tarafından 2008-2013 yılları arasında öğrenme analitikleri ve eğitsel veri madenciliği alanlarında yayınlanan, araştırmacı tarafından belirlenen kriterlere uyan 40 deneysel çalışma araştırma stratejisi, araştırma disiplini, öğrenme ortamı, araştırma hedefleri, veri toplama ve analiz tekniği açısından istatistiksel olmayan yöntemlerle analiz edilmiştir. Araştırma hedefleri (1) öğrenci davranış modellemesi, (2) performans tahmini, (3) öğrenciyi tanıma ve yansıtma, (4) okulu bırakma tahmini, (5) geri bildirim sağlama ve (6) veri saklama, ileme ve madencilik yöntemini değerlendirme kategorilerine ayırmışlardır. Yapılan çalışma sonucunda bu çalışmaların büyük hacimde eğitim verilerinin kullanılabilmesi, kişiselleştirilmiş sistemler için model önerileri, kritik anları ve öğrenme kalıplarını ortaya çıkarması ve öğrenme stratejileri ve davranışları hakkında bilgi vermesi güçlü yönleri; çoğunlukla niceliksel araştırma sonuçlarının olması ve farklı araçlar ve uygulamalar için veri standardizasyonunun olmaması zayıf yönleri olarak belirtilmiştir.

Mohamad ve Tasir (2013) önceki eğitsel veri madenciliği çalışmalarının nasıl ele alındığı ve alanın son eğilimlerini ortaya çıkarmak için yapılan çalışmada 2004 ile 2012 yılları arasında e-öğrenme ortamlarında eğitsel veri madenciliği ile ilgili yayınlanan son dokuz çalışmayı kullanılan veri madenciliği model ve yöntemi, alıntı sayısı, yayının kaynağı ve çalışmanın platformu değişkenlerine göre incelemiştir. Sonuç olarak e-öğrenme ortamlarına önem gösterildiği ve artık platformların kendi araçlarını geliştirdiğini belirtmiş; gelecek çalışmalarda odağı sosyal ağ platformlarına yönelerek log kayıtlarının incelenebileceği önerisinde bulunmuşlardır.

Tekin ve Öztekin (2016), 2006-2016 yılları arasında yayınlanmış 140 çalışmayı yerli-yabancı yazar, yerli-yabancı dergi, yayın yılı, veri madenciliği modelleme yöntemi, araştırma konusu, veri türü, çalışma grubu, veri toplama aracı ve veri analiz yöntemi değişkenleri açısından incelemişlerdir. İnceleme sonuçlarına göre yayınlanan çalışmaların çoğunluğunun çalışma grubunun üniversite öğrencisi olduğu; veri toplama aracı olarak web ortamlarının daha çok tercih edildiği; nicel verilerin nitel verilere göre daha çok tercih edildiği ve verilerin daha çok kümeleme ve sınıflandırma yöntemleri ile

analiz edildiğini belirtmişlerdir.

Dutt, Ismail ve Herawan (2017), 1983-2016 yılları arasında eğitsel veri madenciliği alanında kümeleme yöntemi ile ilgili yayınlanan 166 çalışmayı amaç, eğitsel veri madenciliği tekniği, elde edilen veri ortamı ve alıntı sayısı değişkenleri açısından incelemiştir. Çalışma sonucuna göre kümeleme yöntemi öğrenme için gereken zaman, sınıfta öğrenen davranışlarının farklı değişkenler açısından incelenmesi ve öğrenmeye yönelik motivasyona ilişkin çalışmalarda geçerli ve güvenilir sonuçlar elde ettiğini belirtmişlerdir.

Manjarres, Sandoval ve Suárez (2018), 1993-2015 yılları arasında yayınlanan toplamda 305 yazarın katıldığı ve 12 veri madenciliği tekniğinin tanımlandığı 127 çalışmada eğitsel veri madenciliğinin görevi ve kullanılan veri madenciliği tekniği değişkenlerini incelemişlerdir. Çalışmaların daha çok sanal öğrenme ortamlarında öğrenmeyi etkileyen faktörleri araştırmaya odaklandığı; veri madenciliği tekniği olarak en çok birliktelik kuralı kullanıldığı sonucuna ulaşmışlardır.

Eğitsel veri madenciliği alanı ile ilgili alanın eğilimlerini görmek amacıyla İngilizce alanyazın incelemesi çalışmaları yapılmıştır. Bazı çalışmalar alana yön vermesi açısından çalışmaları elde edilen veri ortamına göre incelemeyi daraltmışlardır. Widyahastuti ve Tjhin (2018) veri madenciliği teknikleri kullanılarak çevrim içi tartışma forumlarında performans tahmini ile ilgili 2011-2016 yılları arasında yayınlanan 22 çalışmayı öğrenci performansını tahmin etme yöntemi ve kullanılan veri madenciliği yöntemleri açısından incelemişlerdir. Öğrenci performansını tahmin etme yöntemlerini; öğrenci etkinliğini kullanarak performans tahmin etme, öğrenci notlarını kullanarak akademik performans tahmin etme ve öğrenci davranışlarını kullanarak akademik performans tahmin etme kategorilerine göre sınıflandırmışlardır. Analizler sonucunda veri madenciliğinde kullanılan sınıflandırma yöntemlerinde Naive Bayes ve Karar Ağaçları tekniklerinin öğrenci performansını tahmin etmede en iyi teknik olduğu sonucuna ulaşmışlardır.

Önceki çalışmalar incelendiğinde eğitsel veri madenciliğinin ilk yıllarından itibaren alanyazın inceleme çalışmaları yapıldığı görülmektedir. Alanın gelişimini takip etmek ve araştırmacılara ışık tutması açısından bu çalışmalar yararlı olmuştur. İlk yıllarda belirlenen yıllar arasında yayınlanan çalışmalar seçilmiş olsa da son yıllarda alanyazın analizi için çalışmaların öğrenme ortamı, araştırma konusu, veri madenciliği yöntemi gibi

farklı özelliklerine göre filtrelenerek seçilmiş olduğu görülmektedir. Eğitsel veri madenciliği konulu yapılan alanyazın incelemesi çalışmalarının genel durumunun daha açık bir şekilde görülebilmesi için çalışmalarla ilgili bilgiler özet halinde Çizelge 2.1’de toplu olarak sunulmuştur. Alanyazında 2013 ile 2018 yılları arasında yayınlanan çalışmaların incelendiği bir çalışmaya rastlanmamıştır.



Çizelge 2.1. EVM konulu alanyazın inceleme çalışmaları

Yazarlar	Amaç	Çalışma sayısı	Değişkenler	Sonuç
Romero ve Ventura (2007)	1995 ile 2005 yılları arasında EVM konulu yayınların incelenmesi	81	Elde edilen veri ortamı, EVM görevleri ve veri madenciliği teknikleri	Sonuç olarak Romeo ve Ventura (2007) eğitsel veri madenciliği alanının ilk günlerindeki gibi olmadığını, hızlı bir şekilde geliştiğini artık olgun bir alan olduğu belirtmişlerdir.
Baker ve Yacef (2009)	Eğitsel veri madenciliğinin ilk yıllarında yayınlanan çalışmaların eğilimlerinin incelenmesi	45	Alıntılanma sayısı, eğitsel veri madenciliği yöntemleri	Halka açık eğitim veri tabanlarının ve çevrim içi kurs araçlarının artması eğitim verilerine daha geniş bir kitlenin ulaşmasını sağladığı böylelikle araştırmacıların önündeki engelin azaldığını, bu nedenle alanla ilgili çalışmaların artmasını beklediklerini ve kullanılan veri madenciliği modellerin de yıllara göre değişime uğradığını belirtmişlerdir.
Romero ve Ventura (2010)	2007 yılında yaptıkları tarama çalışmasına 2010 yılına kadar yapılan 225 çalışmayı daha ekleyerek yeni bir inceleme yapılması	306	Elde edilen veri ortamı, EVM görevleri ve veri madenciliği teknikleri	Yapılan analizler sonucunda EVM araçlarının karmaşık bulunduğu, araçların basitleştirilmesi ile eğitimcilerin kullanmalarının yaygınlaşacağını; daha fazla çalışma yapılacağını ve yapılan her çalışma ile alanın gelişeceğini belirtmişlerdir.
Mohamad ve Tasir (2013)	2004 ile 2012 yılları arasında e-öğrenme ortamlarında eğitsel veri madenciliği ile ilgili yayınlanan son 9 çalışmanın incelenmesi	9	Veri madenciliği model ve yöntem, alıntılanma sayısı, yayının kaynağı, çalışmanın platformu	Sonuç olarak e-öğrenme ortamlarına önem gösterildiği ve artık platformların kendi araçlarını geliştirdiğini belirtmiş; gelecek çalışmalarda odağı sosyal ağ platformlarına yönelerek log kayıtlarının incelenebileceği önerisinde bulunmuşlardır.
Pena-Ayala (2014)	2010 ile 2013 yılının 1. çeyreğinde yayınlanan 240 çalışmanın incelenmesi	240	Eğitsel veri madenciliği görevi, veri madenciliği model, yöntem ve teknik, elde edilen veri ortamı	Yapılan analizler sonucunda EVM araçlarının karmaşık bulunduğu, araçların basitleştirilmesi ile eğitimcilerin kullanmalarının yaygınlaşacağını; daha fazla çalışma yapılacağını ve yapılan her çalışma ile alanın gelişeceğini belirtmişlerdir.
Papamitsiou ve Economides (2014)	2008-2013 yılları arasında öğrenme analitikleri ve eğitsel veri madenciliği alanlarında yayınlanan çalışmaların incelenmesi	40	Araştırma stratejisi, Araştırma disiplini, Öğrenme ortamı, EVM görevleri, Veri toplama ve analiz tekniği	Yapılan çalışma sonucunda bu çalışmaların büyük hacimde eğitim verilerinin kullanılabilmesi, kişiselleştirilmiş sistemler için model önerileri, kritik anları ve öğrenme kalıplarını ortaya çıkarması ve öğrenme stratejileri ve davranışları hakkında bilgi vermesi güçlü yönleri; çoğunlukla niceliksel araştırma sonuçlarının olması ve farklı araçlar ve uygulamalar için veri standardizasyonunun olmaması zayıf yönleri olarak belirtilmiştir.

Çizelge 2.1. (Devam) EVM konulu alanyazın inceleme çalışmaları

Yazarlar	Amaç	Çalışma sayısı	Değişkenler	Sonuç
Tekin ve Öztekin (2016)	2006-2016 yılları arasında yayınlanmış eğitsel veri madenciliği konulu çalışmaların incelenmesi	140	Yerli-yabancı yazar, yerli-yabancı dergi, yayın yılı, veri madenciliği modelleme yöntemi, araştırma konusu, veri türü, çalışma grubu, veri toplama aracı, veri analiz yöntemi	İnceleme sonuçlarına göre yayınlanan çalışmaların çoğunluğunun çalışma grubunun üniversite öğrencisi olduğu; veri toplama aracı olarak web ortamlarının daha çok tercih edildiği; nicel verilerin nitel verilere göre daha çok tercih edildiği ve verilerin daha çok kümeleme ve sınıflandırma yöntemleri ile analiz edildiğini belirtmişlerdir.
Dutt, Ismail ve Herawan (2017)	1983-2016 yılları arasında eğitsel veri madenciliği alanında kümeleme yöntemi ile ilgili yayınlanan çalışmaların incelenmesi	166	Amaç, eğitsel veri madenciliği tekniği, elde edilen veri ortamı, alıntı sayısı	Çalışma sonucuna göre kümeleme yöntemi öğrenme için gereken zaman, sınıfta öğrenen davranışlarının farklı değişkenler açısından incelenmesi ve öğrenmeye yönelik motivasyona ilişkin çalışmalarda geçerli ve güvenilir sonuçlar elde ettiğini belirtmişlerdir.
Manjarres, Sandoval ve Suárez (2018)	1993- 2015 yılları arasında yayınlanan toplamda 305 yazarın katıldığı ve 12 veri madenciliği tekniğinin tanımlandığı çalışmaların incelenmesi	127	Eğitsel veri madenciliği görevi, kullanılan veri madenciliği tekniği	Çalışmaların daha çok sanal öğrenme ortamlarında öğrenmeyi etkileyen faktörleri araştırmaya odaklandığı; veri madenciliği tekniği olarak en çok birliktelik kuralı kullanıldığı sonucuna ulaşmışlardır.
Widyahastuti ve Tjhin (2018)	Veri madenciliği teknikleri kullanılarak çevrim içi tartışma forumlarında performans tahmini ile ilgili 2011-2016 yılları arasında yayınlanan çalışmaların incelenmesi	22	Öğrenci performansını tahmin etme yöntemi, kullanılan veri madenciliği yöntemleri	Analizler sonucunda veri madenciliğinde kullanılan sınıflandırma yöntemlerinde Naive Bayes ve Karar Ağaçları tekniklerinin öğrenci performansını tahmin etmede en iyi teknik olduğu sonucuna ulaşmışlardır.

3. MATERYAL VE YÖNTEM

3.1.Araştırma Modeli

Bu tez çalışması nitel araştırma deseni kullanılarak yapılandırılmıştır. Nitel araştırmalar doğal olarak oluşan olguları ortamlar ve koşullar içerisinde incelemek üzerine yoğunlaşırlar. Gürbüz ve Şahin (2016)' e göre nitel araştırmanın tek bir tanımını yapmak güçtür. Hem farklı disiplinlerde ve bu disiplinde hâkim araştırma deseni ve analiz tekniklerinden yararlanılması, hem de her bir nitel araştırmanın kendine özgü bağlamı olması, ortak bir tanımın yapılmasını güçleştirmektedir (Gürbüz ve Şahin, 2016). Yıldırım ve Şimşek (2016), bunun nedenini “nitel araştırma” kavramının bir şemsiye kavram olarak kullanılmasından ve bu şemsiye altında alabilecek birçok kavramın değişik disiplinlerle yakından ilişkili olmasından kaynaklandığını söylemektedir. “Kültür analizi”, “antropoloji”, “durumsal araştırma”, “yorumlayıcı araştırma”, “eylem araştırması”, “doğal araştırma”, “betimsel araştırma”, “kuram geliştirme”, “içerik analizi” bu kavramlardan sadece birkaç tanesidir. Tüm bu kavramlar araştırma deseni ve analiz teknikleri açılarından birbirine benzer yapılara sahip olduğu için, “nitel araştırma”, bu kavramları içine alan genel bir kavram olarak kabul edilebilir (Yıldırım ve Şimşek, 2016). Her ne kadar tüm bu yönelimleri, yöntemleri, süreçleri ve özellikleri kapsayan bir tanım yapmak güç ise de “nitel araştırma”, gözlem, görüşme ve doküman analizi gibi nitel veri toplama yöntemlerinin kullanıldığı, algıların ve olayların doğal ortamda gerçekçi ve bütüncül bir biçimde ortaya konmasına yönelik nitel bir sürecin izlendiği araştırma olarak tanımlanabilir (Yıldırım ve Şimşek, 2016).

Bu tez çalışmasında veri toplama yöntemi olarak doküman incelemesi kullanılmıştır. Dokümanlar, nitel araştırmalarda etkili bir şekilde kullanılması gereken önemli bilgi kaynaklarıdır. Doküman incelemesi, araştırılması hedeflenen olgu ve olaylar hakkında bilgi içeren yazılı materyallerin analizini kapsar (Yıldırım ve Şimşek, 2016). Doküman incelemesi ile elde edilen verilerin araştırma sorularının ortaya koyduğu temalara göre özetlenmesi ve yorumlanması için nitel araştırma türlerinden betimsel analiz kullanılmıştır.

3.2. Veri Toplama

Çalışmada incelenecek makalelerin seçilmesi için, Hatay Mustafa Kemal Üniversitesi Kütüphanesi'nde erişim sağlanan ve konu ile ilişkili makaleleri içerdiği düşünülen şu veri tabanları taranmıştır:

- Ebscohost
- Business Source Complete
- Dergipark
- Directory of Open Access Journals
- Education Source
- Emerald Insight
- Eric
- Ideal Online
- IEEE Xplore Digital Library
- JSTOR Journals
- Sciencedirect
- Scopus
- Social Sciences Citation Index
- Springer Nature Journals
- Springer New York

Ulusal ve uluslararası alanyazında bilimsel makaleler yayın tarihi ve anahtar kelime sınırlandırmaları ile aranmıştır. Yayın tarihi 1 Ocak 2013 ile 31 Aralık 2018 arası sınırlandırılıp anahtar kelime olarak “Educational Data Mining”, “Eğitsel Veri Madenciliği” ve “Eğitimsel Veri Madenciliği” ifadeleri kullanılmıştır. Bu tarama işlemi ile 183 yayına erişim sağlanmıştır. Ulaşılan yayınlardan alanyazın incelemesi türünde veya nitel çalışmalar kapsam dışı bırakılmış; tam metin erişimi olan ve araştırma sorularına uygun olan 122 makale çalışmaya dâhil edilmiştir.

3.2.1. Makale İnceleme Ölçütleri ve Değerleri

Alanyazın taraması sonucu elde edilen çalışmaların özelliklerini sistemli bir şekilde bir araya getirmek için makale inceleme ölçütleri oluşturulmuştur. Oluşturulan

makale inceleme ölçütleri ve değerleri Çizelge 3.1’de verilmiştir. Bazı yayınlara birden fazla veri tabanı üzerinden erişim sağlanmaktadır, fakat yayınların listede ilk sırada çıkan veri tabanı çalışmaya dâhil edilmiştir.

Çizelge 3.1. Makale inceleme ölçütleri ve değerleri

Ölçütler	Değerleri
Çalışma numarası	
Çalışma adı	
Çalışma yılı	(1) 2013 (2) 2014 (3) 2015 (4) 2016 (5) 2017 (6) 2018
Çalışma veri tabanı	(1) Ebscohost (2) Business Source Complete (3) Dergipark (4) Directory of Open Access Journals (5) Education Source (6) Emerald Insight (7) Eric (8) Ideal Online (9) IEEE Xplore Digital Library (10) JSTOR Journals (11) Sciencedirect (12) Scopus (13) Social Sciences Citation Index (14) Springer Nature Journals (15) Springer New York
Google Scholar alıntılanma sayısı	
Scopus alıntılanma sayısı	
Çalışma yazar bilgileri	
Anahtar kelimeler	
Araştırmanın konusu	(1) Verilerin analizi ve görselleştirilmesi (2) Kurs/öğretmen/yöneticinin iyileştirme yapabilmesi için geri bildirim sağlama (3) Bireyselleştirilmiş öğrenme için öneriler (4) Öğrenci performansını tahmin etme (5) Öğrenci modellemesi (6) İstenmeyen öğrenci davranışlarını belirleme (7) Öğrenci gruplandırma (8) Sosyal ağ analizi (9) Planlama ve zamanlama (10) Kavram haritası geliştirme (11) Kurs içeriği ve eğitim yazılımı yapılandırma
Örneklemin boyutu	(1) 20-100 kişi (2) 100-300 kişi (3) 300-1000 kişi (4) 1000-120000 kişi (5) Belirtilmemiş
Örneklem düzeyi	(1) İlköğretim öğrencisi (2) Ortaokul öğrencisi (3) Lise öğrencisi (4) Üniversite öğrencisi (5) Lisansüstü öğrenci (6) Üniversite mezunu (7) Şirket çalışanı (8) Üniversite öğrencisi ve akademisyenler (9) Çevrim içi kurs (10) Derse ait kurs ve cevaplar (11) Belirtilmemiş

Çizelge 3.1. (Devam) Makale inceleme ölçütleri ve değerleri

Ölçütler	Değerleri
Elde edilen veri ortamı	(1) Geleneksel eğitim (2) Bilgisayar tabanlı eğitim
Veri ortamı bilgisayar tabanlı eğitim ise	(1) Web tabanlı öğrenme (2) Öğrenme yönetim sistemleri (3) Akıllı özel ders sistemleri (4) Uyarlanabilir eğitim sistemleri (5) Testler / anketler (6) Metinler / içindekiler
Kullanılan veri toplama aracı	(1) Öğrenme yönetim sistemi (2) Web ortamı (3) Öğrenci bilgi sistemi verileri (4) Anket (5) Simülasyon verileri (6) Başarı testi (7) Metin (8) Ölçek (9) Diğer (10) Belirtilmemiş
Veri analiz yöntemi	(1) Sınıflama (2) Regresyon (3) Kümeleme (4) İlişki madenciliği (5) Birliktelik kuralı (6) Sıralı örüntü (7) Belirtilmemiş
Veri analiz metodu	(1) Birliktelik Kuralı (2) Bölümlemeli Kümeleme (3) Bulanık Mantık (4) Doğrusal Regresyon (5) Fp Growth Algoritması (6) Genetik Algoritmalar (7) Hiyerarşik Kümeleme (8) İstatistiksel Analiz (9) Karar Ağacı (10) Karar Destek Makineleri (11) K-En Yakın Komşu (12) Korelasyon (13) K-Ortalama (14) Lojistik Regresyon (15) Makine Öğrenmesi (16) Metin Madenciliği (17) Naive Bayes Sınıflandırıcısı (18) Öz düzenleyici Haritalar (19) Pazar Sepeti Analizi (20) Regresyon (21) Yapay Sinir Ağı (22) Belirtilmemiş
Veri analizi için kullanılan yazılım	(1) Coh-Matrix (2) Disco (3) Gephi (4) Google Analitik (5) KnowEdu (6) Matlab (7) Microsoft Excel (8) Rapidminer (9) Spss (10) Spy Us (11) Weka (12) Belirtilmemiş
Sonuçlar	

3.3. Veri Analizi

Bu çalışmada 2013-2018 yılları arasındaki eğitsel veri madenciliği alanında yayınlanmış bilimsel makalelerden elde edilen verilerin analizinde doküman analizi tekniklerinden betimsel analiz tekniği kullanılmıştır. Yıldırım ve Şimşek (2016)'e göre; betimsel analiz önceden belirlenmiş bir çerçeveye bağlı olarak nitel verilerin işlenmesi,

bulguların tanımlanması, tanımlanan bulguların yorumlanması adımlarını içeren analiz yaklaşımıdır.

Betimsel analizde elde edilen verilerin önceden düzenli ve açık bir şekilde betimlenmesi gerekir. Bu betimlemeler açıklanır, yorumlanır, neden-sonuç ilişkileri incelenir ve birtakım sonuçlara ulaşılır. Ortaya çıkan sonuçların ilişkilendirilmesi, anlamlandırılması ve ileriye yönelik tahminlerde bulunulması araştırmanın yapacağı yorumların boyutları arasındadır (Yıldırım ve Şimşek, 2016). Bu tür analizde amaç, elde edilen bulguları düzenlenmiş ve yorumlanmış bir biçimde okuyucuya sunmaktır. Bu amaçla elde edilen veriler, önce sistematik ve açık bir biçimde betimlenir. Daha sonra yapılan bu betimlemeler açıklanır ve yorumlanır, neden-sonuç ilişkileri irdelenir ve birtakım sonuçlara ulaşılır (Yıldırım ve Şimşek, 2016).

Yıldırım ve Şimşek'e göre betimsel analiz dört aşamadan oluşur:

1. Betimsel analiz için bir çerçeve oluşturulması: veri kaybı ve yanlış veri düzenlemesine neden olmamak için araştırmanın boyutlarından yola çıkarak veri analizi için bir çerçeve oluşturulur.
2. Tematik çerçeveye göre verilerin işlenmesi: veriler incelenerek çözümlemeler yapılır. Çözümlemeler sonrasında çerçeveye ilişkili olmayan veriler dışarıda bırakılır ve çerçeveye göre elde edilen veriler tanımlanmak için anlamlı ve mantıklı bir şekilde bir araya getirilir.
3. Bulguların tanımlanması: bu aşamada çerçeveye göre süzülen veriler kolay ve anlaşılır bir dil ile tanımlanır.
4. Bulguların yorumlanması: bu aşamada araştırmacı tanımlanan verileri niteliksel ve nesnel olarak yorumlar.

Betimsel analiz aşamalarından ilki için makaleleri araştırma sorularına göre belirlenen değişkenler açısından incelemek amacıyla daha önce Romero ve Ventura (2007; 2010), Baker ve Yasef (2009) ile Pena-Ayala (2014)'nın alanyazın incelemesi çalışmasındaki değişkenlerden yararlanılarak araştırmacı tarafından "Makale İnceleme Ölçütleri ve Değerleri" hazırlanmıştır (Çizelge 3.1). Bu ölçütlere ve değerlere göre makalelerin verilerinin girilmesi için Ms Excel'de bir çizelge hazırlanmıştır.

Betimsel analiz aşamalarından ikincisi için, her bir makalenin özellikleri düzenli bir biçimde MS Excel'de bir önceki aşamada oluşturulmuş olan çizelgeye girilmiştir.

Betimsel analiz aşamalarından üçüncüsü için, bir önceki aşamada doldurulmuş

olan çizelge Microsoft Excel Programı ile analiz edilmiştir. Veri çözümlemesinde betimsel istatistikte sık kullanılan “frekans” kullanılmıştır. Veri çözümlemesinden elde edilen verileri betimlemek amacıyla çizelgeler ve grafiklerden yararlanılmıştır.

Betimsel analizin sonuncu aşamasında ise ortaya çıkan yüzde ve frekanslar ışığında sonuçlar alanyazınla karşılaştırılarak yorumlanmıştır.

Kısaca özetlemek gerekirse bu çalışmada 2013-2018 yılları arasında “Eğitsel Veri Madenciliği” alanında yayınlanan bilimsel çalışmalar; yayın yılı, veri tabanı, yazar bilgileri, anahtar kelimeleri, atıf sayıları, araştırma konuları, çalışma grubu, çalışma grubu boyutu, veri toplama ortamı, veri toplama aracı, veri madenciliği modelleme yöntemi, veri analiz yöntemi, veri analiz metodu, veri analizi için kullanılan yazılım ve araştırma sonuçlarındaki güncel eğilimleri ortaya çıkartmak amacıyla incelenmiştir.

4. ARAŞTIRMA BULGULARI VE TARTIŞMA

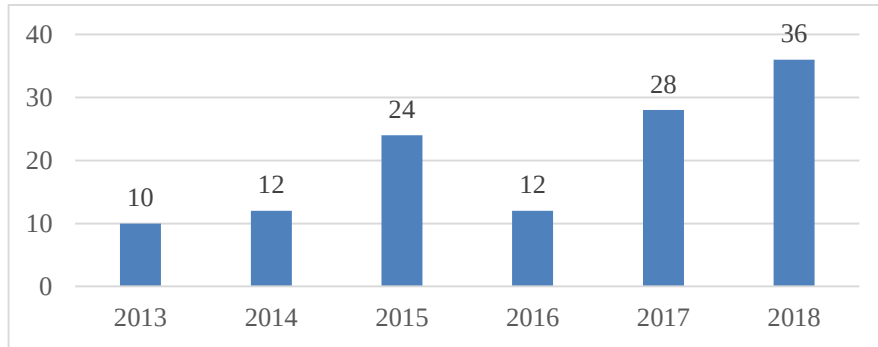
Bu çalışmada 2013-2018 yılları arasında Hatay Mustafa Kemal Üniversitesi Kütüphanesi veri tabanından erişim sağlanan Ebscohost, Business Source Complete, Dergipark, Directory of Open Access Journals, Education Source, Emerald Insight, Eric, Ideal Online, IEEE Xplore Digital Library, JSTOR Journals, Sciencedirect, Scopus, Social Sciences Citation Index, Springer Nature Journals ve Springer New York indekslerinde eğitsel veri madenciliği konusunda yayınlanmış çalışmalar üzerinde alanyazın incelemesi yapılmıştır. Araştırmanın alt problemine göre betimsel analizlerle elde edilen bulgular sistematik ve anlamlı bir şekilde yansıtılmaya çalışılmıştır. Bulgular verildikten sonra, daha önce yapılmış araştırmalar dikkate alınarak tartışılmıştır. Alanyazınla örtüşen ya da farklılık gösteren durumlar belirtilerek yorumlar yapılmıştır.

4.1. Birinci Araştırma Sorusu: Yayınların Demografik Özellikleri

2013-2018 yılları arasında yayımlanan; eğitsel veri madenciliği kullanan yayınların demografik özellikleri kapsamında çalışmaların yayın yılı, erişildiği veri tabanı, yazar(lar), alıntılanma sayısı, anahtar kelime ve araştırma konusu dağılımına ilişkin bulgular bu bölümde yer almaktadır.

Çalışmaların yayın yılı dağılımı

Çalışma kapsamında incelenen çalışmaların yayın yılları Şekil 4.1.'de verilmiştir.



Şekil 4.1. Çalışmaların yayın yılı dağılımı

Çalışma kapsamında incelenen yayınların basım yılları 2013 ile 2018 yılları arasındadır. Yayınların yıllara göre dağılımı incelendiğinde çalışma sayısı özellikle son iki yılda artış eğilimindedir. Eğitsel veri madenciliği alanının önem kazanması ile yayın sayısında artış beklenirken 2016 yılında 2015 yılına göre daha az çalışma yapılmış olması dikkat çekmektedir (Şekil 4.1). Öte yandan Tekin ve Öztekin (2016) 2006 ile 2016 yılları arasındaki eğitsel veri madenciliğine yönelik çalışmaları incelediklerinde çalışma sayısının artış gösterdiğini ve 2016 yılının ilk sırada yer aldığını belirtmişlerdir. Çalışmamızda en çok çalışmanın seçildiği Scopus ve Sciencedirect indeksleri ve diğer farklı 11 indeks, Tekin ve Öztekin (2016)'nın çalışmasında yer almamıştır. Bu çalışmada aksi bir durumun ortaya çıkması taranan veri tabanlarının farklı olmasından kaynaklanmış olabilir.

Çalışmaların Anahtar Kelime Dağılımı

Anahtar kelimeler, yapılmış olan çalışmanın içeriğini araştırmacılara yansıtır ve yapacakları çalışmalar için birçok yayın içinde kendi çalışması ile ilgili olanlara ulaşmasına imkân sunar. Yayınlarda bütünlüğü korumak için anahtar kelimeler İngilizce dilinde incelenmiştir.

Kullanılan anahtar kelimelerin örneklem, eğitim-öğretim, teknoloji, konu/alan/disiplin, öğrenen özellikleri ve veri madenciliği ile ilgili durumların dikkate alınarak seçildiği görülmüştür. Yapılan analizler sonucunda 122 çalışmada 351 farklı anahtar kelime kullanıldığı belirlenmiştir. Anahtar kelimeler betimsel analiz kullanılarak 16 kategoriye ayrılmıştır. Anahtar kelime kategorileri aşağıda verilmiştir.

- Eğitim-Öğretim ile İlgili Kavramlar: Bu kategoriye eğitim- öğretim kavramları ile ilgili anahtar kelimeler dâhil edilmiştir.
- E-Öğrenme Ortamı: Bu kategoriye eğitim sürecinde kullanılan öğrenme ortamı ile ilgili anahtar kelimeler dâhil edilmiştir.
- Konu/ Alan/ Disiplin: Bu kategoriye ders, disiplin, meslek ve uzmanlık eğitimi ile ilgili anahtar kelimeler dâhil edilmiştir.
- Merkezi Öğrenci Değerlendirme Sınavı: Bu kategoriye ulusal ve uluslararası düzenlenen sınavlar ile ilgili anahtar kelimeler dâhil edilmiştir.
- Öğrenci Atipik Davranışları: Bu kategoriye olumsuz öğrenci davranışları ile ilgili

anahtar kelimeler dâhil edilmiştir. Romero ve Ventura (2010) eğitsel veri madenciliği yayınlarını incelediği çalışmada eğitsel veri madenciliğinin görevi olarak belirttiği olumsuz öğrenci davranışlarını tespit etme ile ilgili anahtar kelimeler bu kategoriye dâhil edilmiştir.

- Öğrenci Performansı: Bu kategoriye öğrenci akademik performansı ile ilgili anahtar kelimeler dâhil edilmiştir. Romero ve Ventura (2013) eğitimde veri madenciliği konulu çalışmada eğitsel veri madenciliğinin en eski ve en popüler görevi olarak belirttiği öğrenci performansını tahmin etme ile ilgili anahtar kelimeler bu kategoriye dâhil edilmiştir.
- Öğrenci Stil/ Beceri/ Davranış: Bu kategoriye öğrenci bilişsel/ öğrenme stil, beceri, tutum ve davranış özellikleri ile ilgili anahtar kelimeler dâhil edilmiştir.
- Öğrenme Kuram ve Modelleri: Bu kategoriye öğrenmenin nasıl gerçekleştiği ile ilgili kuram ve modellere dair anahtar kelimeler dâhil edilmiştir.
- Örneklem Düzeyi: Bu kategoriye çalışmalarda kullanılan örneklem grubu ile ilgili anahtar kelimeler dâhil edilmiştir.
- Teknoloji: Bu kategoriye çalışmada kullanılan teknolojik ürün, yazılım ve sistem ile ilgili anahtar kelimeler dâhil edilmiştir.
- Teori/ Model/ Süreç/ Teknik: Bu kategoriye veri analizi ile ilgili olan teori/ model/ süreç/ teknik ile ilgili anahtar kelimeler dâhil edilmiştir.
- Veri Analiz Yazılımı: Bu kategoriye veri analizinde kullanılan yazılımlar ile ilgili anahtar kelimeler dâhil edilmiştir.
- Veri ile İlgili Kavramlar: Bu kategoriye veri madenciliği kavramları ile ilgili anahtar kelimeler dâhil edilmiştir.
- Veri Madenciliği Görevi: Bu kategoriye veri madenciliğinin çalışmadaki görev durumu ile ilgili anahtar kelimeler dâhil edilmiştir.
- Veri Analizi: Bu kategoriye veri madenciliği tahmin ve tanımlayıcı modelleme yöntemleri, teknikleri ve algoritmaları ile ilgili anahtar kelimeler dâhil edilmiştir.
- Veri Türleri: Bu kategoriye elde edilen veri türleri ile ilgili anahtar kelimeler dâhil edilmiştir.

Kategoriler ve anahtar kelime sayısı dağılımı Çizelge 4.1’de gösterilmiştir. Makalelerin büyük bir kısmında veri ile ilgili kavramlar (141), veri analizi (84) ve eğitim-öğretim ile ilgili kavramlar (69) kategorilerinde anahtar kelimelere yer verildiği tespit

edilmiştir. Bazı anahtar kelimeler birkez kullanılmasına karşın, bazıları birden fazla kullanılmıştır. Eğitsel veri madenciliği konulu yayınlara ulaşmak için kullanılan “Educational data mining” (122) ifadesi veri ile ilgili kavramlar kategorisinde yer almıştır.

Çizelge 4.1. Kategoriler ve anahtar kelime sayısı dağılımı

Anahtar Kelime Kategorisi	Anahtar Kelime Sayısı
Veri ile İlgili Kavramlar	141
Veri Analizi	84
Eğitim- Öğretim ile İlgili Kavramlar	69
E-Öğrenme Ortamı	46
Veri Türleri	39
Öğrenci Performansı	35
Öğrenme Kuram ve Modelleri	33
Veri Madenciliği Görevi	33
Teori/ Model/ Süreç/ Teknik	24
Konu/ Alan/ Disiplin	23
Öğrenci Stil/ Beceri/ Davranış	20
Teknoloji	17
Öğrenci Atipik Davranışları	16
Örnekleme	11
Veri Analiz Yazılımı	5
Merkezi Öğrenci Değerlendirme Sınavı	2

Analizlerde “Educational data mining” ifadesinden sonra en fazla kullanılan anahtar kelimeler “Learning Analytics (Öğrenme Analitikleri)” (24), “Cluster Analysis (Kümeleme Analizi)” (14), ve “Learning Management System (Öğrenme Yönetim Sistemi)” (13) olduğu görülmüştür. Agudo- Peregrina ve ark. (2013) eğitsel veri madenciliği ile ilişkilendirilen öğrenme analitiklerinin, öğretmenlerin ve öğrencilerin öğrenmeyi geliştirmeleri için ihtiyaç duydukları eylemin zekâ durumları, verilerin yorumlanması ve bağlamsallaştırılmasını içerdiğini belirtmişlerdir. Öğrenme analitikleri öğrencilerin çevrim içi öğrenme aktivitelerindeki davranışlarını analiz etmek ve ÖYS'deki spesifik etkileşim modellerini tespit etmek, EVM ve ÖA araştırma toplulukları için büyük ilgi çeken bir konu olduğunu belirtmişlerdir (Luhanak ve ark., 2017). Veri analiz yöntemi olan “Cluster Analysis (Kümeleme Analizi)” çalışmalarda en çok

kullanılan ikinci veri analiz yöntemi olmuştur.

Çizelge 4.2’de görüldüğü gibi bu çalışmada örneklem ile ilgili altı farklı kelime ve bu kelimelerin toplamda 11 frekansa sahip olduğu tespit edilmiştir. Bunlar arasında “Higher Education (Yüksek Öğrenim)” en çok kullanılan kelimedir. Tekin ve Öztekin (2016) yaptıkları çalışmada en çok üniversite düzeyinde EVM çalışması yapıldığını belirtmiştir. Bu çalışmada da örneklem özellikleri analizi bulgularında çalışmalarda en çok kullanılan örneklem düzeyinin üniversite öğrencisi olduğu sonucuna ulaşılmıştır.

Teknoloji kategorisine dâhil edilen 10 farklı kelime olduğu ve bu kelimelerin toplamda 17 frekansa sahip olduğu tespit edilmiştir (Çizelge 4.2). Bunlar arasında “Internet (İnternet)”, “Artificial Intelligence in Education (Eğitimde Yapay Zekâ)”, “Knowledge Management System (Bilgi Yönetim Sistemi)”, “Social Media (Sosyal Medya)” ve “Stroke Grouping (Kontur Gruplaması)” en çok kullanılan anahtar kelimelerdir. Veri madenciliği yöntemlerinden makine öğrenmesi ile alakalı ve son zamanların popüler teknolojilerinden olan yapay zekânın eğitim alanında da kullanıldığı görülmüştür.

Ulusal ve uluslararası sınavların yer verildiği kategori olan merkezi öğrenci değerlendirme sınavı kategorisine “LYS (Liseye Geçiş Sınavı)” ve “TIMSS (Trends in International Mathematics and Science Study) (Uluslararası Matematik ve Fen Eğilimleri Araştırması)” kelimeleri dâhil edilmiştir. Bu kategori iki farklı anahtar kelime ile en az anahtar kelime bulunan kategori olmuştur.

Çizelge 4.2’de görüldüğü gibi konu/alan/disiplin kategorisinde 14 farklı anahtar kelime olduğu ve toplam frekansı 23 olduğu tespit edilmiştir. Bunlar arasında “Computer Science Education (Bilgisayar Bilimi Eğitimi)”, “Chemical Education (Kimya Eğitimi)” ve “Science Education (Bilim Eğitimi)” en çok tercih edilen alan olmuştur (Çizelge 4.2).

Çizelge 4.2. Örneklem, teknoloji, merkezi öğrenci değerlendirme sınavı ve konu/alan/disiplin kategorileri ve kodları

Örneklem	Teknoloji	Merkezi Öğrenci Değerlendirme Sınavı	Konu/ Alan/ Disiplin
- Higher Education (6) - Adult Education - Elementary Education - General Education - Postsecondary Education - Undergraduate Students	- Internet (4) - Artificial Intelligence in Education (2) - Knowledge Management System (2) - Social Media (2) - Stroke Grouping (2) - Computer Modeling Tools - ICT - Oop - Radio Frequency Identification - Technological Tools	- LYS - TIMSS	- Computer Science Education (6) - Chemical Education (2) - Science Education (2) - Technology Acceptance Model (2) - Accounting - Activity Theory - Ambient Intelligence - Bag of Phonemes - Complex Systems - Engineering Education - Mathematics Education - Medical Education - Pedagogical Issues - Problem Solving

Çizelge 4.3'te görüldüğü gibi e-öğrenme ortamı ile ilgili anahtar kelimelerde 15 farklı anahtar kelimeye ve toplamda 46 frekansa sahip olduğu tespit edilmiştir. “Learning Management Systems (Öğrenme Yönetim Sistemi)”, “Smart e-Learning (Akıllı e-Öğrenme)”, “Massive Open Online Course (Kitleli Açık Çevrim İçi Dersler)”, “Computer-Based Assessment (Bilgisayar Tabanlı Değerlendirme)” ve “Virtual Learning Environment (Sanal Öğrenme Ortamı)” çalışmalarda e-öğrenme kategorisinde en çok yer alan anahtar kelimeler olmuştur. Bu kategoride yer alan “Learning Management System” çalışmalarda belirtilen 351 anahtar kelime içerisinde frekansı en yüksek üçüncü anahtar kelimedir.

Öğrenme kuram ve modelleri kategorisinde yer alan anahtar kelimeler Çizelge 4.3'te gösterilmiştir. Bu kategoride 19 farklı kelime ve toplamda 33 frekansa sahip olduğu tespit edilmiştir. “E-Learning (E-Öğrenme)”, “Blended Learning (Harmanlanmış Öğrenme)” ve “Traditional Educational Environments (Geleneksel Eğitim Ortamları)” bu kategoride en çok kullanılan anahtar kelimelerdir.

Çizelge 4.3'te görüldüğü gibi veri ile ilgili kavram kategorisinde dokuz farklı anahtar kelime ve toplamda 141 frekansa sahip olduğu görülmüştür. EVM konulu yayınları indekslemek için kullanılan “Education Data Mining (Eğitsel Veri Madenciliği)” ve “Eğitsel Veri Madenciliği” anahtar kelimeleri de bu kategoride yer almıştır. Bu kategorinin toplam frekansının yüksek olmasının sebebi yayın indekslemek için kullanılan anahtar kelimenin bu alanda olmasıdır. “Educational Data Mining”, “Data Mining (Veri Madenciliği)”, “Big Data (Büyük Veri)” ve “Knowledge Discovery in Databases (Veri Tabanında Bilgi Keşfi)” kelimeleri bu kategoride en çok kullanılan anahtar kelimelerdir.

Çizelge 4.3. E-öğrenme ortamı, öğrenme kuram ve modelleri ve veri ile ilgili kavramlar kategorileri ve kodları

E-Öğrenme Ortamı	Öğrenme Kuram ve Modelleri	Veri ile İlgili Kavramlar
• Learning Management System (13) • Smart E-Learning (7) • Massive Open Online Course (5) • Computer-Based Assessment (4) • Virtual Learning Enviroment (3) • Interactive Learning Environment (2) • Online Learning (2) • Personal Learning Environment (2) • Simulation (2) • Computer-Assisted Pronunciation Training System • Ecoliteracy • Maritime E-Learning • Security • PDCA Cycle • Tutoring Action Plan	• E-Learning (9) • Blended Learning (4) • Traditional Educational Environments (4) • Bloom's Taxonomy • Cognitive Tutor • Collaborative Learning • Computer Supported Collaborative Learning • Constructivism • Educational Games • Evidence-Based Education • Game-Theory • Gamification • Lifelong Learning • Meta-Learning • Multiple İntelligences • Problem-Based Learning • Social Learning • Students Activity Model • Technology Supported Learning	• Educational Data Mining (122) • Data Minig (9) • Big Data (3) • Knowledge Discovery in Databases (2) • Change Mining • Data İntegration • Data İntegration • Data Mining For Small Data Set • High-Dimensional Data Classification

Çizelge 4.4'te gösterildiği gibi veri analiz ile ilgili model, yöntem ve algoritmaların yer aldığı veri analizi kategorisi toplamda 84 frekans ile en çok anahtar kelime içeren ikinci kategoridir. Bu kategoride 18 farklı anahtar kelime yer almıştır. Çalışmalarda genel olarak veri analizinde kullanılan model, yöntem veya algoritmanın belirtildiği görülmüştür; bu durumun çalışmanın içeriğini yansıtmaya açısından araştırmacılara yardımcı olacağı düşünülmektedir. “Cluster Analysis (Kümeleme Analizi)”, “Classification (Sınıflandırma)”, “Prediction (Tahmin)”, “Association Rule Mining (Birliktelik Kuralı Madenciliği)” ve “Decision Support (Karar Destek Sistemi)” en çok kullanılan anahtar kelimelerdir. Veri analiz yöntemi olan “Cluster Analysis” veri analizinde en çok kullanılan ikinci yöntem olduğu sonucuna ulaşmamıza rağmen, ilk yöntem olan “Classification”ne göre daha fazla kullanılmıştır. Bunun sebebinin çoğu yayında sınıflandırma yöntemi kullanılmasına rağmen sınıflandırma yöntemini dâhil olduğu modelleme yöntemi olan “Prediction” olarak yer verilmesi olduğu düşünülmektedir. Veri analiz kategorisinde yer alan anahtar kelimelerin çoğunluğunun tahmin modeli ile ilişkili yöntem, teknik ve algoritma olduğu görülmüştür. Bu kategorideki kelimelerin birçoğunun çalışmalarda tekrar edildiği tespit edilmiştir.

Veri analizinde kullanılan yazılımları içeren veri analiz programı kategorisinde 4 farklı anahtar kelime dâhil edilmiştir (Çizelge 4.4). Bu kategori toplamda beş frekansa sahip olduğu tespit edilmiştir. Bu kategoride en çok kullanılan anahtar kelime “Weka” olmuştur. Yaptığımız veri analizinde kullanılan program sıklık analizinde de Weka programının veri analizinde en fazla kullanılan yazılım olduğu sonucuna ulaşılmıştır. Çalışmalarda veri analizinde kullanılan yazılımla ilgili bilginin anahtar kelime olarak yer verilmediği gözlemlenmiştir.

Çizelge 4.4'te görüldüğü gibi çalışmada veri türleri kategorisinin 26 farklı anahtar kelime ve toplamda 39 frekansa sahip olduğu görülmüştür. “Online Behavior (Çevrim İçi Davranış)”, “Search Patterns (Arama Kalıpları)”, “Computer-Based Testing (Bilgisayar Tabanlı Test)” ve “Log (Günlük)” en çok kullanılan anahtar kelimelerdir. Eğitsel veri madenciliği çalışmalarında öğrenci davranışları, ödünç kitap kaydı, sayfa görüntüleme, çevrim içi tartışma, proxy bilgileri, ziyaret saati, görüntüleme, arama kalıpları, çevrim içi ortam öğrenci günlükleri ve öğrenci cevapları gibi veri türleri veri analizinde kullanılmıştır.

Çizelge 4.4. Veri türleri, veri analiz yazılımı ve veri analizi kategorileri ve kodları

Veri Türleri	Veri Analiz Yazılımı	Veri Analizi
• Online Behavior (4) • Search Patterns (3) • Computer-Based Testing (3) • Log (3) • Laboratory Dataset (2) • Longitudinal Student Data (2) • Page View (2) • Online Discussion (2) • Work Patterns • Book-Borrowing Record • Clinical Reasoning • Cross Section • Cross Table • Frequency Domain • Imbalanced Data • Intelligent Activity • Live Video Streaming • Online Interaction • Pedagogical Data • Proxy Variable • Student's Answers • Sketch Understanding • Temporal Data • Unix Command History • Visit Time • Visualization	• Weka (2) • Clickstream Analytics • Discrete Fourier Transforms • Rapid Miner	• Cluster Analysis (14) • Classification (12) • Prediction (11) • Association Rule Mining (8) • Decision Support (8) • Genetic Programming (5) • Naive Bayes (5) • Machine Learning (4) • Particle Swarm Classification (3) • Neural Network (3) • Time Series (3) • Regression (2) • Dragonfly Algorithm • Frequent Pattern Analysis • Fuzzy Representations • Market Basket Analysis • Predictive Apriori Algorithm • Support Vector Machine

Çizelge 4.5'te görüldüğü gibi, bu çalışmada veri madenciliği analizi ve sürecinde kullanılan teori/model/süreç/teknikkategorisinin 21 farklı anahtar kelime ve toplamda 24 frekansa sahip olduğu tespit edilmiştir. Bunlar arasında “CRISP-DM (Cross-Industry Standard Process Model for Data Mining) (Veri Madenciliği için Standart Süreç Modeli)” ve “Feature Selection (Öznitelik Seçimi)” makalelerde en çok kullanılan anahtar kelimelerdir. Bu kelimeler en çok kullanılan anahtar kelime olmasına rağmen frekanslarının düşük olduğu görülmüştür.

Veri madenciliği görevleri ile ilgili anahtar kelime dağılımı ve frekansı Çizelge 4.5'te verilmiştir. Bu çalışmada veri madenciliği görevleri kategorisinin 20 farklı anahtar kelime ve toplamda 33 frekansa sahip olduğu tespit edilmiştir. “Text mining (Metin Madenciliği)”, “Learner Modeling (Öğrenci Modelleme)” ve “Process Mining (İşlem Madenciliği)” en çok kullanılan anahtar kelimeler olmuştur.

Eğitsel veri madenciliğinin popüler konularından olan öğrenci performansını tahmin etme ile ilgili anahtar kelimeler Çizelge 4.5'te verilmiştir. Öğrenci performansını tahmin etme kategorisinin 11 farklı kelime ve toplam 35 frekansa sahip olduğu tespit edilmiştir. “Performance Prediction (Performans Tahmini)”, “Students’ Academic Performance (Öğrenci Akademik Performansı)”, “Predict Achievement of Student (Öğrenci Performansını Tahmin Etme)”, “Performance Assessment (Performans Değerlendirme)” ve “Student Failure (Öğrenci Hatası)” bu kategoride en çok kullanılan anahtar kelimelerdir.

Çizelge 4.5. Teori/model/süreç/teknik, veri madenciliği görevi ve öğrenci performansı kategorileri ve kodları

Teori/ Model/ Süreç/ Teknik	Veri Madenciliği Görevi	Öğrenci Performansı
• Crisp-Dm (2) • Feature Selection (2) • A/B Testing • Centre Particle Swarm Optimisation • Dynamic Time Warping • Evolutionary Computation • Feature Transformation • Gravitational Models • Inductive Miner • Latent Class Analysis • Map-Reduce Framework • Markov Chain • Model Discovery Algorithms • Partial Least Squares Structural Equation Modeling • Pathfinder Network • Pivot Table • Principles Component Analysis (Pca) • Reference model • Rocchio Algorithm • Synthetic Attributes • Temporal Dimension	• Text Mining (6) • Learner Modeling (5) • Process Mining (3) • Automatic Prediction (2) • Classifying Students (2) • Clustering Students • Designing Causal Explanations • Dialog Analysis • Discussion Mining • Disposition Analytics • Enrolment Prediction • Ensemble Clustering • Information Visualization • Log File Analysis • Predicting Marks • Social Learning Analytics • Student Interaction Analysis • Video Analytics • Word Recommendation • Words Subgroup Discovery	• Performance Prediction (10) • Students' Academic Performance (9) • Predict Achievement of Students (4) • Performance Assessment (3) • Student Failure (3) • Academic Performance Modeling • AGPA • Classification Performance Measures • Inquiry Assessment • Mathematics Achievement • Student's Success Rate

Çizelge 4.6’da görüldüğü gibi, bu çalışmada öğrenci atipik davranış ile ilgili dört farklı anahtar kelime tespit edilmiştir. Anahtar kelimelerin toplam frekansının 16 olduğu görülmüştür. Bu kategoride en çok kullanılan anahtar kelime “Student Dropout (Öğrenci Terki)” olmuştur. Kategoride yer alan kelimelerinin hepsinin birden çok kez kullanıldığı görülmüştür. Öğrenci stil/beceri/davranış ile ilgili anahtar kelime dağılımı Çizelge 4.6’da verilmiştir. Bu kategoride 16 farklı kelimenin yer aldığı ve toplam 20 frekansa sahip olduğu görülmüştür. “Skills (Beceriler)”, “Retention (Alıkoyma)” ve “Teaching Style (Öğretim Stili)” kategorinin en çok kullanılan kelimeleridir.

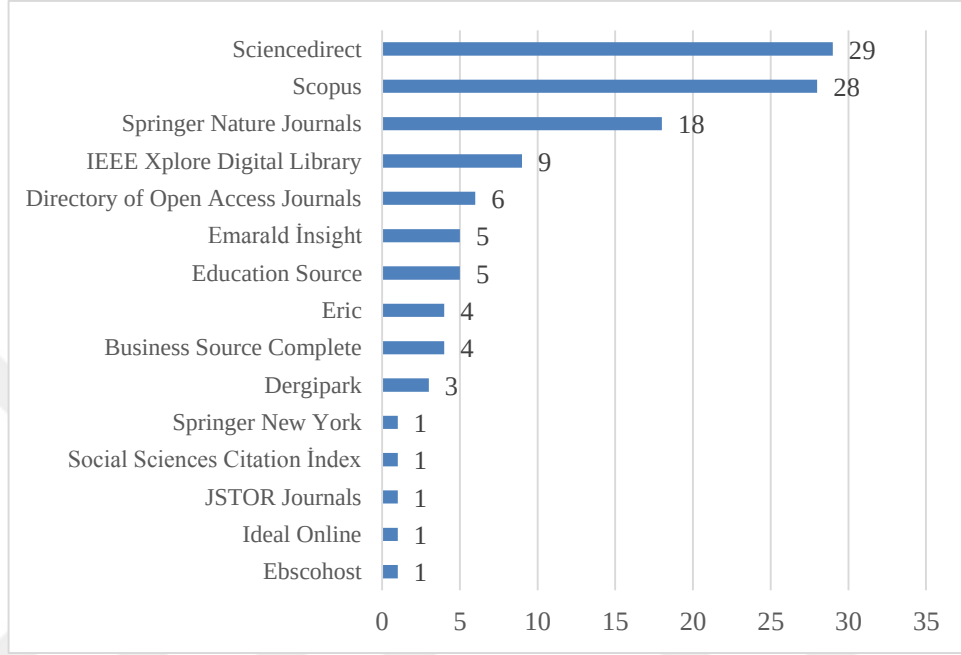
Çizelge 4.6’da çalışmalarda yer alan eğitim-öğretim kavramları ile ilgili anahtar kelime dağılımı gösterilmiştir. Bu kategoride 18 farklı kelime olduğu ve toplam 69 frekansa sahip olduğu tespit edilmiştir. Kategori toplam frekansı bakımından veri ile ilgili kavramlar ve veri analizi kategorisinden sonra üçüncü sıradadır. Bu kategoride “Learning Analytics (Öğrenme Analitikleri)”, “Student Evaluation (Öğrenci Değerlendirme)”, “Educational Technology (Eğitim Teknolojisi)”, “Learning and Teaching Methods (Öğrenme ve Öğretim Metotları)” ve “Learning Outcomes (Öğrenme Çıktıları)” en çok kullanılan anahtar kelimelerdir. “Learning Analytic” çalışmalarda en çok kullanılan anahtar kelimedir.

Çizelge 4.6. Öğrenci atipik davranış, öğrenci stil/beceri/davranış ve eğitim-öğretim ile ilgili kavramlar

Öğrenci Atipik Davranış	Öğrenci Stil/ Beceri/ Davranış	Eğitim-Öğretim ile İlgili Kavramlar
• Student Dropout (9) • Academic Dishonesty and Plagiarism (3) • Attrition (2) • Procrastination (2)	• Skills (4) • Retention (2) • Teaching Style (2) • Affect • Behavior • Carelessness • Designing causal explanations • Misconceptions • Motivation • Robust Learning • Social Interaction • Student Careers • Students Satisfaction • Students' Learning Path • Time Management Strategy • Types of Participant Behavior	• Learning Analytics (24) • Student Evaluation (5) • Educational Technology (4) • Learning and Teaching Methods (4) • Learning Outcomes (3) • Concept Map (2) • Course Design (2) • Feedback (2) • Learning Concepts (2) • Learning Design (2) • Procedural Training (2) • Automated Essay Scoring • Broader Impacts • Curriculum • Educational Certification • Educational Knowledge Graph • Educational Processes • Educational Relation
		• Index of Learning Styles • Learning Behavior • Learning Communities • Learning History • Learning Technologies • Measurement Models • Multi-Task Learning • P2P • Profile Learning • Student Heterogeneity • Transfer Learning • Video Learning

Çalışmaların Veri Tabanı Dağılımı

Çalışma kapsamında incelenen yayınlara erişim sağlayan veri tabanı Şekil 4.2’de verilmiştir.



Şekil 4.2. Çalışmaların veri tabanı dağılımı

Çalışma kapsamında incelenen yayınlar Hatay Mustafa Kemal Üniversitesi Kütüphane Veri Tabanı’nda “Educational Data Mining” sözcük öbeği ile anahtar kelime dizininde arama sonucu çalışmaya dâhil edilmiştir. Bazı yayınlara birden fazla veri tabanı üzerinden erişim sağlanmaktadır fakat yayınların listede ilk sırada çıkan veri tabanı çalışmaya dâhil edilmiştir. Eğitsel veri madenciliği konulu çalışmaların; Sciencedirect (29) ve Scopus (28) veri tabanlarında diğer veri tabanlarına göre daha fazla indekslendiği görülmektedir (Şekil 4.2).

Tekin ve Öztekin (2016) yaptıkları çalışmada Google Scholar, Eric, Elsevier, Proquest, Ulakbim, Ebschost ve Springer veri tabanlarından erişim sağladıkları 140 makaleyi incelemiştir fakat veri tabanına göre yayın sayısını belirtmemişlerdir. Dutt, Ismail ve Herawan (2017) kümeleme yöntemi ile ilgili yaptıkları çalışmada IEEEExplore, ACM Digital, Journal of Educational Data Mining, ProQuest Education Journals (ABI/Inform Complete, ERIC, ProQuestEducation Journals ve ScienceDirect veri

tabanlarında erişim sağladıkları 166 makaleyi incelemiştir. Makalelerin sayı dağılımı IEEEExplore (58), ACM Digital (38), ScienceDirect (36), Proquest Education Journals (27) ve Journal of Educational Data Mining (7) şeklindedir.

Çalışmaların Yazar Dağılımı

Çalışma kapsamında incelenen çalışmaların yazar dağılımı Çizelge 4.7’de verilmiştir.

Çizelge 4.7. Çalışmaların yazar dağılımı

Yazar	f	Yazar	f
Cristobal Romero	8	María A. Martínez	2
Sebastian Ventura	7	Mariluz Guenaga	2
Carlos Márquez-Vera	3	Meehyun Yoon	2
Gökhan Akçapınar	3	Miguel Sánchez-Santillán	2
Ryan S. Baker	3	Natthakan Iam-On	2
Il-Hyun Jo	3	Nemanja Maček	2
Ahmet Tekin	2	Öyküm Esra Aşkın	2
Alberto Cano	2	Pedro J. Munoz-Merino	2
Anwar Ali Yahya	2	Petar Spalević	2
Bratislav Predić	2	Rebeca Cerezo	2
Charoula Angeli	2	Tal Soffer	2
Cristian Olivaresrodríguez	2	Tali Kahan	2
David Lizcano	2	Tossapon Boongoen	2
Dejan Rančić	2	Yeonjeong Park	2
Dongho Kim	2	Yong Wang	2
José A. Ruipérez- Valiente	2	Gabrijela Dimić	2
Juan A. Lara	2		

Toplam 358 farklı yazarın eğitsel veri madenciliği konusunda yayın yaptığı sonucuna ulaşılmıştır. Çizelge 4.7’de birden fazla çalışma yapan yazarların dağılımı verilmiştir. En fazla yayında ismi geçenlerin İspanya Cordoba Üniversitesi Bilgisayar Bilimleri profesörlerinden Cristobal Romero (8) ve Sebastian Ventura (7) olduğu görülmektedir. Aynı zamanda bu iki yazar 2007 ve 2010 yıllarında yayınladıkları

alanyazın inceleme çalışmaları ile eğitsel veri madenciliği alanındaki gelişmeleri ortaya koyarak alanın gelişmesine ve önem kazanmasına katkı sağlamışlardır.

Çalışmaların Alıntılanma Sayısı Dağılımı

Baker ve Yasef (2009) yaptıkları çalışmada eğitsel veri madenciliği alanının ne durumda olduğunu görebilmek için en çok alıntılanan çalışmalara bakmak gerektiğini belirtmiş; çalışmalarında Romero ve Ventura (2007)'nin çalışmasında bulunan eğitsel veri madenciliğinin ilk yıllarında yayınlanmış en çok alıntılanan çalışmaları analiz etmişlerdir. Bu çalışma kapsamında ise 1 Ekim 2019 tarihi itibarıyla Google Scholar ve Scopus İndeksi incelenmiştir. Her iki indekste de ortak dokuz çalışma bulunurken, sadece birer çalışma farklılık göstermektedir. Atıf sayısı sıralamasında ilk 10'da bulunan çalışmaların uluslararası indekste olduğu görülmektedir.

Çalışma kapsamında incelenen yayınların 1 Ekim 2019 tarihi itibarıyla Google Scholar ve Scopus dizininde en fazla atıf alan 10 makalenin atıf sayısı Çizelge 4.8'de verilmiştir.

Çizelge 4.8. 1 Ekim 2019 tarihi itibarıyla Google Scholar ve Scopus veri tabanında ilk 10 makale

Yayın Adı	Yazar	Google Scholar Atıf Sayısı	Scopus Atıf Sayısı	Yayın Yılı
Web Usage Mining For Predicting Final Marks Of Students That Use Moodle Courses	Cristobal Romero, Pedro G. Espejo, Amelia Zafra, Sebastian Ventura, Jose Raul Romero	241	90	2013
Predicting Student Performance By Using Data Mining Methods For Classification	Dorina Kabakchieva	164	81	2013
Examining Students' Online Interaction In A Live Video Streaming Environment Using Data Mining And Text Mining	Wu He	154	65	2013
Participation-Based Student Final Performance Prediction Model Through Interpretable Genetic Programming: Integrating Learning Analytics, Educational Data Mining And Theory	Xing Wanli, Guo Rui, Petakovic Eva, Goggins Sean	153	69	2015

Çizelge 4.8. (Devam) 1 Ekim 2019 tarihi itibarıyla Google Scholar ve Scopus veri tabanı ilk 10 makale

Yayın Adı	Yazar	Google Scholar Atıf Sayısı	Scopus Atıf Sayısı	Yayın Yılı
Predicting Student Failure At School Using Genetic Programming And Different Data Mining Approaches With High Dimensional And İmbalanced Data	Carlos Márquez-Vera, Alberto Cano, Cristóbal Romero, Sebastián Ventura	142	58	2013
Student Data Mining Solution– Knowledge Management System Related To Higher Education Institutions	Srecko Natek, Moti Zwilling	124	57	2014
A System For Knowledge Discovery In E-Learning Environments Within The European Higher Education Area – Application To Student Data From Open University Of Madrid, Udima	Juan A. Lara, David Lizcano, María A. Martínez, Juan Pazos, Teresa Riera	100	43	2014
Classification And Prediction Based Data Mining Algorithms To Predict Slow Learners In Education Sector	Parneet Kaur, Manpreet Singh, Gurpreet Singh Josanc	91	58	2015
Predicting School Failure And Dropout By Using Data Mining Techniques	Carlos Márquez-Vera, Cristóbal Romero Morales, Sebastián Ventura Soto	88	41	2013
Early Dropout Prediction Using Data Mining: A Case Study With High School Students	Carlos Márquez-Vera, Alberto Cano, Cristobal Romero, Amin Yousef Mohammad Noaman, Habib Mousa Fardoun, Sebastian Ventura	73	55	2016
Evaluating The Effectiveness Of Educational Data Mining Techniques For Early Prediction Of Students' Academic Failure İn Introductory Programming Courses	Evandro B. Costa, Balduino Fonseca, Marcelo Almeida Santana, Fabrisia Ferreira De Araújo, Joilson Rego	70	43	2017

İncelenen indekslerde en çok alıntı yapılan makale, Romero ve ark. (2013)'ün öğrenci performansını tahmin etmek için Moodle sisteminden elde edilen ödev, forum ve ara sınav bilgilerini kendi geliştirdikleri veri madenciliği aracı ile analiz ettikleri çalışma olmuştur.

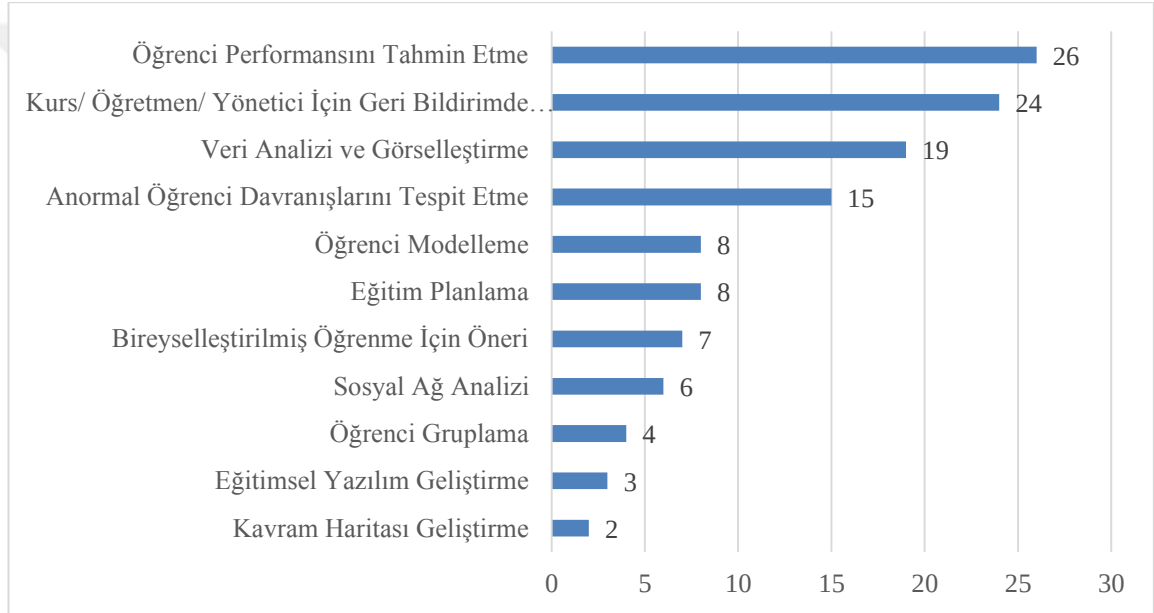
En çok alıntı yapılan ikinci makalede, Kabakchieva (2013) üniversiteler için veri madenciliğinin yüksek potansiyelini ortaya koymak için öğrencinin kişisel ve üniversite öncesi verilerini kullanarak öğrenci performansını tahmin etmeye örnek bir çalışma sunmaktadır.

En çok alıntı yapılan diğer çalışmalarda ise; He (2013)'nin çevrim içi bir kursta paylaşılan sohbet mesajları ve çevrim içi soruları ile öğrencinin kurs başarısı arasında anlamlı bir durumun olup olmadığını göstermeye odaklanmıştır. Wanli ve ark. (2015) bilgisayar sisteminde elde edilen veri analizini otomatikleştirmeye odaklanan tahmin veri madenciliği modellerinin en etkin bütünleştirme potansiyeli olanını ortaya çıkarmaktadır. Marquez-Vera ve ark. (2013) öğrenci başarısızlığını tahmin etmek için yapay zeka alt alanı olan genetik programlama yaklaşımı ile geliştirdikleri algoritma ile 10 farklı sınıflandırma algoritmasının karşılaştırmaktadır. Zwilling ve Natek (2014) yüksek öğretim kurumlarında öğrenci bilgi sisteminden elde edilen veriler ile öğrenci başarısının tahmin edilip edilmeyeceğine ve öğrencinin başarı oranıyla ilişkilendirilebilecek belirli öğrenci özellikleri olup olmadığını ortaya çıkarmaya odaklanan bir çalışma sunmuştur. Lara ve arkadaşları (2014) online eğitim sisteminden ayrılan öğrencileri sınıflandırmak için e-öğrenme ortamları ile etkileşimi sonucu ortaya çıkan verileri analiz ederek öğretmenlerin yararlı bulabilecekleri bilgisini ortaya çıkarmak için veri tabanlarında bilgi keşfi kullanımını önermektedir. Kaur ve ark. (2015) sınıflandırma tabanlı algoritmalar kullanarak tahmin veri madenciliği modeli ile öğrenciler arasında yavaş öğrenenleri belirlemeyi göstermeye odaklanmaktadır. Marquez-Vera ve ark. (2013) okuldaki başarısızlığı ve okuldan ayrılmayı tahmin etmek için veri madenciliği tekniklerini kullanmayı önermektedir. Marquez-Vera ve ark. (2016), öğrencinin okuldan ayrılmasını en erken tahmin eden sınıflandırma algoritması önermektedir. Costa ve ark. (2017) programlama derslerinde başarısız olma ihtimali olan öğrencileri önceden tahmin etmek için eğitimsel veri madenciliği tekniklerinin etkinliği üzerine karşılaştırmalı bir çalışma sunmaktadır.

Çalışmaların alıntılanma sayılarına bakıldığında birçok yazarın bir tane çalışması ilk 10'da yer alırken, Cristobal Romero, Sebastian Ventura, Carlos Márquez-Vera ve Alberto Cano isimli yazarların birden fazla çalışması olduğu görülmektedir. Baker ve Yasef (2009) 1995 ile 2005 yılları arasında yayınlanan çalışmaları alıntı sayısı açısından incelediğinde Romero, Ventura ve ark. (2003)'nin eğitsel veri madenciliğini desteklemek için araçlar sundukları çalışmanın en çok alıntılanma yapılan çalışmalardan olduğunu belirtmişlerdir. Cristobal Romero ve Sebastian Ventura'nın eğitsel veri madenciliği alanına ilk yıllardan itibaren katkı sunan yazarlardan olduğu görülmektedir.

Araştırma Konusu Dağılımı

Çalışmalar daha önceki eğitsel veri madenciliği alanyazın çalışmalarında olduğu gibi alanın eğilimlerini görmek için araştırma konusu bakımından incelenmiştir. Çalışmalar, Romero ve Ventura (2010)'nın araştırma konusu sınıflandırılmasına göre sınıflandırılmıştır. Öğrenci performansını tahmin etme (26) ve eğitimi iyileştirmek için geri bildirimde bulunma (24) konularında diğer alanlara göre daha fazla yoğunlaştığı görülmektedir (Şekil 4.3). Eğitimsel yazılım geliştirme (3) ve kavram haritası geliştirme (2) en az çalışılan araştırma konuları olmuştur.



Şekil 4.3. Çalışmaların araştırma konusu dağılımı

Bu bulguları destekler nitelikte bir çalışma olan Romero ve Ventura (2010), en çok çalışma yapılan alanların öğrenci performanslarını tahmin etme (76) ve geri bildirim sağlama olduğunu (40); en az çalışmanın ise kurs içeriği ve eğitim yazılımının oluşturulması (9) ve kavram haritası geliştirme (10) olduğunu belirtmişlerdir. Çalışmalar en fazla ve en az araştırma yapılan konular açısından benzeşmektedir.

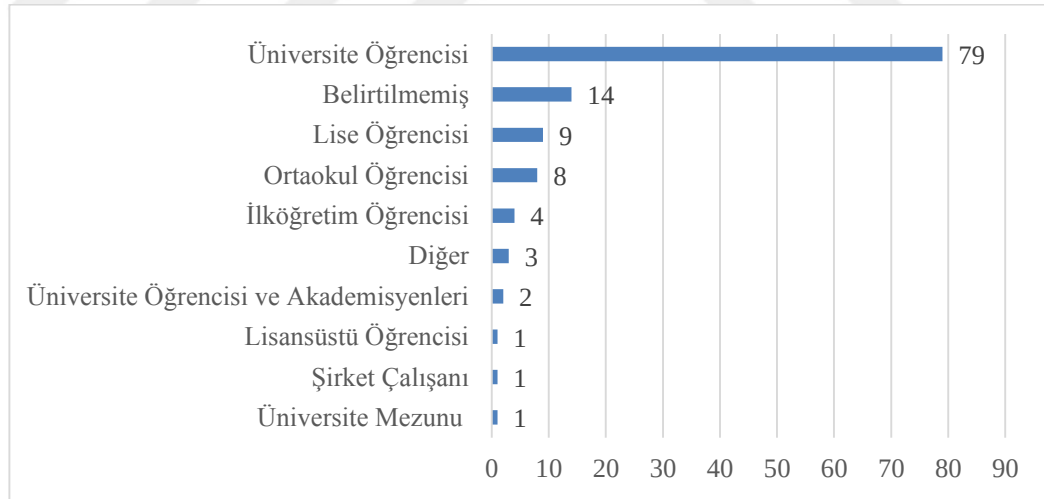
Sin ve Muthu (2015) da yaptıkları alanyazın inceleme çalışmasında makalelerin başlıca eğilimlerinin öğrenci performansını tahmin etme üzerine olduğunu belirtmişlerdir. Widyahastuti ve Tjhin (2018) tarafından yapılan çalışmada alanın eğilimi

olan öğrenci performansını tahmin etme üzerine alanyazın inceleme çalışması yapmışlardır. Çalışmaya sadece çevrim içi artışma forumunda gerçekleştirilen çalışmaları dâhil etmişlerdir. Öğrenci performansını tahmin etmek için 17 çalışmada “mesaj gönder”, yedi çalışmada mesaj okurken harcanan zaman, altı çalışmada ise öğrencinin harcadığı zaman değişkenleri kullanılmıştır. Çevrim içi tartışma mesajları, öğrenme teşhisi için bilgi taşıyabilen tipik dijital verilerdir (Lu ve Jeng, 2006).

4.2. İkinci Araştırma Sorusu: Çalışmaların Örneklem Özellikleri

Örneklem Grubu Dağılımı

Çalışma grubu düzeyine göre yapılan analizler Şekil 4.4’te gösterilmiştir. Elde edilen grafiğe göre en fazla üniversite düzeyinde (79) çalışma yapılmıştır. Üniversite grubunu lise (9), ortaokul (8) ve ilköğretim (4) öğrenci grupları izlemektedir. Çalışma grubu olarak toplamda 21 çalışma k12 öğrenci grubundan elde edilen verilerle yapılmıştır. Çalışmaların 14 tanesinde ise örneklem düzeyi belirtilmemiştir.

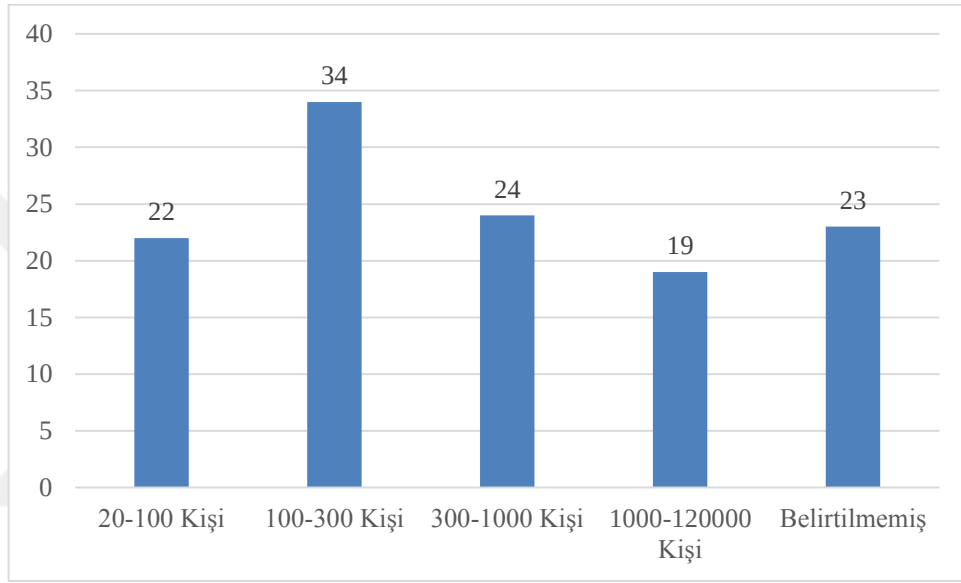


Şekil 4.4. Çalışmaların örneklem grubu dağılımı

İncelenen makalelerde çalışma grubu olarak en çok üniversite öğrencileri (79) tercih edilmiştir. Bu sonuç, Tekin ve Öztekin (2016) tarafından yapılan çalışmada çalışma grubu %68.5 oranla üniversite öğrencileri ile yapılan çalışmaların en yüksek orana sahip olması sonucu ile benzeşmektedir.

Çalışma Grubu Boyutu

Çalışmalar çalışma grubu boyutu açısından 20-100 kişi, 100-300 kişi, 300-1000 kişi ve 1000-120000 kişi aralıklarına göre sınıflandırılmıştır (Şekil 4.5). En fazla tercih edilen çalışma grubu boyutu 100-300 kişi (34) olmuştur. Bunu sırasıyla 300-1000 kişi (24), 20-100 kişi (22) ve 1000-120000 (19) takip etmektedir. Öte yandan 23 çalışmanın çalışma grubu boyutu makalelerde belirtilmemiştir.



Şekil 4.5. Çalışmaların çalışma grubu boyutu

Tekin ve Öztekin (2016) yaptıkları çalışmada, çalışma grubu dağılımının 101-301 kişi (29), 301-1000 kişi (28), 38- 100 kişi (26) ve 1000'den fazla kişi (9) olduğunu belirtmiştir. Çalışmalar çalışma grubu dağılımı bakımından benzeşmektedir. Yüksek boyutlu veri analizinde kullanılan yöntemlerin gelişmesi ile birlikte son yıllarda daha büyük çalışma gruplarının dâhil edildiği çalışma sayısında artış olduğu düşünülmektedir.

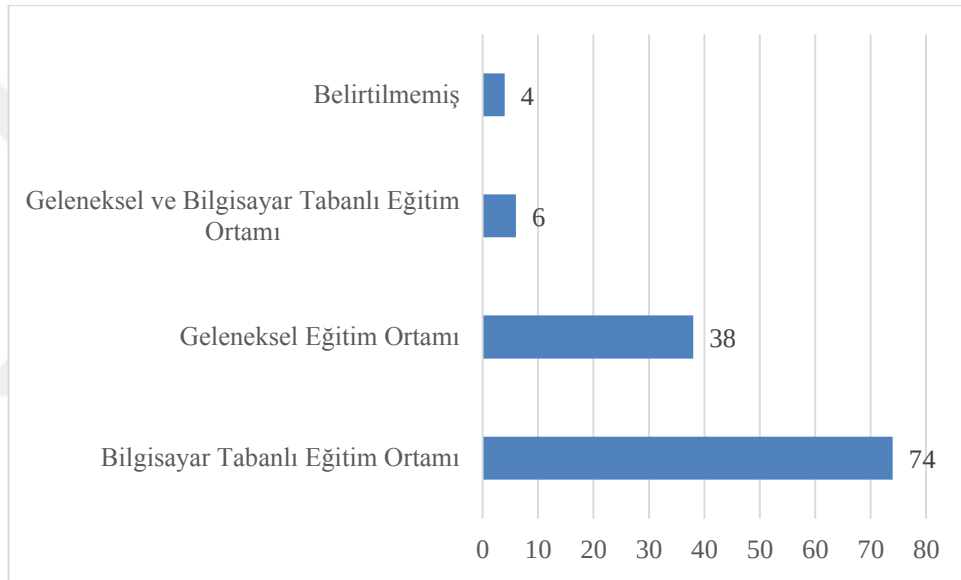
4.3.Üçüncü Araştırma Sorusu: Veri Toplama Süreci Özellikleri

Veri Toplama Ortamı Dağılımı

Eğitsel veri madenciliği, veri ortamı olarak geleneksel ve bilgisayar tabanlı eğitim ortamlarından elde edilen verileri analiz edebilmektedir. Bu iki ortamın elde edilen veri

türü ve boyutu açısından birbirlerine göre avantaj ve dezavantajları olabilmektedir. Bazı çalışmalar hem bilgisayar tabanlı ve hem de geleneksel eğitim ortamlarından elde ettikleri verileri kullanabilmektedir.

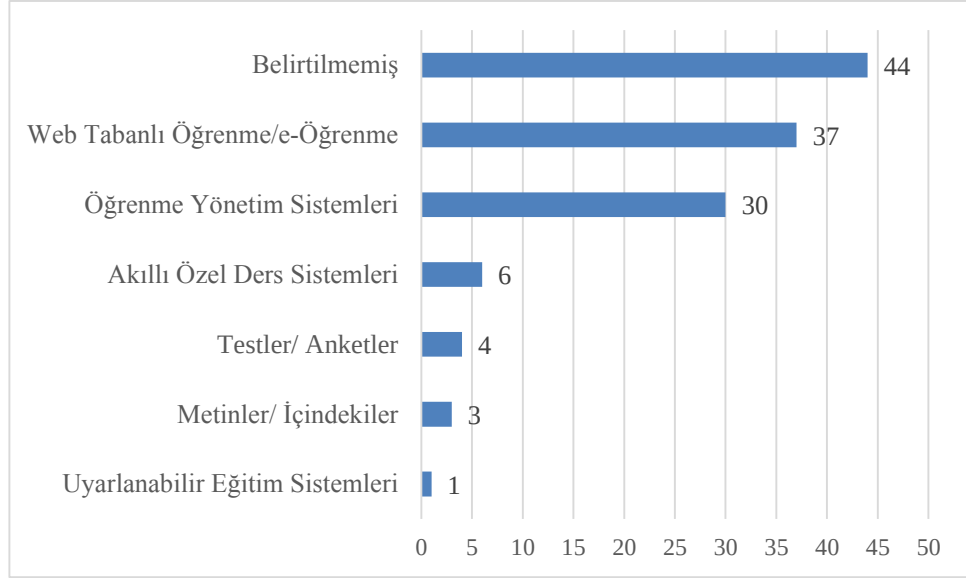
İncelenen çalışmalarda veri elde edilen ortam olarak en fazla bilgisayar tabanlı eğitim ortamı (74) olduğu görülmektedir (Şekil 4.6). Romero ve Ventura (2010) yaptıkları çalışmada, en çok veri elde edilen ortamın bilgisayar tabanlı eğitim ortamı olduğunu belirtmiştir. Geleneksel eğitim ortamından veri toplama sürecinin zor olmasının veri toplama ortamı tercihini etkilediği ve çevrim içi ortamlara yönelimi artırdığı düşünülmektedir.



Şekil 4.6. Çalışmaların veri toplama ortamı dağılımı

Bilgisayar Tabanlı Eğitim Ortam Türü Dağılımı

İncelenen çalışmalarda bilgisayar tabanlı eğitim ortam türlerinden e-öğrenme ortamları (37) ve öğrenme yönetim sistemleri (30) en çok tercih edilen ortam türü olduğu sonucuna ulaşılmıştır (Şekil 4.7). Akıllı ders sistemleri, çevrim içi anket ve test, çevrim içi metin ve uyarlanabilir eğitim sistemleri çok fazla tercih edilmeyen ortamlardır. Bunun sebebi bu ortam türlerinde geliştirilen sistemlerin veri toplama süreçlerinin gelişmiş olmaması olabilir. Çalışmaların 44 tanesinde ortam türü belirtilmemiştir.

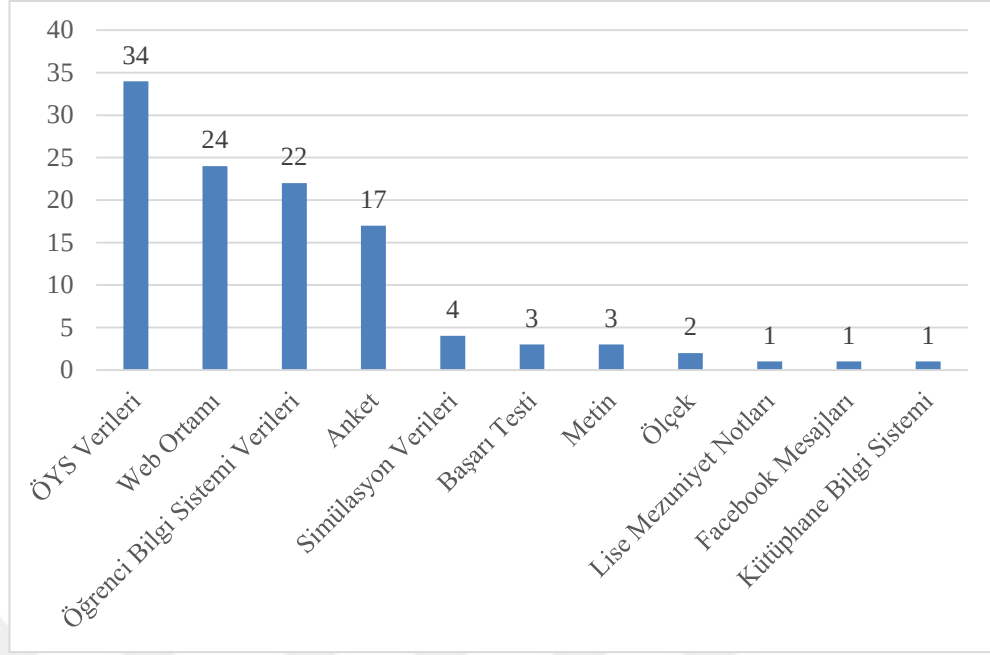


Şekil 4.7. Çalışmaların bilgisayar tabanlı ortam türü dağılımı

Çalışmaların üç tanesinde farklı iki ortam beraber kullanılmıştır. Bunlardan iki tanesinde Web Tabanlı Öğrenme ve Metin ortamları kullanılırken, diğerinde ise Öğrenme Yönetim Sistemi ve Metin ortamı kullanılmıştır.

Veri Toplama Aracı Dağılımı

Çalışmaların veri toplama aracına göre dağılımı Şekil 4.8’de gösterilmiştir. Veri toplama aracı ÖYS verileri, web ortamı, öğrenci bilgi sistemi verileri, anket, simülasyon verileri, başarı testi, lise mezuniyet notları, Facebook mesajları, kütüphane bilgi sistemi, makale metni, ölçek, envanter ve akademik teklif mektubu olmak üzere 13 kategoriye ayrılmıştır. Öğrenme yönetim sistemi verileri (34) en fazla kullanılan veri toplama aracıdır. Bunun yanı sıra Web ortamı (24), öğrenci bilgi sistemi (22) ve anket (17) veri toplama aracı olarak kullanılmıştır.



Şekil 4.8. Çalışmaların veri toplama aracı dağılımı

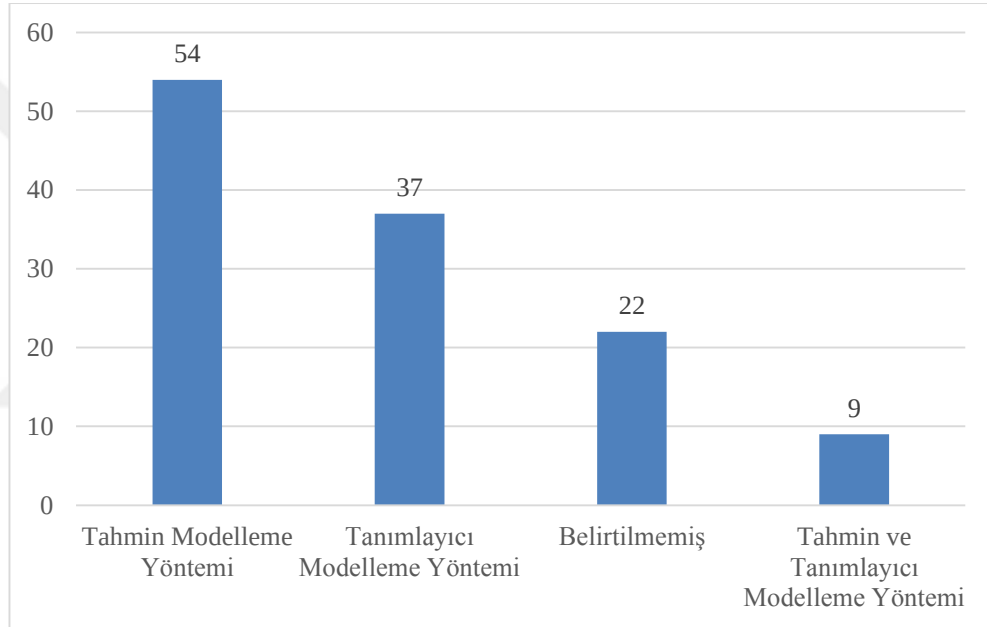
Öğrenme yönetim sistemi verileri Moodle (23), Massive Open Online Course (MOOC) (8), Sakai (1) ortamlarından elde edilmiştir. Mohamad ve Tasir (2013) eğitsel veri madenciliği konusundaki araştırmaların çoğunun Moodle ve e-öğrenme kullanımına önem verdiğini belirtmişlerdir. Manjarres, Sandoval ve Suárez (2018) tarafından yapılan alanyazın inceleme çalışmasında eğitsel veri madenciliği konusunda yapılan çalışmaların özellikle öğrenme yönetim sistemlerinde toplanan verilerin analizine odaklandığını tespit etmişlerdir.

Poptropica, MyEducator, Model-it, Geogebra ve Bioworld veri toplanan bazı web ortamlarıdır. Veri toplama aracı olarak kullanılan metin; makale metni (1), derse ait sorular (1) ve öğrenci tarafından üniversiteden kabul almak için yazılan akademik teklif mektubu (1) olmak üzere üç tanedir. Veri toplama aracı olarak kullanılan ölçekler “Öğrenme Stili Envanteri” ve “Çoklu Zekâ Ölçeği”dir. Veri toplama aracı olarak başarı testi kullanan çalışmalarından bir tanesi ülkemizde 2011 yılında düzenlenen İstanbul’da bulunan Ortaöğretim kurumlarına ait Liseye Yerleştirme Sınavı sonuçlarıdır.

4.4.Dördüncü Araştırma Sorusu: Veri Madenciliğine İlişkin Kullanılan Yöntemlerin Özellikleri

Veri Madenciliği Modelleme Yöntemine Göre Dağılım

Çalışmaların veri madenciliği modelleme yöntemine göre dağılımı Şekil 4.9’da verilmiştir. Grafiğe göre çalışmalarda en çok tahmin edici model (54) kullanılmıştır. Bazı çalışmalarda tahmin edici ve tanımlayıcı model (9) beraber kullanılmıştır. Çalışmaların 22’sinde ise veri madenciliği modelleme yöntemi belirtilmemiştir.



Şekil 4.9. Çalışmaların veri madenciliği modelleme yöntemi dağılımı

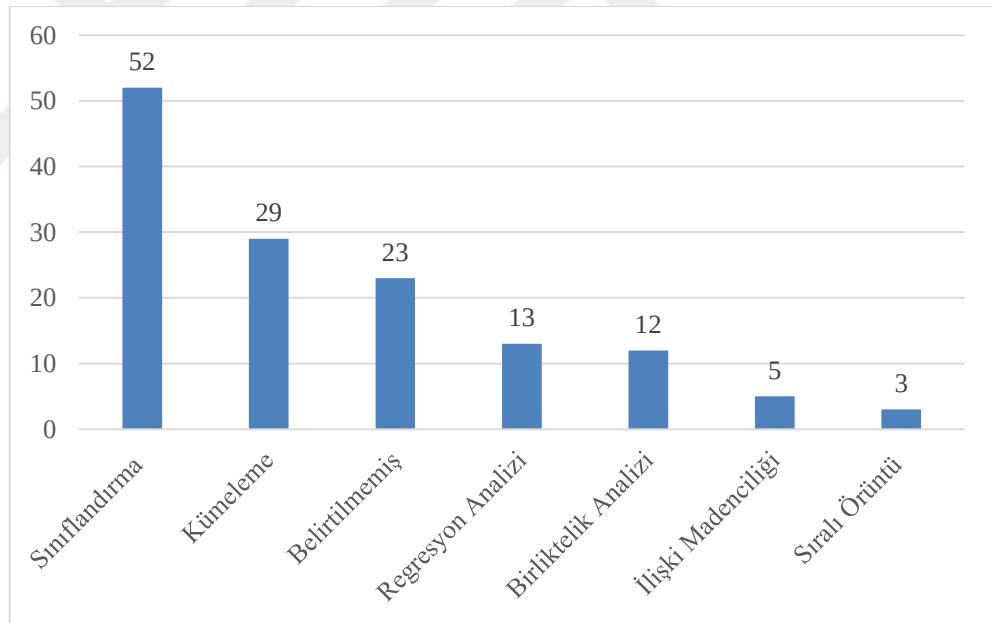
Benzer Şekilde Pena-Ayala (2014), 2010-2012 yılları arasında yayınlanan çalışmaların %60’ının tahmin edici, %40’nın tanımlayıcı model olduğunu belirtmiştir. Pena-Ayala (2014) çalışmalarda en fazla kullanılan veri madenciliği modelleme yönteminin tahmin edici modelleme yöntemi olduğu sonucuna ulaşmıştır. Tekin ve Öztekin (2016) de inceledikleri 92 çalışmalarda kullanılan veri madenciliği modelleme yöntemlerinden tahmin edici model (45) %49 ile en yüksek orana sahipken, tanımlayıcı modelin (20), %21.7 oranında olduğunu belirtmişlerdir.

Tahmin edici modellerin daha fazla tercih edilmesi EVM’nin doğası ile uyumludur. Baker (2010) tahmin yöntemlerinin EVM çalışmalarında kullanılarak öğrenci

çıktılarını ve öngörülmecek olayları önceden tahmin etmeyi sağlaması ile önemli konulara açıklık getirdiğini belirtmiştir.

Veri Madenciliği Veri Analiz Yöntemine Göre Dağılımı

Çalışmaların kullanılan veri analiz yöntemine göre dağılımı Şekil 4.9’da verilmiştir. Grafiğe göre veri madenciliği analiz yöntemi olarak en çok sınıflandırma analizi; en az ise sıralı örüntü madenciliği kullanılmıştır. Çalışmaları 16’sı veri analizi için iki farklı, ikisi ise üç farklı yöntem kullanmıştır. Özellikle sınıflandırma yöntemi (12), regresyon analizi (8), kümeleme (8), ilişki madenciliği (2), birliktelik analizi (4) ve sıralı örüntü (3) çalışmada başka yöntemlerle birlikte veri analizinde kullanılmıştır. Sınıflandırma ve kümeleme yöntemleri en çok birlikte kullanılan veri analiz yöntemleridir.



Şekil 4.9. çalışmaların veri madenciliği analiz yöntemi dağılımı

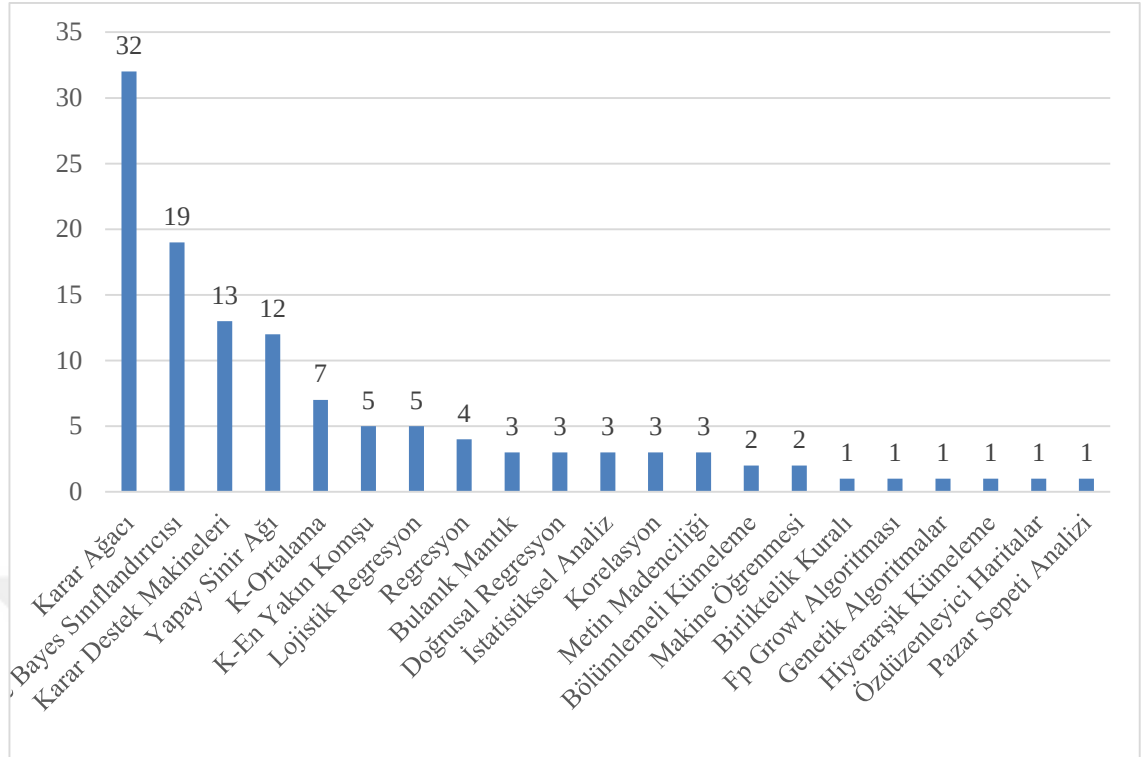
Pena-Ayala (2014) en çok kullanılan veri analiz yöntemlerinin sırasıyla sınıflandırma ve kümeleme olduğu ve çalışmaların %60’ında bu yöntemlerin kullanıldığını belirtmiştir. Bunun yanısıra Papamitsiou ve Economides (2014) tarafından yapılan çalışmada, çalışmalar veri madenciliği yöntemine göre değerlendirilmiştir. En popüler yöntemin sınıflandırma ardından da kümeleme olduğu belirlenmiştir. Bu

bakımdan bulgular benzeşmektedir.

Baker ve Yasef (2009) 1995-2005 yılları arasında en önde gelen veri madenciliği modelleme yönteminin %43 ile ilişki madenciliği olduğunu, ikinci araştırma yönteminin ise %28 ile tahmin yöntemi olduğunu; bu modelleme yöntemlerini %15 ile kümeleme yönteminin izlediğini belirtmişlerdir. Yazarlar, 1995-2005 yılları arasında ilişki madenciliği çalışmalarda yaygınken, 2008 ile 2009 yıllarında yayınlanan çalışmaların sadece %9'unun ilişki madenciliği içerdiğini ve ilişki madenciliğinin beşinci sıraya gerilediğini; çalışmaların %42'sini kapsayan tahmin yönteminin en yaygın yöntem haline geldiğini; kümeleme yönteminin %15 ile aynı konumunda kaldığını belirtmişlerdir.

Veri Madenciliği Veri Analizi Metodu Dağılımı

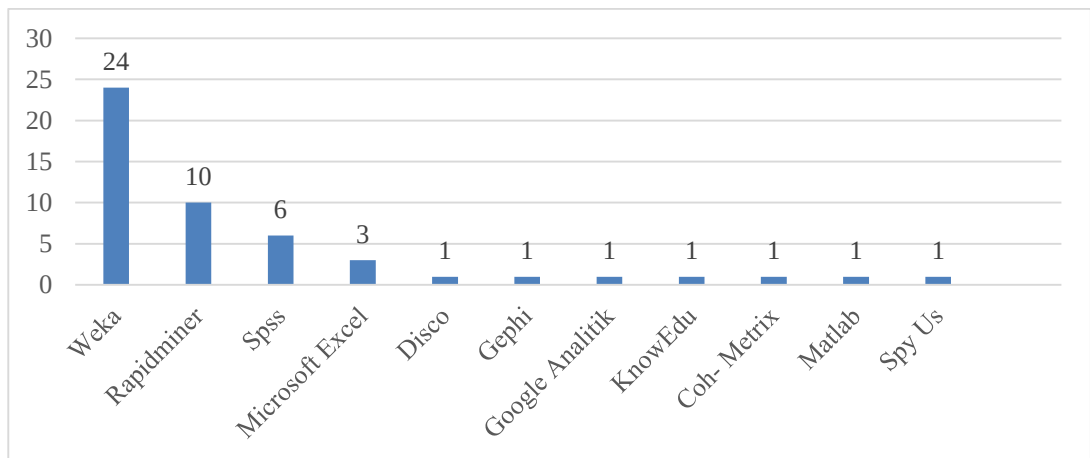
Çalışmaların 69 tanesinde veri analiz metodu belirtilmiştir (Şekil 4.10). Bazı çalışmalarda birden fazla veri analiz metodu kullanılmıştır. Veri analiz metodu belirtilen çalışmalarda yapılan sıklık analizinde sınıflandırma yöntemlerinden olan karar ağacının (32) en fazla kullanılan veri analiz metodu olduğu görülmüştür. Çalışmaların veri analiz metoduna bakıldığında karar ağacı (32), Naive Bayes sınıflandırıcısı (19), karar destek makineleri (13) ve yapay sinir ağları (12) en çok kullanılmış olup, 52 tane çalışmada sınıflandırma yöntemine ait analiz metodu kullanıldığı görülmektedir. Sınıflandırma analiz yöntemini kullanan çalışmalardan 32 tanesi karar ağacı veri analiz metodunu kullanmıştır.



Şekil 4.10. Çalışmaların veri madenciliği veri analiz metodu dağılımı

Veri Analizi İçin Kullanılan Yazılım Dağılımı

Çalışmaların 73 tanesinde veri analizi için kullanılan yazılım belirtilmemiş, geri kalan 49 tanesinde ise kullanılan yazılım belirtilmiştir (Şekil 4.11). Weka (24) ve Rapidminer (10) veri analizi için en fazla kullanılan yazılımlar olmuştur.



Şekil 4.11. Çalışmalarda kullanılan yazılım dağılımı

Yeni Zellanda Waikato Üniversitesi tarafında makine öğrenmesindeki en son tekniklere kolay erişim sağlayabilecek bir program ihtiyacından dolayı ilk 1992 yılında geliştirilmeye başlayan, temelde makine öğrenmesi algoritmalarını ve veri ön işleme işlemlerini amaçlayan Java ile geliştirilen açık kaynak kodlu veri madenciliği programıdır. Weka yazılımı regresyon, sınıflandırma, birliktelik kuralı madenciliği ve kümeleme algoritmalarını içermektedir. Weka yazılımının açık kaynak kodlu olması ve java ile uyumlu olması yaygın olarak kullanılmasını sağladığı düşünülmektedir (Hall ve ark., 2009)

4.5.Beşinci Araştırma Sorusu: Çalışmaların Sonuç Dağılımı

Eğitsel veri madenciliğinde yapılan çalışmalar son yıllarda artış göstermiş olsa da eğitsel veri madenciliğinin birçok görevi ve veri analizi için farklı metodları vardır. Bu nedenle eğitsel veri madenciliğinde öğrenci performansını tahmin ederken hangi metod, hangi veri, verinin hangi zamanda toplanması, hangi analiz sürecinin daha üstün olduğu ile ilgili genel yargılar oluşmamıştır. Çalışmalarda öğrenci performansını tahmin etmek için uygulanan tek veri analiz yönteminin her açıdan üstün performans göstermediği (Helal ve ark.,2018); tek bir algoritmadan ziyade birçok algoritmanın birlikte kullanılmasının daha doğru tahmin edeceği belirtilmiştir (Romero ve ark., 2013; Predic ve ark., 2018).

Öğrenme yönetim sisteminde öğrenci performansını tahmin etmek için bu sistemden elde edilen öğrenci erişim süreci verileri kullanarak öğrenci performansını tahmin etmenin mümkün olduğu (Gamulin ve ark., 2016; Jo ve ark., 2015); öğrenme yönetim sistemini etkin kullanmaları ile final notu arasında güçlü bir ilişki olduğu (Ayub ve ark., 2017), Naive Bayes metodu ile analiz edilen ara sınav notlarının final performansı için güçlü bir yordayıcı olduğu (Al-Barrak ve ark., 2015); öğrenci performansını tahmin etmek için öğrenme yönetim sisteminden elde edilen farklı veri türlerinin kullanılabileceği sonucuna ulaşılmıştır. Öğrencilerin performansının yapay sinir ağı (Şengür ve ark.,2014), karar ağacı (Natek ve ark.,2014), Naive Bayes (Al-Barrak ve ark., 2015); genetik algoritmalar (Wanli ve ark., 2015), birliktelik kuralı madenciliği (Luna ve ark.,2015), lojistik regresyon (Bucos ve ark., 2018) yöntemleri kullanılarak tahmin edilebileceği sonucuna ulaşılmıştır.

Eğitsel veri madenciliği ile öğrencinin dersi/okulu bırakma durumu (Marquez-

Vera ve ark.,2013; Sangodiah ve ark.,2015; Marquez-Vera ve ark., 2016; Iam-On ve ark., 2017), dersteki başarısızlığı (Costa ve ark., 2017), dikkatsizlik hataları (Gobert, 2015); yavaş öğrenme durumu (Kaur ve ark., 2015), akademik statü kaybı (Guarin ve ark., 2015) ve erteleme davranışı (Cerezo ve ark., 2017) gibi istenmeyen öğrenci davranışlarını etkileyen faktörleri ve istenmeyen durumların erken zamanda tespit edilebileceği sonucuna ulaşmıştır.

EVM'nin en önemli görevlerinden bir tanesi de yöneticiler, eğitimciler ve öğretmenler için geri bildirimde bulunmaktır. Romero ve ark. (2013) Moodle'da elde ettikleri sınavları öğretmenlere geri bildirim sağlamak için analiz ettikleri çalışmada gramer tabanlı genetik algoritmanın ilginç kuralları keşfetmek için yararlı olduğunu; Beemt ve ark. (2018) MOOC'daki öğrenme davranışındaki ve öğrenme sürecindeki kalıpları araştırdıkları çalışmada, süreç madenciliğinin bu tür çalışmalar için verimli bir yaklaşım olduğunu belirtmiştir. Chalaris ve ark. (2014), yüksek öğretim kurumları yönetimlerine EVM yöntemlerinin olumlu geri bildirimler sağlayabileceğini gördüklerini; Joorabchi ve ark. (2016) web sitesinin içeriğini analiz ettikleri çalışmalarında, önerdikleri metin madenciliği yaklaşımının yükseköğretimde verdikleri derslerde önemli ve zorlu konuları bilgi verme potansiyeline sahip olduğunu; Chamio-Gonzalez ve ark. (2015) EVM aracılığı ile alınan geri bildirimlerin yükseköğretimde öğretim elemanlarının öğretim yöntemlerini gözden geçirmesine ve öğrenme davranışlarını keşfetmelerine yardımcı olduğunu; Bautista ve ark. (2018) yüksek öğretimde öğrencilerin başarısız oldukları dersler arasındaki ilişkileri bulmak için yaptığı çalışmada kullandıkları birliktelik kuralı madenciliği yönteminden elde edilen sonuçlarının öğretim yöntemlerinde iyileştirme, ders ön şartlarının değiştirilmesi veya öğretimi geliştirmek için öğrencilere tanımlayıcı faaliyetlerin geliştirilmesi için yardımcı olabileceğini belirtmiştir. Wallden ve ark. (2014) probleme dayalı öğrenme yöntemine kullanıp kullanamama durumunu inceledikleri çalışmada log analizleri sonucunda, eğitsel veri madenciliğinin problem dayalı öğrenme yöntemini takip etme durumlarını izlemek için yararlı olduğunu belirtmişlerdir.

5. SONUÇ VE ÖNERİLER

5.1. Sonuç

Eğitsel veri madenciliği alanı ilk çalışmanın yapıldığı 1995 yılından itibaren kendini geliştirmekte ve yenilemekte olan bir alan olarak görülmektedir. Alanda yapılan çalışma sayısının hızlı bir şekilde artmasına rağmen eğitsel veri madenciliğinin eğitime daha fazla katkı sağlayabilmesi için daha fazla çalışma yapılması beklenmektedir. Eğitsel veri madenciliği konulu 2013-2018 yılları arasında yayımlanan bilimsel çalışmaların demografik özelliklerine ilişkin bulgulara bakıldığında;

Çalışmalar basım yıllarına göre incelendiğinde çalışma sayısında özellikle son yıllarda artış eğilimi olduğu görülmektedir. Alanın gelişmesine olan katkının da son yıllardaki çalışmalarla ile artış gösterdiği düşünülmektedir.

Erişim sağlanan veri tabanına bakıldığında önemli veri tabanlarından olan Scimedirect ve Scopus'tan erişim sağlanan çalışma sayısının daha fazla olduğu görülmektedir.

Çalışmaların yazar dağılımı analizinde toplamda 358 farklı yazarın eğitsel veri madenciliği konusunda yayın yaptığı sonucuna ulaşılmıştır. İspanya Cordobo Üniversitesi profesörlerinden Cristobal Romero (8) ve Sebastian Ventura (7) en fazla yayında ismi geçen yazarlardır. Aynı zamanda bu iki yazarın Google Scholar ve Scopus Veri Tabanlarında eğitsel veri madenciliği konulu en çok alıntılanma sayısına sahip çalışmanın yazarlarından olduğu görülmektedir.

Çalışmaların Google Scholar ve Scopus veri tabanlarında alıntılanma sayısına bakıldığında her iki veri tabanında da ortak 9 çalışma bulunurken, 1'er çalışma farklılık göstermektedir. Alıntılanma sayısı sıralamasında ilk 10'da bulunan çalışmaların uluslararası indekste olduğu görülmektedir. En çok alıntılanma yapılan çalışmaların öğrenci performansını tahmin etme (Romero ve ark., 2013; Kabakchieva, 2013; He, 2013; Zwilling ve Natek, 2014;), veri analizi için veri madenciliği modellerinin en etkin bütünleştirme potansiyelini ortaya çıkarma (Wanli ve ark., 2015), öğrenci başarısızlığını tahmin etme (Marquez-Vera ve ark., 2013; Costa ve ark., 2017), eğitim ortamlarından erken ayrılan öğrenciyi tespit etme (Lara ve ark. 2014; Marquez- Vera ve ark., 2013; Marquez- Vera ve ark., 2016), yavaş öğrenen öğrenciyi tespit etme (Kaur ve ark., 2015;)

konularında çalıştığı görülmektedir.

Eğitsel veri madenciliği konulu 2013-2018 yılları arasında yayımlanan çalışmaların araştırma konusuna ilişkin bulgulara bakıldığında en fazla öğrenci performansını tahmin etme, geri bildirim sağlama ve veri analizi ve görselleştirilmesi konularında araştırma yapıldığı sonucuna ulaşılmıştır.

Eğitsel veri madenciliği konulu 2013-2018 yılları arasında yayımlanan çalışmaların örneklem özelliklerine ilişkin bulgulara bakıldığında;

Üniversite öğrencileri en fazla yayın yapılan örneklem grubu olmuştur. Üniversite düzeyinde bilgisayar tabanlı eğitim ortamlarının daha sık kullanılması ile birlikte veri toplama sürecinin daha kolay olmasından dolayı çalışmalarda üniversite öğrencisi düzeyinin seçildiği düşünülmektedir. Üniversite öğrencilerini lise öğrencileri takip etmektedir.

Çalışmalarda çalışma grubu boyutu 34 yayın ile en fazla 100-300 kişilik grup olmuştur. Bunu 24 yayında 300-1000 kişi, 22 yayında 20-100 kişi ve 19 yayında 1000-120000 kişi takip etmektedir. Çalışmaların 58'inde örneklem büyüklüğü 100 ile 1000 kişi, 80'inde 20 ila 1000 kişilik örneklem grubu kullanılmıştır.

Eğitsel veri madenciliği konulu 2013-2018 yılları arasında yayımlanan yayınların veri toplama süreci özelliklerine ilişkin bulgulara bakıldığında;

Bilgisayar tabanlı eğitim ortamları geleneksel eğitim ortamlarında göre daha fazla tercih edilmiştir. Bunun sebebinin geleneksel eğitim ortamlarında veri toplama sürecinin daha zor olduğu düşünülmektedir. Bazı çalışmalarda bilgisayar tabanlı eğitim ortamı ve geleneksel eğitim ortamından elde edilen veriler birlikte kullanılmıştır.

Bilgisayar tabanlı eğitim ortamlarında yapılan çalışmaların çoğunda veri toplama ortam türü belirtilmemiştir. Belirtilen çalışmalarda e-öğrenme ortamları ve öğrenme yönetim sistemleri en çok tercih edilen ortam türleri olduğu sonucuna ulaşılmıştır.

Öğrenme yönetim sistemi kayıtları en fazla kullanılan veri toplama aracı olmuştur. Öğrenme yönetim sistemi olarak en fazla Moodle sistemi kullanılmıştır. Bunun sebebinin Moodle'un ücretsiz olması, erişim ve kullanımının kolay olması olduğu düşünülmektedir. Poptropica, MyEducator, Model-it, Geogebra ve Bioworld veri toplanan bazı web ortamlarıdır. Web ortamı kullanan çalışmaların birçoğu web ortamını belirtmemişlerdir.

Eğitsel veri madenciliği konulu 2013-2018 yılları arasında yayımlanan çalışmaların veri madenciliğine ilişkin kullanılan yöntem özelliklerine bakıldığında;

Tahmin edici model en çok kullanılan veri madenciliği modelleme yöntemi olmuştur. Çalışmaların dokuz tanesinde tanımlayıcı ve tahmin edici modelleme yöntemi birlikte kullanılmıştır.

Veri madenciliği veri analiz yöntemine göre tahmin edici modelleme yöntemlerinden olan sınıflandırma en fazla kullanılan veri analiz yöntemi olmuştur. Tanımlayıcı modelleme yöntemlerinden olan kümeleme yöntemi ise en fazla kullanılan ikinci veri analiz yöntemi olmuştur. Çalışmaların bazılarında farklı veri analiz yöntemleri birlikte kullanılmıştır. Sınıflandırma ve kümeleme veri analiz yöntemleri en çok birlikte kullanılan veri analiz yöntemleridir.

Çalışmaların 69 tanesinde veri analiz metodu belirtilmiştir. Bazı çalışmalarda birden fazla veri analiz metodu kullanılmıştır. Sınıflandırma yöntemlerinden olan karar ağacı, Naive Bayes sınıflandırıcısı, karar destek makineleri ve yapay sinir ağı en çok kullanılan veri analiz metotları olmuştur. Kümeleme yöntemlerinden ise k-ortalama en çok kullanılan veri analiz metodu olmuştur.

Veri analizi için kullanılan yazılım çoğu çalışmada belirtilmemiştir. Belirtilen çalışmalarda en çok kullanılan yazılım Weka ve Rapidminer olmuştur. Rapidminer'in tercih edilmesinin sebebi açık kaynak kodlu olması, Weka'nın tercih edilmesinin sebebi ise birçok platformda çalışabilen ve veri madenciliği yöntemlerinin hemen hemen hepsinin kullanılabileceği bir yazılım olması olabilir.

5.2. Öneriler

- Eğitsel veri madenciliği alanında 2013-2018 yılları arasında yapılan çalışmalarda;
- Araştırma konusu olarak öğrenci gruplama, kavram haritası geliştirme ve kurs içeriği ve eğitim yazılımı yapılandırılması konularında az sayıda çalışma yapıldığı sonucuna ulaşılmıştır. Bu araştırma konularında yapılacak çalışma sayısı artırılabilir.
 - Çalışma grubu olarak 79'unda üniversite öğrencileri tercih edilmiştir. Lise, ilköğretim, lisans üstü gibi çalışma grupları ile çalışma sayıları artırılabilir.
 - Çalışma grubu 27 çalışmada belirtilmemiştir. Veri madenciliğinde kullanılacak veri madenciliği yöntemleri veri büyüklükleri ile ilgilidir. Bu sebeple yazarların

alışmalarda kullandıkları alışma grubu boyutlarını belirtmeleri, arařtırmacılara faydalı olabilir.

- Bilgisayar tabanlı eēitim ortamları ve geleneksel eēitim ortamlarından elde edilen veri trlerinde farklılıklar olabilmektedir. Elde edilen veri tr, verinin toplanma sreci ve verinin analizinde yapılan farklılıklar, farklı sonulara ulařılmasına sebep olmaktadır. Geleneksel eēitim ortamından elde edilen veri trleride eēitimin iyileřtirilmesine farklı durumlarda katkı saēlayacaēı dřlmektedir. Bu sebeple geleneksel eēitim ortamında yapılacak alışma sayısı artırılabilir.
- Bilgisayar tabanlı eēitim ortam tr alışmaların oēunluēunda belirtilmemiřtir. Ortam trnn belirtilmesi arařtırmacıların ortam seimine yardımcı olacaēı iin ortam trnn belirtilmesi nerilmektedir.
- Veri toplama aracı olarak dijital ortamlar daha ok tercih edilmiřtir. Geleneksel eēitim ortamında elde edilen verinin dijital ortama aktarılmasını kolaylařtıracak aralar geliřtirilebilir.
- alışmaların oēunluēunda veri analiz metodu ve kullanılan yazılım belirtilmemiřtir. Yazarların alışmalarda kullandıkları veri analiz metodunu ve kullandıkları yazılımı belirtmeleri, hem alışmanın gvenirliēini ve geerliēini artıracak, hem de bu alanda alışan arařtırmacılara yntemi daha net sunacaktır. Bu nedenle eēitsel veri madenciliēi konulu yayınlarda veri analizinin kullanılan veri analizi yntemi ve yazılımı da ierecek řekilde daha ayrıntılı aıklanması nerilir.
- Alanyazın incelemesi ile farklı veri tabanlarında eēitsel veri madenciliēi eēilimleri incelenebilir.

KAYNAKLAR

- Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á. and 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.
- Akçapınar, G., 2014. Çevrim içi öğrenme ortamındaki etkileşim verilerine göre öğrencilerin akademik performanslarının veri madenciliği yaklaşımı ile modellenmesi. Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Doktora Tezi.
- Akçapınar, G., 2015. How automated feedback through text mining changes plagiaristic behavior in online assignments. **Computers & Education**, 87, 123-130.
- Akpınar, H., 2000. Veri tabanlarında bilgi keşfi ve veri madenciliği. **İÜ İşletme Fakültesi Dergisi**, 29(1), 1-22.
- Aksoy, E. and Narlı, S., 2015. An Examination of Mathematically Gifted Students' Learning Styles by Decision Trees. **Turkish Journal of Giftedness & Education**, 5(2).
- Al-Barrak, M.A. and Al-Razgan, M.A., 2015. Predicting students' performance through classification: a case study. **Journal of Theoretical and Applied Information Technology**. 75(2), 167-175.
- Alexandron, G., Ruipérez-Valiente, J. A., Chen, Z., Muñoz-Merino, P. J. and Pritchard, D. E., 2017. Copying@ Scale: Using harvesting accounts for collecting correct answers in a MOOC. **Computers & Education**. 108, 96-114.
- Ali Yahya, A., 2018. Centroid particle swarm optimisation for highdimensional data classification. **Journal Of Experimental & Theoretical Artificial Intelligence**, 30(6), 857-886.
- Aydın, S., 2007. Veri madenciliği ve anadolu üniversitesi uzaktan eğitim sisteminde bir uygulama. Anadolu Üniversitesi Sosyal Bilimler Enstitüsü Doktora Tezi.
- Ayub, M., Toba, H., Yong, S. and Wijanto, M.C., 2017. Modelling students' activities in programming subjects through educational data mining. **Global Journal of Engineering Education Volume**. 19(3), 249-255.
- Baker, R. S., 2010. Data mining for education. International encyclopedia
- Baker, R. S. and Yacef, K., 2009. The state of educational data mining in 2009: A review and future vision. **Journal of Educational Data Mining**, 1(1), 1–15.
- Bautista, R. M., 2018. Clustering failed courses of engineering students using association rule mining. **Journal of Theoretical and Applied Information Technology**, 96(4), 875-886.
- Bucos, M. and Drăgulescu, B., 2018. Predicting student success using data generated in traditional educational environments. **TEM Journal**. 7(3), 617-625.
- Castro, F., Vellido, A., Nebot, A. and Mugica, F., 2007. Applying data mining techniques to e-learning problems. In Evolution of teaching and learning paradigms in

- intelligent environment, Springer, Berlin, Heidelberg, 183-221.
- Cerezo, R., Esteban, M., Sánchez Santillán, M. and Núñez, J.C., 2017. Procrastinating behavior in computer-based learning environments to predict performance: a case study in moodle. **Frontiers in Psychology**, 8, 1403.
- Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C. and Tsolakidis, A., (2014). Improving quality of educational processes providing new knowledge using data mining techniques. **Procedia-Social and Behavioral Sciences**, 147, 390-397.
- Chamizo-Gonzalez, J., Cano-Montero, E.I., Urquia-Grande, E. and Muñoz Colomina, C.I., 2015. Educational data mining for improving learning outcomes in teaching accounting within higher education. **The International Journal of Information and Learning Technology**, 32(5), 272-285.
- Costa, E.B., Fonseca, B., Santana, M.A., Araújo, F.F. and 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.
- Dunham, M.H., 2003. **Data mining: introductory and advanced topics**. Prentice Hall, Pearson Education Inc., New Jersey.
- Dutt, A., Ismail, M.A. and Herawan, T., 2017. A systematic review on educational data mining. **IEEE Access**, 5, 15991- 16005.
- Emel, G.G. ve Taşkın, Ç., 2005. Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması. **Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi**, 6 (2), 221-239.
- Gamulin, J., Gamulin, O. and Kermek, D., 2016. Using Fourier coefficients in time series analysis for student performance prediction in blended learning environments. **Expert Systems**, 33(2), 189-200.
- Gobert, J.D., Kim, Y.J., Sao Pedro, M.A., Kennedy, M. and Betts, C.G., 2015. Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld. **Thinking Skills And Creativity**, 18, 81-90.
- Goldsmith, G.R., Fulton, A.D., Witherill, C.D. and Espeleta, J.F., 2014. Improving the efficacy of web-based educational outreach in ecology. **Ecosphere**, 5(10), 1-9.
- Guarín, C.E.L., Guzmán, E.L. and González, F.A., 2015. A model to predict low academic performance at a specific enrollment using data mining. **Ieee Revista Iberoamericana de tecnologías del Aprendizaje**, 10(3), 119-125.
- Gürbüz, S. ve Şahin, F., 2016. **Sosyal bilimlerde nitel araştırma yöntemleri**, 3. Baskı. Seçkin Yayıncılık, 401 s, Ankara.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H., 2009. The WEKA data mining software: an update. **ACM SIGKDD Explorations Newsletter**, 11(1), 10-18.

- Hamoud, A.K., 2018. Classifying students' answers using clustering algorithms based on principle component analysis. **Journal of Theoretical and Applied Information Technology**, 96(7), 1813-1825.
- Han, J. and Kamber, M., 2006. **Data mining: concepts and techniques (Second edition)**. Kaufmann Publisher, 2st Ed., San Francisco, USA.
- 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.
- Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D.J. and Long, Q., 2018. Predicting academic performance by considering student heterogeneity. **Knowledge-Based Systems**, 161, 134-146.
- Iam-On, N. and Boongoen, T., 2017. Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings. **International Journal Machine Learning and Cybernetics**, 8(29), 497–510.
- Ismail, S., 2015. Design and implementation of an intelligent system to predict the student graduation agpa. **Australian Educational Computing**, 30(2).
- Jo, I.H., Kim, D. and Yoon, M., 2015. Constructing proxy variables to measure adult learners' time management strategies in lms. **Journal of Educational Technology & Society**, 18(3), 214–225.
- Joorabchi, A., English, M. and Mahdi, A.E., 2016. An insight into challenges and subject-related difficulties faced by computer science learners. **Journal of Enterprise Information Management.**, 29(2), 255-275.
- Jugo, I., Kovačić, B. and Tijan, E., 2015. Cluster analysis of student activity in a web-based intelligent tutoring system. **Pomorstvo**, 29(1), 75-83.
- Kabakchieva, D., 2013. Predicting student performance by using data mining methods for classification. **Cybernetics and Information Technologies**, 13(1), 61-72.
- Kamber, J. H., and Morgan, M., 2000. **Data mining concept and techniques**. Kaufmann Publishers, 1st Ed., San Francisco, USA.
- Kaur, P., Singh, M. and Josanc, G.S., 2015. Classification and prediction based data mining algorithms to predict slow learners in education sector. **Procedia Computer Science**, 57, 500-508.
- Kausar, S., Huahu, X., Hussain, I., Wenhao, Z. and Zahid, M., 2018. Integration of data mining clustering approach in the personalized e-learning system. **IEEE Access**, 6, 72724-72734.
- Kumar, S.A. and Vijayalakshmi, M.N., 2011. Efficiency of decision tree in predicting student's academic performance. **Computer Science & Information Techonology**, 2, 335-343.
- Lara, J.A., Lizcano, D., Martínez, M.A., Pazos, J. and Riera, T. 2014. A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University Of Madrid, UDIMA. **Computers And Education**, 72, 23-36.

- Lavigne, G., Ruiz, G.R., Salas, L.M. and Sandoval, J.O., 2015. Log Analysis in a virtual learning environment for engineering students. **International Journal of Educational Technology in Higher Education**, 12(3), 113-128.
- Liñán, L.C. and Pérez, Á.A.J., 2015. Educational data mining and learning analytics: differences, similarities, and time evolution. **International Journal of Educational Technology in Higher Education**, 12(3), 98-112.
- Lu, L.F.L. and Jeng, I., 2006. Knowledge construction in inservice teacher online discourse: Impacts of instructor roles and facilitative strategies. **Journal of Research on Technology in Education**, 39(2), 183–202.
- Luna, J.M., Romero, C., Romero, J.R. and Ventura, S., 2015. An evolutionary algorithm for the discovery of rare class association rules in learning management systems. **Applied Intelligence**, 42(3), 501-513.
- Manjarres, A.V., Sandoval, L.G.M. and Suárez, M.J.S., 2018. Data mining techniques applied in educational environments: Literature Review. **Digital Education Review**, 33, 235-266.
- Márquez-Vera, C., Cano, A., Romero, C. and 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.
- Márquez- Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H. and Ventura, S., 2016. Early dropout prediction using data mining: a case study with high school students. **Expert Systems**, 33(1), 107-124.
- Márquez-Vera, C., Morales, C. R. and Soto, S. V., 2013. Predicting school failure and dropout by using data mining techniques. **IEEE Revista Iberoamericana de Tecnologías del Aprendizaje**, 8(1), 7-14.
- Miguéis, V.L., Paulo, A.F., Garcia, J.V. and Silva, A., 2018. Early segmentation of students according to their academic performance: A predictive modelling approach. **Decision Support Systems**, 115: 36-51.
- Millî Eğitim Bakanlığı (MEB), 2018. 2023 Eğitim Vizyon Belgesi. <http://2023vizyonu.meb.gov.tr/>.
- Mohamad, S.K. and Tasir, Z., 2013. Educational data mining: A review. **Procedia-Social and Behavioral Sciences**, 97, 320-324.
- Moore, A., 2006. Statistical data mining tutorials. Downloaded 1 August 2009 from <http://www.autonlab.org/tutorials/>
- Mühling, A., 2016. Aggregating concept map data to investigate the knowledge of beginning CS students. **Computer Science Education**, 26(2-3), 176-191.
- Natek, S. and Zwilling, M., 2014. Student data mining solution–knowledge management system related to higher education institutions. **Expert Systems With Applications**, 41(14), 6400-6407.
- Northcutt, C.G., Ho, A.D. and Chuang, I.L., 2015. Detecting and preventing “multiple-account” cheating in massive open online courses. **Computers & Education**, 100, 71-80.

- Özbay, Ö., 2015. Öğretim yönetim sistemi üzerinde üniversite (lisans) düzeyindeki öğrenci hareketliliğinin veri madenciliği yöntemleriyle analizi. Başkent Üniversitesi Eğitim Bilimleri Enstitüsü Yüksek Lisans Tezi.
- Özekeş, S., 2003. Veri madenciliği modelleri ve uygulama alanları. **İstanbul Ticaret Üniversitesi Dergisi**, 3, 65-82.
- Papamitsiou, Z. and Economides, A., 2014. Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. **Journal of Educational Technology & Society**, 17 (4), 49–64.
- Pena-Ayala, A., 2014. Educational data mining: A survey and a data mining-based analysis of recent works. **Expert Systems With Applications**, 41(4), 1432-1462.
- Poitras, E.G., Lajoie, S.P., Doleck, T. and Jarrell, A., 2016. Subgroup discovery with user interaction data: An empirically guided approach to improving intelligent tutoring systems. **Educational Technology & Society**, 19(2), 204-214.
- Predić, B., Dimić, G., Rančić, D., Štrbac, D., Maček, N. and Spalević, P., 2018. Improving final grade prediction accuracy in blended learning environment using voting ensembles. **Computer Applications In Engineering Education**, 26(6), 2294-2306.
- Priyambada, S.A., Mahendrawathi, E.R. and Yahya, B.N., 2017. Curriculum assessment of higher educational institution using aggregate profile clustering. **Procedia Computer Science**, 124, 264–273.
- Riofrio-Luzcando, D., Ramirez, J. and Berrocal-Lobo, M., 2017. Predicting student actions in a procedural training environment. **IEEE Transactions on Learning Technologies**, 10(4), 463-474.
- Romero, C. and Ventura, S., 2007. Educational data mining: a survey from 1995 to 2005. **Expert Systems with Applications**, 33(1), 135-146.
- Romero, C. And Ventura, S., 2010. Educational data mining: a review of the state of the art. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, 40(6), 601–618.
- Romero, C., Espejo, P., Zafra, A., Ventura, S. and Romero, J.R., 2013. Web usage mining for predicting final marks of students that use moodle courses. **Computer Applications in Engineering Education**, 21(1), 135-146.
- Romero, C. and Ventura, S., 2013. Data mining in education. **WIREs Data Mining Knowledge Discovery**, 3, 12–27.
- Sangodiah, A.V., Beleya, P., Muniandy, M., Heng, L.E. and Ramendran, C., 2015. Minimizing student attrition in higher learning institutions in Malaysia using support vector machine. **Journal of Theoretical and Applied Information Technology**, 71(3).
- Shao, Z., Li, Y., Wang, X., Zhao, X. and Guo, Y., 2018. Research on a new automatic generation algorithm of concept map based on text analysis and association rules mining. **Journal Of Ambient Intelligence And Humanized Computing**, 11, 1-13.

- Siemens, G. and d Baker, R.S., 2012. Learning analytics and educational data mining: towards communication and collaboration. In **Proceedings of the 2nd International Conference on Learning Analytics and Knowledge**, 252-254.
- Silahtaroğlu, G., 2016. Veri madenciliği kavram ve uygulamaları. **Papatya Yayıncılık**, 304 s, ISBN: 978-975-6797-81-5, İstanbul.
- Soffer, T., Kahan, T. and Livne, E., 2017. E-assessment of online academic courses via students' activities and perceptions. **Studies in Educational Evaluation**, 54, 83-93.
- Şengür, D. ve Tekin, A., 2013. Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini. **Gazi Üniversitesi Bilişim Teknolojileri Dergisi**, 6(3), 7-16.
- Tekin, A. ve Öztekin, Z., 2018. Eğitsel Veri Madenciliği ile ilgili 2006-2016 yılları arasında yapılan çalışmaların incelenmesi. **Eğitim Teknolojisi Kuram ve Uygulama**, 8(2), 67-89.
- Van den Beemt A., Buijs, J. and van der Aalst, W., 2018. Analysing structured learning behaviour in massive open online courses (MOOCs): An approach based on process mining and clustering. **International Review of Research in Open and Distributed Learning**, 19(5), 37-60.
- Verma, P., Sood, S.K. and Kalra, S., 2017. Smart computing based student performance evaluation framework for engineering education. **Computer Applications In Engineering Education**, 25(6), 977-991.
- Volk, H., Kellner, K. and Wohlhart, D., 2015. Learning analytics for english language teaching. **Journal of Universal Computer Science**, 21(1), 156-174.
- Wallden, S. and Mäkinen, E., 2014. Educational data mining and problem-based learning. **Informatics in Education**, 13(1), 141-156.
- Widyahastuti, F. and Tjhin, V.U., 2018. Performance prediction in online discussion forum: state of the art and comparative analysis. **Procedia Computer Science**, 135, 302-314.
- Xing, W., Guo, R., Petakovic, E. and Goggins, S., 2015. Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. **Computers Human in Behavior**, 47, 168-181.
- Yıldırım, A. ve Şimşek, H., 2016. **Sosyal bilimlerde nitel araştırma yöntemleri**. Seçkin Yayıncılık.
- Yoo, J. and Kim, J., 2014. Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns. **International Artificial Intelligence in Education Society**, 24 (1), 8-32.

ÖZGEÇMİŞ

Yazar, 1993 yılında Amasya/Taşova’da doğdu. İlkokul eğitimini Taşova’da, ortaokul eğitimini Tokat/Erbaa ‘da ve lise eğitimini Amasya Anadolu Öğretmen Lisesi’nde tamamladı. Gazi Üniversitesi Gazi Eğitim Fakültesi Bilgisayar ve Öğretim Teknolojileri Öğretmenliği Bölümünden 2015 yılında mezun oldu. 2016 yılı şubat ayında Hatay iline Bilişim Teknolojileri öğretmeni olarak atandı. Artvin/Borçka ilçesinde öğretmenlik görevini sürdürmektedir.



EK 1: Çalışma Kapsamında İncelenen Yayınlar

1. Akçapınar, G., 2015. How automated feedback through text mining changes plagiaristic behavior in online assignments. *Computers & Education*, 87, 123-130.
2. Akçapınar, G., Altun, A., and Aşkar, P., 2015. Modeling students' academic performance based on their interactions in an online learning environment. *İlköğretim Online*, 14(3).
3. Aksoy, E., and Narlı, S., 2015. An examination of mathematically gifted students' learning styles by decision trees. *Turkish Journal of Giftedness & Education*, 5(2).
4. Al-Barrak, M.A. and Al-Razgan, M.A., 2015. Predicting students' performance through classification: a case study. *Journal of Theoretical and Applied Information Technology*. 75(2), 167-175.
5. Alexandron, G., Ruipérez-Valiente, J. A., Chen, Z., Muñoz-Merino, P. J., and Pritchard, D. E., 2017. Copying@ Scale: Using harvesting accounts for collecting correct answers in a MOOC. *Computers & Education*. 108, 96-114.
6. Ali Yahya, A., 2018. Centroid particle swarm optimisation for high-dimensional data classification. *Journal of Experimental & Theoretical Artificial Intelligence*, 30(6), 857-886.
7. Alsaffar, A. H., 2017. Empirical study on the effect of using synthetic attributes on classification algorithms. *International Journal of Intelligent Computing and Cybernetics*, 10(2), 111-129.
8. Angeli, C., and Valanides, N., 2013. Using educational data mining methods to assess field-dependent and field-independent learners' complex problem solving. *Educational Technology Research and Development*, 61(3), 521-548.
9. Angeli, C., Howard, S. K., Ma, J., Yang, J., and Kirschner, P. A., 2017. Data mining in educational technology classroom research: Can it make a contribution?. *Computers & Education*, 113, 226-242.
10. Ayub, M., Toba, H., Yong, S. and Wijanto, M.C., 2017. Modelling students' activities in programming subjects through educational data mining. *Global Journal of Engineering Education Volume*. 19(3), 249-255.
11. Aziz, A. A., Ismail, N. H., and Ahmad, F., 2013. Mining students' academic performance. *Journal of Theoretical & Applied Information Technology*, 53(3).
12. Baker, R. S., Clarke- Midura, J., and Ocumpaugh, J, 2016. Towards general models of effective science inquiry in virtual performance assessments. *Journal of Computer Assisted Learning*, 32(3), 267-280.

13. Baker, R. S., Gowda, S. M., Corbett, A. T., and Ocumpaugh, J., 2012. Towards automatically detecting whether student learning is shallow. In International Conference on Intelligent Tutoring Systems, 444-453.
14. Bautista, R. M., 2018. Clustering failed courses of engineering students using association rule mining. *Journal of Theoretical and Applied Information Technology*, 96(4), 875-886.
15. Beemer, J., Spoon, K., Fan, J., Stronach, J., Frazee, J. P., Bohonak, A. J., and Levine, R. A., 2018. Assessing Instructional Modalities: Individualized Treatment Effects for Personalized Learning. *Journal of Statistics Education*, 26(1), 31-39.
16. Bharara, S., Sabitha, S., and Bansal, A., 2018. Application of learning analytics using clustering data Mining for Students' disposition analysis. *Education and Information Technologies*, 23(2), 957-984.
17. Bilen, Ö., Hotaman, D., Aşkın, Ö. E., ve Büyüklü, A. H., 2014. LYS başarılarına göre okul performanslarının eğitsel veri madenciliği teknikleriyle incelenmesi: 2011 İstanbul örneği. *Eğitim ve Bilim*, 39(172).
18. Bogarín Vega, A., Cerezo Menéndez, R., and Romero, C., 2018. Discovering learning processes using Inductive Miner: A case study with Learning Management Systems (LMSs). *Psicothema* 30(3), 322-329.
19. Bucos, M. and Drăgulescu, B., 2018. Predicting student success using data generated in traditional educational environments. *TEM Journal*. 7(3), 617-625.
20. Burgos, C., Campanario, M. L., de la Pena, D., Lara, J. A., Lizcano, D., and Martínez, M. A., 2018. Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers & Electrical Engineering*, 66, 541-556.
21. Bushati, S., Ninka, I., and Bushi, F., 2013. Assimilation Techniques and Technological Methods, Implemented in the Teaching and Learning Process. *Mediterranean Journal of Social Sciences*, 4(3), 763.
22. Campagni, R., Merlini, D., Sprugnoli, R., and Verri, M. C., 2015. Data mining models for student careers. *Expert Systems with Applications*, 42(13), 5508-5521.
23. Cerezo, R., Esteban, M., Sánchez Santillán, M., and Núñez, J.C., 2017. Procrastinating behavior in computer-based learning environments to predict performance: a case study in moodle. *Frontiers in Psychology*, 8, 1403.
24. Chalaris, M., Gritzalis, S., Maragoudakis, M., Sgouropoulou, C., and Tsolakidis, A., 2014. Improving quality of educational processes providing new knowledge using data mining techniques. *Procedia-Social and Behavioral Sciences*, 147, 390-397.
25. Chamizo-Gonzalez, J., Cano-Montero, E. I., Urquia-Grande, E., and Muñoz-Colomina, C. I., 2015. Educational data mining for improving learning outcomes in

- teaching accounting within higher education. *The International Journal of Information and Learning Technology*, 32(5), 272-285.
26. Chen, G., Davis, D., Krause, M., Aivaloglou, E., Hauff, C., and Houben, G. J., 2016. From Learners to Earners: Enabling MOOC Learners to Apply Their Skills and Earn Money in an Online Market Place. *IEEE Transactions on Learning Technologies*, 11(2), 264-274.
 27. Chen, P., Lu, Y., Zheng, V. W., Chen, X., and Yang, B., 2018. Knowedu: A system to construct knowledge graph for education. *IEEE Access*, 6, 31553-31563.
 28. Costa, E.B., Fonseca, B., Santana, M.A., Araújo, F.F. and 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.
 29. Cuadros, J., Artigas, C., Guitart, F., and Martori, F., 2015. Analyzing a virtual-lab based contextualized activity from action logs. *Procedia-Social and Behavioral Sciences*, 182, 441-447.
 30. Depren, S. K., Aşkın, Ö. E., and Öz, E., 2017. Identifying the classification performances of educational data mining methods: a case study for TIMSS. *Educ Sci Theory Pract*, 17(5), 1605-1623.
 31. DiCerbo, K. E., 2014. Game-based assessment of persistence. *Journal of Educational Technology & Society*, 17(1), 17-28.
 32. Dimić, G., Predić, B., Rančić, D., Petrović, V., Maček, N., and Spalević, P., 2018. Association analysis of moodle e- tests in blended learning educational environment. *Computer Applications in Engineering Education*, 26(3), 417-430.
 33. Dobashi, K., 2015. Time series analysis of the in class page view history of digital teaching materials using cross table. *Procedia Computer Science*, 60, 1032-1040.
 34. Dobashi, K., 2017. Automatic data integration from Moodle course logs to pivot tables for time series cross section analysis. *Procedia computer science*, 112, 1835-1844.
 35. Eguiluz, A., Guenaga, M., Garaizar, P., and Olivares-Rodriguez, C., 2017. Exploring the progression of early programmers in a set of computational thinking challenges via clickstream analysis. *IEEE Transactions on Emerging Topics in Computing*.
 36. Evale, D., 2016. Learning management system with prediction model and course-content recommendation module. *Journal of Information Technology Education: Research*, 16, 437-457.
 37. Fratamico, L., Conati, C., Kardan, S., and Roll, I., 2017. Applying a framework for student modeling in exploratory learning environments: Comparing data

- representation granularity to handle environment complexity. *International Journal of Artificial Intelligence in Education*, 27(2), 320-352.
38. Gamulin, J., Gamulin, O. and Kermek, D., 2016. Using Fourier coefficients in time series analysis for student performance prediction in blended learning environments. *Expert Systems*, 33(2), 189-200.
 39. Gobert, J. D., Kim, Y. J., Sao Pedro, M. A., Kennedy, M., and Betts, C. G., 2015. Using educational data mining to assess students' skills at designing and conducting experiments within a complex systems microworld. *Thinking Skills and Creativity*, 18, 81-90.
 40. Goldsmith, G.R., Fulton, A.D., Witherill, C.D., and Espeleta, J.F., 2014. Improving the efficacy of web-based educational outreach in ecology. *Ecosphere*, 5(10), 1-9.
 41. Gómez- Rey, P., Fernández- Navarro, F., and Barberà, E., 2016. Ordinal regression by a gravitational model in the field of educational data mining. *Expert Systems*, 33(2), 161-175.
 42. Guarín, C.E.L., Guzmán, E.L., and González, F.A., 2015. A model to predict low academic performance at a specific enrollment using data mining. *Ieee Revista Iberoamericana de tecnologias del Aprendizaje*, 10(3), 119-125.
 43. Hamoud, A.K., 2018. Classifying students' answers using clustering algorithms based on principle component analysis. *Journal of Theoretical and Applied Information Technology*, 96(7), 1813-1825.
 44. 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.
 45. Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D.J., and Long, Q., 2018. Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, 161, 134-146.
 46. Hlosta, M., Zdrahal, Z., and Zendulka, J., 2018. Are we meeting a deadline? classification goal achievement in time in the presence of imbalanced data. *Knowledge-Based Systems*, 160, 278-295.
 47. Huang, T. C. K., Huang, C. H., and Chuang, Y. T., 2016. Change discovery of learning performance in dynamic educational environments. *Telematics and Informatics*, 33(3), 773-792.
 48. Iam-On, N. and Boongoen, T., 2017. Improved student dropout prediction in Thai University using ensemble of mixed-type data clusterings. *International Journal Machine Learning and Cybernetics*, 8(29), 497-510.
 49. Iam-On, N., and Boongoen, T., 2017. Generating descriptive model for student dropout: a review of clustering approach. *Human-centric Computing and Information Sciences*, 7(1).

50. Ismail, S. and Abdullah S., 2015. Design and implementation of an intelligent system to predict the student graduation agpa. *Australian Educational Computing*, 30(2).
51. Ji, H., Park, K., Jo, J., and Lim, H., 2016. Mining students activities from a computer supported collaborative learning system based on peer to peer network. *Peer-to-Peer Networking and Applications*, 9(3), 465-476.
52. Jo, I. H., Kim, D., and Yoon, M., 2015. Constructing proxy variables to measure adult learners' time management strategies in LMS. *Journal of Educational Technology & Society*, 18(3), 214-225.
53. Joorabchi, A., English, M., and Mahdi, A.E., 2016. An insight into challenges and subject-related difficulties faced by computer science learners. *Journal of Enterprise Information Management.*, 29(2), 255-275.
54. Jugo, I., Kovačić, B., and Tijan, E., 2015. Cluster analysis of student activity in a webbased intelligent tutoring system. *Pomorstvo*, 29(1), 75-83.
55. Juhaňák, L., Zounek, J., and Rohlíková, L., 2019. Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*, 92, 496-506.
56. Kabakchieva, D., 2013. Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61-72.
57. Kahan, T., Soffer, T., and Nachmias, R., 2017. Types of participant behavior in a massive open online course. *The International Review of Research in Open and Distributed Learning*, 18(6). 1-18.
58. Karabulut, B., Cihan, Ş., Ünver, H. M., and Ergüzen, A., 2018. Electrolab: a new dataset for educational data mining. *Technological Applied Sciences*, 13(4), 318-328.
59. Kaur, P., Singh, M., and Josanc, G.S., 2015. Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia Computer Science*, 57, 500-508.
60. Kausar, S., Huahu, X., Hussam, I., Wenhao, Z. and Zahid, M., 2018. Integration of data mining clustering approach in the personalized e-learning system. *IEEE Access*, 6, 72724-72734.
61. Khan, A., and Ghosh, S. K., 2018. Data mining based analysis to explore the effect of teaching on student performance. *Education and Information Technologies*, 23(4), 1677-1697.
62. Kim, D., Park, Y., Yoon, M., and Jo, I. H., 2016. Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *The Internet and Higher Education*, 30, 30-43.

63. Lara, J.A., Lizcano, D., Martínez, M.A., Pazos, J. and Riera, T. 2014. A system for knowledge discovery in e-learning environments within the European Higher Education Area – Application to student data from Open University Of Madrid, UDIMA. *Computers And Education*, 72, 23-36.
64. Lavigne, G., Ruiz, G.R., Salas, L.M., and Sandoval, J.O., 2015. Log analysis in a virtual learning environment for engineering students. *International Journal of Educational Technology in Higher Education*, 12(3), 113-128.
65. Lee, J., Lee, C. H., Kim, D. W., and Kang, B. Y., 2016. Smartphone-assisted pronunciation learning technique for ambient intelligence. *IEEE Access*, 5, 312-325.
66. Lee, Y. J., 2017. Modeling students' problem solving performance in the computer-based mathematics learning environment. *The International Journal of Information and Learning Technology*, 34(5), 385-395.
67. Lei, C. U., Man, K. M., and Ting, T. O., 2014. Using coh-metrix to analyse writing skills of students: A case study in a technological common core curriculum course. *Lecture Notes in Engineering and Computer Science.*, 41(3):193-197.
68. Lei, C. U., Man, K. M., and Ting, T. O., 2014. Using coh-metrix to analyse writing skills of students: A case study in a technological common core curriculum course. *Lecture Notes in Engineering and Computer Science*.
69. Lian, D. F., and Liu, Q., 2018. Jointly Recommending Library Books and Predicting Academic Performance: A Mutual Reinforcement Perspective. *Journal of Computer Science and Technology*, 33(4), 654-667.
70. Luna, J. M., Romero, C., Romero, J. R., and Ventura, S., 2015. An evolutionary algorithm for the discovery of rare class association rules in learning management systems. *Applied Intelligence*, 42(3), 501-513.
71. Maitra, S., Madan, S., Kandwal, R., and Mahajan, P., 2018. Mining authentic student feedback for faculty using Naïve Bayes classifier. *Procedia computer science*, 132, 1171-1183.
72. Márquez-Vera, C., Cano, A., Romero, C., and 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), 315330.
73. Márquez- Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., and Ventura, S., 2016. Early dropout prediction using data mining: a case study with high school students. *Expert Systems*, 33(1), 107-124.
74. Márquez-Vera, C., Morales, C. R., and Soto, S. V., 2013. Predicting school failure and dropout by using data mining techniques. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 8(1), 7-14.

75. Miguéis, V.L., Paulo, A.F., Garcia, J.V., and Silva, A., 2018. Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115: 36-51.
76. Mühling, A., 2016. Aggregating concept map data to investigate the knowledge of beginning CS students. *Computer Science Education*, 26(2-3), 176-191.
77. Nagao, K., Morita, N., and Ohira, S., 2017. Evidence-based education: case study of educational data acquisition and reuse. *J. Syst. Cybern. Inform.: JSCI*, 15(7), 77-84.
78. Natek, S., and Zwillling, M., 2014. Student data mining solution–knowledge management system related to higher education institutions. *Expert Systems With Applications*, 41(14), 6400-6407.
79. Northcutt, C.G., Ho, A.D., and Chuang, I.L., 2015. Detecting and preventing “multipleaccount” cheating in massive open online courses. *Computers & Education*, 100, 71-80.
80. Nyland, R., Davies, R. S., Chapman, J., and Allen, G., 2017. Transaction-level learning analytics in online authentic assessments. *Journal of Computing in Higher Education*, 29(2), 201-217.
81. Oeda, S., and Hashimoto, G., 2017. Log-Data Clustering Analysis for Dropout Prediction in Beginner Programming Classes. *Procedia Computer Science*, 112, 614-621.
82. Olivares-Rodríguez, C., Guenaga, M., and Garaizar, P., 2018. Using children’s search patterns to predict the quality of their creative problem solving. *Aslib Journal of Information Management*, 70(5), 538-550.
83. Olivares-Rodríguez, C., Guenaga, M., and Garaizar, P., 2018. Using children’s search patterns to predict the quality of their creative problem solving. *Aslib Journal of Information Management*, 70(5), 538-550.
84. Olmo, J. L., Romero, C., Gibaja, E., and Ventura, S., 2015. Improving Meta-learning for Algorithm Selection by Using Multi-label Classification: A Case of Study with Educational Data Sets. *International Journal of Computational Intelligence Systems*, 8(6), 1144-1164.
85. Park, Y., Yu, J. H., and Jo, I. H., 2016. Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education*, 29, 1-11.
86. Paule Ruiz, M. P., Riestra González, M., Sánchez Santillán, M., and Pérez Pérez, J. R., 2015. The Procrastination related indicators in e-learning platforms. *Journal of Universal Computer Science*.
87. Pintar, D., Begušić, D., Škopljanać-Maćina, F., and Vranić, M., 2018. Automatic extraction of learning concepts from exam query repositories.

88. Poitras, E. G., Lajoie, S. P., Doleck, T., and Jarrell, A., 2016. Subgroup discovery with user interaction data: An empirically guided approach to improving intelligent tutoring systems. *Journal of Educational Technology & Society*, 19(2), 204-214.
89. Predić, B., Dimić, G., Rančić, D., Štrbac, D., Maček, N., and Spalević, P., 2018. Improving final grade prediction accuracy in blended learning environment using voting ensembles. *Computer Applications In Engineering Education*, 26(6), 2294-2306.
90. Priyambada, S.A., Mahendrawathi, E.R., and Yahya, B.N., 2017. Curriculum assessment of higher educational institution using aggregate profile clustering. *Procedia Computer Science*, 124, 264–273.
91. Reamer, A. C., Ivy, J. S., Vila-Parrish, A. R., and Young, R. E., 2015. Understanding the evolution of mathematics performance in primary education and the implications for STEM learning: A Markovian approach. *Computers in Human Behavior*, 47, 4-17.
92. Riofrio-Luzcando, D., Ramirez, J., and Berrocal-Lobo, M., 2017. Predicting student actions in a procedural training environment. *IEEE Transactions on Learning Technologies*, 10(4), 463-474.
93. Riofrío-Luzcando, D., Ramírez, J., Moral, C., de Antonio, A., and Berrocal-Lobo, M., 2019. Visualizing a collective student model for procedural training environments. *Multimedia Tools and Applications*, 78(8), 10983-11010.
94. Romero, C., Espejo, P., Zafra, A., Ventura, S., and Romero, J.R., 2013. Web usage mining for predicting final marks of students that use moodle courses. *Computer Applications in Engineering Education*, 21(1), 135-146.
95. Romero, C., Zafra, A., Luna, J. M., and Ventura, S., 2013. Association rule mining using genetic programming to provide feedback to instructors from multiple- choice quiz data. *Expert Systems*, 30(2), 162-172.
96. Rowe, E., asbell-Clarke, J, Baker, R.S., Eagle, M., Hicks, A.G., Barnes, T.M., Brown, R.A. and Edwards, T., 2017. Assessing implicit science learning in digital games. *Computers In Human Behavior*. 76:617-630.
97. Ruipérez- Valiente, J. A., Muñoz- Merino, P. J., and Delgado Kloos, C., 2018. Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators. *Expert Systems*, 35(6), e12298.
98. San Pedro, M. O. Z., d Baker, R. S., and Rodrigo, M. M. T., 2014. Carelessness and affect in an intelligent tutoring system for mathematics. *International Journal of Artificial Intelligence in Education*, 24(2), 189-210.
99. Sandoval, A., Gonzalez, C., Alarcon, R., Pichara, K., and Montenegro, M., 2018. Centralized student performance prediction in large courses based on low-cost variables in an institutional context. *The Internet and Higher Education*, 37, 76-89.

100. Sangodiah, A.V., Beleya, P., Muniandy, M., Heng, L.E., and Ramendran, C., 2015. Minimizing student attrition in higher learning institutions in Malaysia using support vector machine. *Journal of Theoretical and Applied Information Technology*. 71(3).
101. Sarra, A., Fontanella, L., and Di Zio, S., 2019. Identifying students at risk of academic failure within the educational data mining framework. *Social Indicators Research*, 146(1-2), 41-60.
102. Shao, Z., Li, Y., Wang, X., Zhao, X., and Guo, Y., 2018. Research on a new automatic generation algorithm of concept map based on text analysis and association rules mining. *Journal of Ambient Intelligence and Humanized Computing*, 1-13.
103. Shaojie Qu, S., Li, K., Zhang, S. and Wang, Y., 2018. Predicting achievement of students in smart campus. *Ieee Acces*. 6:60264-60273.
104. Shrestha, R.M., Orgun, M.A. and Busch, P., 2016. Offer acceptance prediction of academic placement. *Neural Computing And Applications*. 27(8):2351-2368.
105. Soffer, T., Kahan, T. and Livne, E., 2017. E-assessment of online academic courses via students' activities and perceptions. *Studies in Educational Evaluation*. 54: 83-93.
106. Stahovich, T. F., and Lin, H., 2016. Enabling data mining of handwritten coursework. *Computers & Graphics*, 57, 31-45.
107. Şengür, D., ve Tekin, A., 2013. Öğrencilerin mezuniyet notlarının veri madenciliği metotları ile tahmini. *Gazi Üniversitesi Bilişim Teknolojileri Dergisi*, 6(3), 7-16.
108. Tang, H., Xing, W. and Pei, B., 2018. Exploring the temporal dimension of forum participation in MOOCs. *Exploring The Temporal Dimension Of Forum Participation In MOOCs, Distance Education*. 39(3): 353-372.
109. Tekin, A., 2014. Early prediction of students' grade point averages at graduation: A data mining approach. *Eurasian Journal of Educational Research*, 54, 207-226.
110. Urbina Nájera, A. B., de la Calleja, J., and Medina, M. A., 2017. Associating students and teachers for tutoring in higher education using clustering and data mining. *Computer Applications in Engineering Education*, 25(5), 823-832.
111. Van den Beemt A., Buijs, J., and van der Aalst, W., 2018. Analysing structured learning behaviour in massive open online courses (MOOCs): An approach based on process mining and clustering. *International Review of Research in Open and Distributed Learning*. 19(5), 37-60.
112. Veeramanickam, M. R. M., Mohanapriya, M., Pandey, B. K., Akhade, S., Kale, S. A., Patil, R., and Vigneshwar, M., 2018. Map-reduce framework based cluster

- architecture for academic student's performance prediction using cumulative dragonfly based neural network. *Cluster Computing*, 22(1), 1259-1275.
113. Verma, P., Sood, S.K., and Kalra, S., 2017. Smart computing based student performance evaluation framework for engineering education. *Computer Applications In Engineering Education*, 25(6), 977-991.
 114. Viswanathan, S. A., and VanLehn, K., 2017. Using the tablet gestures and speech of pairs of students to classify their collaboration. *IEEE Transactions on Learning Technologies*, 11(2), 230-242.
 115. Volk, H., Kellner, K., and Wohlhart, D., 2015. Learning analytics for english language teaching. *Journal of Universal Computer Science*, 21(1), 156-174.
 116. Wallden, S., and Mäkinen, E., 2014. Educational data mining and problem-based learning. *Informatics in Education*, 13(1), 141–156.
 117. Wook, M., Yusof, Z.M. and Nazr, M.Z.A., 2017. Educational data mining acceptance among undergraduate students. *Education And Information Technologies: The Official Journal Of The IFIP Technical Committee On Education*, 22(3):1195-1216.
 118. Wu, J. Y., Hsiao, Y. C., and Nian, M. W., 2018. Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment. *Interactive Learning Environments*, 1-16.
 119. Xing, W., Guo, R., Petakovic, E., and Goggins, S., 2015. Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers Human in Behavior*, 47, 168-181.
 120. Yahya, A. A., Osman, A., and El-Bashir, M. S., 2017. Rocchio algorithm-based particle initialization mechanism for effective PSO classification of high dimensional data. *Swarm and evolutionary computation*, 34, 18-32.
 121. Yoo, J., and Kim, J., 2014. Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns. *International Journal of Artificial Intelligence in Education*, 24(1), 8-32.
 122. Zhou, Q., Quan, W., Zhong, Y., Xiao, W., Mou, C. and Wang, Y., 2018. Predicting high-risk students using Internet access logs. *Knowledge And Information Systems*, 55(2):393-413.