

EĞİTSEL VERİ MADENCİLİĞİ YÖNTEMİ İLE İLGİLİ YAPILMIŞ ÇALIŞMALARIN İNCELENMESİ: İÇERİK ANALİZİ

Kübranur AKGÜN*, Müzeyyen BULUT ÖZEK**

Makale Geliş Tarihi: 16.07.2020

Makale Kabul Tarihi: 27.12.2020

Özet

Veri madenciliği (VM) uygulamaları teknolojik gelişmelerin yaşanması ve internetin hayatımıza girmesi ile birlikte oldukça önem kazanmıştır. Bu durum konu ile ilgili yapılan çalışmaların sayısının artmasına neden olmuştur. Özellikle bireyselleştirilmiş öğrenme, öğrenci davranışlarını gruplandırarak modelleme ve bazı durumların (akademik başarı, okulu bırakma, çalışma alışkanlığı vs.) tahmininde oldukça başarılı sonuçlar vermesi, eğitsel anlamda bu uygulamaların kullanımı ile ilgili çalışma konusunda araştırmacıları cesaretlendirmiştir. Bu çalışmada veri madenciliği uygulamaları ile ilgili yıl sınırlaması olmaksızın Google akademik, EBSCO, Elsevier, ERIC ve YÖKTez veri tabanlarında yayınlanmış olan araştırmaların incelenmesi amaçlanmıştır. Araştırmaya toplam 102 çalışma dâhil edilmiş, nitel araştırma yöntemlerinden doküman inceleme tekniği kullanılarak analiz edilmiştir. Sonuç olarak belirlenen yıl aralığında konu ile ilgili yapılan çalışmaların artış eğiliminde olduğu görülmüştür. Ayrıca yapılan çalışmaların ana odağının farklı veritabanları ve öğrenme yönetim sistemleri aracılığıyla öğrencilerin ileri dönemlerdeki akademik başarısının tahmin etmeye yönelik oluşturulduğu gözlenmiştir. Bu bulguyu takiben çalışmaların ikinci odak noktası ise öğrenci tutum ve davranışlarını belirlemek amacıyla gerçekleştirilmesidir. Son olarak konuyla ilgili çalışmalarda sıklıkla başvurulan veri madenciliği aracı WEKA olurken en fazla başvurulan veri madenciliği teknikleri ise karar ağaçları ve yapay sinir ağları olarak ortaya çıkmıştır. Araştırma sonuçlarının gelecekte bu konuda çalışma yapacak araştırmacılara yol göstermesi beklenmektedir. Bu nedenle araştırmacılar tarafından farklı değişkenler üzerinden farklı bakış açıları derinleştirilerek yeni çalışmaların yapılması önerilebilir.

Anahtar Kelimeler: Eğitsel veri madenciliği, veri madenciliği, içerik analizi

INVESTIGATION OF THE STUDIES ABOUT THE EDUCATIONAL DATA MINING METHOD: CONTENT ANALYSIS

Abstract

Data mining (VM) applications have gained importance with technological developments and the internet entering our lives. This situation caused an increase in the number of studies on the subject. Especially individualized learning, grouping student behaviors, modeling and giving quite successful results in the prediction of some situations (academic achievement, dropout, work habit, etc.) encouraged researchers to work

* Yüksek Lisans Öğrencisi, Fırat Üniversitesi, kubranur.akgun7@gmail.com, ORCID:0000-0002-5129-6345

** Dr. Öğr. Üyesi, Fırat Üniversitesi, muzeyyenbulut@gmail.com, ORCID:0000-0001-7594-8937

on the use of these applications in educational sense. In this study, it is aimed to examine the researches published in Google academic, EBSCO, Elsevier,

ERIC and YOKTez databases without year limitation regarding data mining applications. A total of 102 studies were included in the study and analyzed using the document analysis technique, which is one of the qualitative research methods. As a result, it has been observed that the studies on the subject in the determined year range are in an increasing trend. In addition, it was observed that the main focus of the studies was created to predict students' academic success in the future through different databases and learning management systems. Following this finding, the second focus of the studies is to be conducted to determine student attitudes and behaviors. Finally, WEKA is the data mining tool that is frequently used in studies on the subject, while the most frequently used data mining techniques have emerged as decision trees and artificial neural networks. The results of the research are expected to guide the researchers who will work on this subject in the future. For this reason, new studies may be suggested by researchers by deepening different perspectives on different variables.

Keywords: Educational data mining, data mining, content analysis

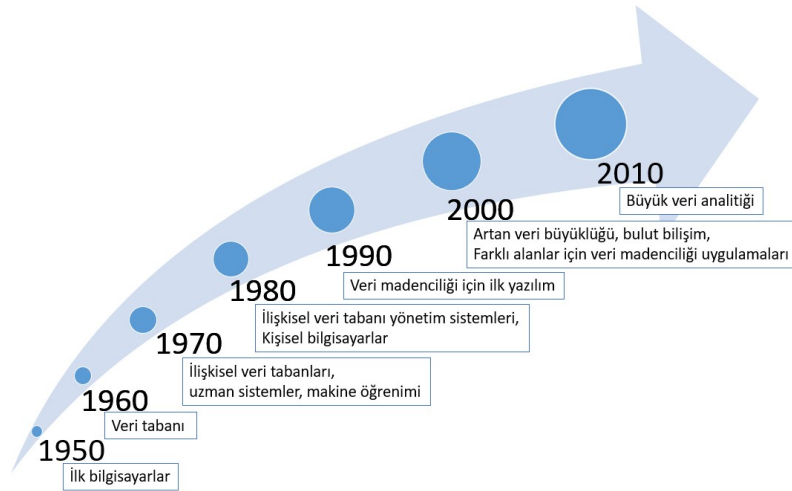
1. GİRİŞ

Bilgisayarların icat edilmesi ve hayatımıza girmesi ile birlikte bu ortamlar üzerinde verilerin depolanması süreci de başlamıştır. Teknolojinin gelişmesi ve internetin hayatımıza girmesi bir taraftan büyük verilerin saklanması kolaylaştırırken diğer taraftan çok büyük miktarda ve çok farklı türlerde veri yığınlarının oluşmasına neden olmuştur. Oluşan veri yığınlarının boyutu ile paralel olarak verilerin tutulduğu veri tabanlarının boyutlarında da artışlar gözlenmiştir (Taşdemir, 2012). Bunun yanı sıra verinin hacim (volume) ve çeşitliliğinin (variety) yanında veriye ait hız (velocity), doğrulama (verification), değer (value), gerçeklik (veracity), değişkenlik (variability), hassaslık (vulnerability) görselleştirme (visualization) gibi yeni kavramlar da ortaya çıkmıştır (Atalay ve Çelik, 2017). Büyük veri tabanlarında saklanan bu veriler, çok çeşitli özelliklere sahip olmakla birlikte tek başlarına değersiz olup ancak bir amaç doğrultusunda işlendiği zaman bilgiye dönüşüp değer kazanabilmektedir (Savaş, Topoğlu ve Yılmaz, 2012). Ancak yüksek miktardaki bu ham verilerin analiz edilerek içerisinden anlamlı bilgilerin çıkarılması noktasında klasik istatistiksel yöntemler yetersiz kalmaktadır (Ayık, Özdemir ve Yavuz, 2007; Şengür ve Tekin, 2013; Türel ve Baz, 2016). Bu yüzden ham veriyi anlamlı hale getirme konusunda araştırmacılar tarafından farklı disiplinlere ihtiyaç duyulmuştur (Çalış, Kayapınar ve Çetinyokuş, 2014). Bunun sonucunda da veri madenciliğine başvurulmuştur.

Farklı disiplin arayışı sonucunda ulaşılan yöntemler arasında son yıllarda oldukça yaygın bir kullanıma sahip olan veri madenciliği ile ilgili alan yazında birçok tanımlamaya rastlamak mümkündür. Bu tanımlamalardan bazıları ise şu şekildedir; veri madenciliği, çok büyük miktarda ham verinin oluşturulup depolandığı sağlıktan alışverişe, bankacılıktan eğitime kadar hemen her alanda kullanılabilen bir disiplindir (Erten, 2015 akt. Özbay, 2015). Türel ve Baz (2016) tarafından yapılan tanımlamada veri madenciliği, “Veri yığınları içerisinde anlamı bilinmeyen ve avantaj oluşturabilecek bilgilerin çıkarıldığı yapay zekâ, veri tabanı yönetimi ve makine öğrenme gibi birçok işlemin olduğu veri analizi yöntemleridir.” şeklinde ifade edilmiştir. Türel ve Baz(2016)’ın yapmış olduğu tanımlamaya benzer olarak Ching ve Michael (2002) ‘den aktaran Çalış, Kayapınar ve Çetinyokuş (2014)’a göre veri madenciliği “makine öğrenme, istatistik ve veri tabanları alanlarındaki teknikleri birleştirerek büyük veri tabanlarından faydalı ve değerli bilgiyi çıkarmamıza imkân tanımaktadır”. Bununla birlikte

literatürde yer alan veri madenciliği tanımlamalarından yola çıkarak Irmak (2009) tarafından yapılan kapsamlı tanımlamada ise veri madenciliği, büyük veri yığınlarında tutulan veriler arasındaki ve mevcut yöntemler ile ayrıştırılamayan ilişkileri, örüntüleri veya yönelimleri tespit etmek amacıyla bilgisayar uygulamaları, istatistik ve makine öğrenimi alanlarındaki tekniklerin birlikte kullanılarak analiz edilmesi ve görselleştirilmesi süreci olarak ifade edilmiştir.

Veri madenciliğinin tarihi, Şekil 1’de görüldüğü gibi 1950’deki II. Dünya Savaşı sırasında ABD ordusu için geliştirilen ilk bilgisayarlar (ENIAC) başlamaktadır. Önceleri karmaşık hesaplamaları gerçekleştirmek amacıyla kullanılan bilgisayarlar daha sonra veri depolama amacıyla kullanılmıştır. Bu durum 1960’da veri tabanının ortaya çıkmasını sağlamıştır. 1970’de veri tabanlarında daha fazla bilginin depolanma ve veri tabanlarının düzenlenmesi ihtiyaçları veri modellemeyi doğurmuştur. Bugün de en çok tercih edilen veri modelleme yöntemi, ilişkisel veri modellemesidir. Yine aynı yıllarda geçmiş deneyimleri ve örnekleri problem çözmede kullanarak bilgisayarları programlayan makine öğrenmesi de gelişimine başlamıştır. İstatistik, makine öğrenmesi ve veri tabanı disiplinlerinin birlikte çalışması ile 1990’da ilk veri madenciliği yazılımı geliştirilmiştir. 2000’li yıllarda yazı, ses, video gibi farklı türlerdeki veriler bulut ortamlarda depolanmaya başlanmıştır. İlerleyen yıllarda bu veri türlerine sürekli artan ve değişen hızda sosyal medya paylaşımları, log dosyaları, bloglar, konum bilgisi dâhil olmuştur. Hacim, hız ve çeşitlilik bakımından artan günümüz verileri Büyük Veri Analitiği yaklaşımına neden olmuştur.



Şekil 1: Veri madenciliğinin tarihsel gelişimi

Biriktirilen veriler arasından elde edilen değerli bilgilerin gelecek için tahmin fırsatı vermesi, ortaya çıkması muhtemel olumsuz durumlar için önceden tedbir alma olanağı tanınması veri madenciliğinin kullanımını önemli kılmış ve kullanım alanlarını genişletmiştir. Verilerin kaydedildiği her alanda veri madenciliği tekniklerini kullanmak mümkündür ve Şekil 2’de görüldüğü gibi veri madenciliği istatistik, yapay zekâ, makine öğrenmesi, veri

tabanı teknolojileri ve görselleştirme (veriyi doğru analiz etmek için görsel destek) gibi birçok disiplin ile bağlantılıdır. Her ne kadar makine öğrenmesi ve istatistik ile güçlü bağlara sahip olsa da gerçekleştirilen çalışma amacına göre diğer disiplinlerle de farklı derecelerde bağlantılar söz konusu olmaktadır.



Şekil 2: Veri madenciliği ile ilgili diğer disiplinler

Literatürdeki en yaygın tanımıyla “eğitim, bireyin davranışlarında kendi yaşantısı yoluyla kasıtlı olarak istendik yönde değişiklik meydana getirme süreci” olarak ifade edilmektedir (Ertürk 1997; Akt. Kartal ve Afacan 2011: 77). Bu süreci etkileyen faktörler göz önüne alındığında ise kolay yönetilebilen bir süreç olduğu söylenemez. Aynı zamanda eğitimi yalnızca öğretmen ve öğrenci arasındaki bir süreç olarak ifade etmek de doğru değildir. Aile, arkadaşlar, öğretmenler, okul ortamı, sınıf ortamı, dersler gibi faktörler bu sürecin gerçekleşmesi esnasında öğrenci üzerinde olumlu ya da olumsuz etkilere sahiptir. Bu noktada önemli olan öğrencinin bu faktörlerden hangisinden veya hangilerinden ne düzeyde etkilendiğinin belirlenmesidir. Ancak öğrencinin eğitim sürecine etki eden faktörler belirlendikten sonra gerekli önlemler alınarak eğitimin etkinliği artırılabilir. Bununla birlikte bu faktörlerin belirlenip gerekli önlemlerin alınması için değerlendirilmesi gereken veriler, oldukça çok ve karmaşıktır. Bu noktada toplanan verilerin analiz edilip değerlendirilmesi ve anlamlı verilere dönüştürülmesi de bir hayli zahmetlidir. Oldukça fazla olan bu verilerin anlamlı hale getirilip kullanılması için veri madenciliği yöntemlerinden faydalanılabilir. Bu nedenle Türel ve Baz(2016)’da veri yığınları üzerinden yapılacak irdelemelerde eğitim-öğretim sürecinde aktif rol alan katılımcıların ihtiyaçlarına cevap verecek veri madenciliği teknikleri eğitim sistemlerinde kullanılması gerektiğini ifade etmiştir. Çünkü veri madenciliği yöntemleri veri yığınlarının bulunduğu her alanda kullanılabilir. Nitekim eğitim alanında da veri madenciliği yöntemleri kullanılarak anlamlı ilişkilerin belirleneceği, geleceğe ışık tutabilecek bilgilerin türetileceği veri tabanları mevcuttur (Birtül, 2011). Özellikle öğrenme yönetim sistemleri, akıllı öğretim sistemleri, uyarlanabilir hipermedya sistemleri gibi 21. yy öğrenme ortamlarında birçok veri türü bulunmaktadır. Bununla birlikte bu ortamları kullanan tüm öğrenciler motivasyon, hazırbulunuşluluk, özdüzenleme gibi becerilerine ait farklılıklardan dolayı aynı öğrenme performansını sergilememektedir (Şahin, 2018). Bu durum 2005 yılında “Eğitim Verileri Madenciliği” olarak adlandırılan ilk atölye çalışmasına konu olmuş ve 2008’de Uluslararası Eğitim Verileri Madenciliği Çalışma Grubu tarafından

düzenlenen Uluslararası Eğitim Verileri Madenciliği Konferansı'nın kurulmasıyla sonuçlanmıştır (Romero ve Ventura, 2013).

Araştırmacılar tarafından eğitim sürecinde kullanılması tavsiye edilen eğitsel veri madenciliği ile ilgili tanımların birçoğuna bakıldığında, bu kavramın Narli, Aksoy ve Ercire (2014) tarafından “Eğitimsel veri tabanlarından elde edilen veriye, eğitimsel amaçlar doğrultusunda veri madenciliği yöntemlerinin uygulanmasıyla ortaya çıkan bir araştırma alanıdır.” şeklinde tanımlandığı görülmüştür. Diğer taraftan Uluslararası Eğitimsel Veri Madenciliği Topluluğu eğitsel veri madenciliğini, eğitim ortamlarındaki benzersiz veri türlerini araştırmak ve bu yöntemleri kullanarak öğrencileri ve öğrendiklerini daha iyi anlamak için yöntemler geliştirmek olarak tanımlamaktadır.



Şekil 3: Eğitsel veri madenciliğinin kullanım amaçları

Eğitsel veri madenciliğinin kullanım amaçları, Romero ve Ventura (2013)'ın araştırmasından uyarlanan Şekil 3'te gösterilmektedir. Bu şekilden yola çıkılarak eğitsel veri madenciliği ile eğitim sürecinin en önemli ögesi olan öğrencilerin kişisel bilgileri, not durumları, devam devamsızlık bilgileri, başarılı ve başarısız olduğu dersler gibi verilerine farklı modeller uygulanarak başarı durumunun sebeplerinin belirlenmesi, başarı durumlarının yükseltilmesi, devamsızlıklarının önüne geçilmesi, alacakları derslerin seçiminde ve kariyer hedeflerine ilişkin tavsiyelerde bulunulması sağlanabilir (Rizvi, Rienties ve Khoja, 2019). Ayrıca öğrencilerin kişisel özelliklerine, bireysel öğrenme benzerliklerine göre grupların oluşturulmasında, düşük motivasyon, devamsızlık, okulu bırakma, okul kurallarına uymama gibi istenmeyen öğrenci davranışlarının tespit edilmesinde, önceden tahmin edilmesinde ve gerekli tedbirlerin alınmasında da eğitimde veri madenciliği tekniklerinden yararlanılmaktadır (Aksoy, 2014). Eğitim öğretim ortamında ortaya çıkabilecek sorunların önceden tespit edilmesi ve eğitim ve öğretim ortamlarının iyileştirilmesi gerçekleştirilebilir (Can, 2017). Eğitimciler için de eğitimde veri madenciliği teknikleri kullanılarak öğrencilerin öğrenme

sürecini daha verimli hale nasıl getirebilecekleri, düzeltici önlemleri almalarını sağlayacak geri bildirimlere ulaşılabilir (Yakupoglu, 2018).

Eğitim alanında da öğretmenlere, öğrencilere ve yöneticilere yol gösterici faydalı bilgileri ortaya koyacak doğru bir eğitim süreci planlamada veri yığınlarının doğru bir şekilde irdelenmesi oldukça önemlidir. Dolayısıyla günümüzde pazarlama, sağlık, endüstri, bilişim, mühendislik, telekomünikasyon, borsa, biyoloji gibi pek çok alanda kullanılan veri madenciliğinin son zamanlarda önemli kullanım alanlarından biri de eğitim/öğretim alanı olarak karşımıza çıkmaktadır. Psikometri, öğrenme analitiği, istatistik gibi farklı birçok disiplini bir arada kullanan eğitsel veri madenciliğinden (Türel ve Baz, 2016), öğrenci başarısı ile ilgili tespitlerde bulunmak, eğitim-öğretim ortamındaki aksaklıkları ve bunların nedenleri ile ilgili çıkarımlarda bulunmak, doğru ve ihtiyacı karşılayan eğitim-öğretim ortamlarının oluşturulması amacıyla yararlanılabilmektedir (Özbay, 2015). Eğitim anlamında geniş bir çalışma alanı sunan veri madenciliği kavramı ile ilgili araştırmaların sayısı ulusal ve uluslararası alan yazında her geçen gün artmasına karşın yerli kaynaklarda henüz doyuma ulaşmamış ve yeterli sayıda çalışma ortaya konulmamıştır. Bu nedenle alanda yapılmış çalışmalar eğitimde veri madenciliği disiplininin kullanım durumunu ortaya koyması ve alandaki eksikliklerin belirlenmesi açısından önemlidir. Ayrıca, bu çalışmalar alanın derinlemesine incelenmesine ve tanınmasını sağladığı için gelecekteki çalışmalarında yönelim alanını belirlemesi açısından araştırmacıların fikir sahibi olması sağlamaktadır. Bu çalışmada öncelikle veri madenciliği ile ilgili kuramsal çerçeve verilmiş daha sonra alanda yapılmış araştırmaların dünyadaki ve Türkiye’deki durumu ortaya konulmuştur.

2. YÖNTEM

Bu çalışma kapsamında, eğitsel veri madenciliği yöntemini konu alan ve bununla ilgili örnekler sunan Türkiye’de ve uluslararası yapılmış makale ve tezler incelenmiştir. Bu amaç doğrultusunda tarama modeliyle yapılan çalışma, nitel araştırma yöntemine göre desenlenmiştir. Veri madenciliği, eğitsel veri madenciliği ve veri madenciliğinin alt başlıkları olan karar ağaçları, kümeleme, birliktelik analizi gibi kavramları içeren Türkçe ve İngilizce çalışmalar incelenmiştir. İncelenen makale ve tezler doküman analizi tekniği önceki bölümde belirtilen kriterler açısından analiz edilmiştir. “Doküman analizi, üzerinde çalışma yapılmak istenen olgu ve bu olgular ile ilgili bilgi sunan yazılı materyallerin incelenmesini temel alır” (Yıldırım ve Şimşek, 2008 akt. Özgür ve Akgün, 2018). Bu çalışmada herhangi bir yıl sınırlaması olmadan Google Akademik, EBSCO, Elsevier, ERIC ve YÖKTez tarama merkezinde eğitsel veri madenciliği, karar ağaçları, kümeleme, veri madenciliği anahtar kelimeleri ile ulaşılmış olan toplam 102 çalışma incelenmiştir.

2.1. Veri Toplama Aracı

Bu araştırma sürecinde verileri toplamak için Göktaş ve arkadaşları (2012) tarafından eğitimde teknolojilerindeki eğilimleri tespit etmek amacıyla yürütmüş oldukları çalışma kapsamında geliştirilen makalenin künyesi, türü, konusu, metodolojisi, veri toplama araçları, veri analiz yöntemi ve örneklem (boyutu, düzeyi ve örneklem seçimi) başlıklarını içeren “Eğitim Teknolojileri Yayın İnceleme Formu” incelenmiş olup araştırmanın kapsamı doğrultusunda bazı başlıkların çıkarılması (Makalenin Konusu, Makale Türü, verilerin

analizi) ve eklenmesi (Yayın Türü, veri madenciliği metodu, veri madenciliği ile incelenen durum) sonucu güncellenerek son şekli verilmiş ve kullanılmıştır.

Google Akademik, EBSCO, Elsevier, ERIC ve YÖKTez tarama merkezi üzerinden ulaşılan makale ve tezler;

- ✓ eğitimde veri madenciliği metotlarının kullanımına yönelmiş olması,
- ✓ veri madenciliği, eğitsel veri madenciliği, kümeleme, karar ağacı, birliktelik analizi gibi konuları işlemiş olması,
- ✓ yüksek lisans tezi, makale ve bildiri olması,
- ✓ İngilizce veya Türkçe Tam metin olması

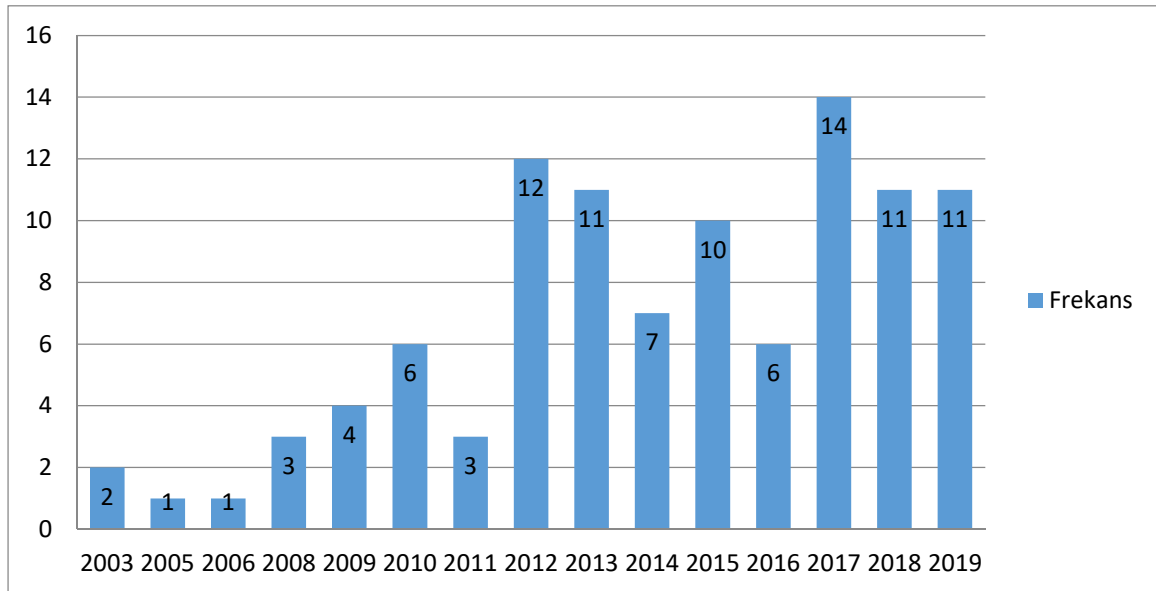
ölçütleri doğrultusunda araştırmaya dâhil edilmiş ve yayın yılı, araştırma yöntemi, örneklem düzeyi, kullanılan veri toplama aracı, kullanılan veri madenciliği yöntemi bakımından incelenmiştir.

3. BULGULAR

Türkiye ve Dünya’da eğitsel veri madenciliği araştırmaların durumu ile ilgili elde edilen bulgular ise bu başlık altında toparlanarak sunulmuştur.

3.1. Araştırmaların Yıllara Göre Dağılımı

Araştırma kapsamında yapılmış olan çalışmaların genel olarak 2008-2019 yılları arasında yayınlandığı gözlenmiş olup incelenen 102 çalışmanın yayın yıllarına göre dağılımı Şekil 4’te gösterilmiştir.



Şekil 4: Çalışmaların yayın yılına göre dağılımı

Şekil 4’te de görüldüğü üzere 2008 yılından 2019 yılına kadar geçen sürede her sene araştırmacılar tarafından çalışmaların ortaya konulduğu bununla birlikte en fazla çalışmanın

ise 2017 yılında yapıldığı ($f=8$) sonucuna ulaşılmış bu sonucu ise 12 çalışma ile 2012 yılı takip etmiştir. 2008 öncesinde yapılan çalışmalar incelendiğinde ise 2003, 2005 ve 2006 verildiği gözlenmiş olup bunlar ise diğer ($f=4$) kategorisi altında incelemeye tabi tutulmuştur.

3.2.Araştırmaların Dergi Eğilimleri

Araştırma kapsamında incelenen eğitsel veri madenciliği konusunu işlemiş olan 102 çalışmanın yayınlandığı dergi listesi aşağıda Tablo 1’ de sunulmuştur.

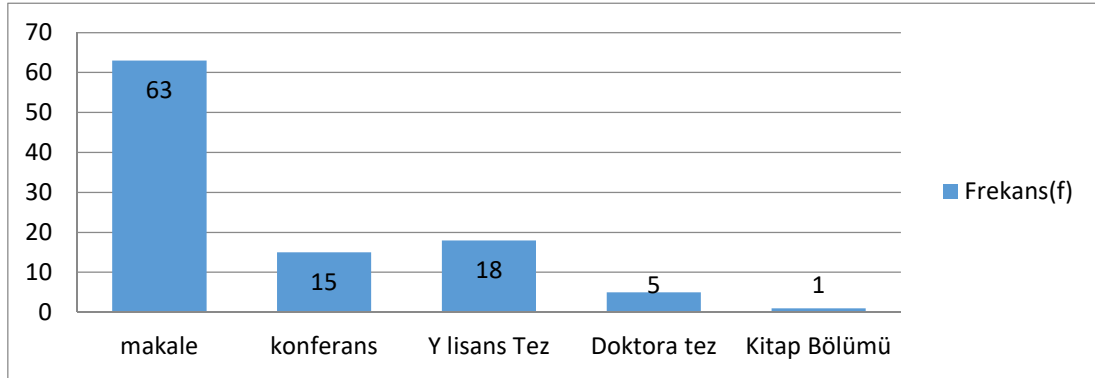
Tablo 1. Eğitsel veri madenciliği konulu çalışmalara ilişkin dergi listesi

Dergi Adı	Sayı
Bilişim Teknolojileri Dergisi	2
Trakya Üniversitesi Sosyal Bilimler Dergisi	1
Computers in Human Behavior	5
100.yıl Üniversitesi Eğitim Fakültesi Dergisi	1
Telematics and Informatics	2
Mediterranean Journal of Humanities	1
PIVOLKA Dergisi	1
YÖKTez Tarama Merkezi	23
Atatürk Üniversitesi İktisadi Ve İdari Bilimler Dergisi	1
Elektrik-Elektronik,Bilgisayar Ve Biyomedikal Müh. Sempozyumu	1
International Symposium On Digital Forensks And Security	1
Gazi üniv. iktisadi ve idari bilimler Dergisi	1
ITTES-2016	2
Politeknik Dergisi	1
Internet and Higher Education	2
Inesjournal Dergisi	1
Social and Behavioral Sciences	3
GEFAD	1
Computers & Education	11
İstanbul Ticaret Üniv. Fen Bilimleri Dergisi	1
Procedia Computer Science	7
Turkish Journal Of Computer And Mathematics Education	1
Eğitim ve Öğretim Araştırmaları Dergisi	1
Fırat Üniv. Mühendislik Bilim. Dergisi	1
Diğer	30
Toplam	42

Tablo 1’e bakıldığında 2003-2019 yılları arasında konu ile ilgili olarak birçok farklı dergide çalışmaların yayınlanmış olduğu görülmektedir. Makale, konferans bildirisi, tez veya kitap bölümü olarak farklı yayın türünde olan bu çalışmaların dergilere düzenli bir şekilde dağıldığı görülse de bazı ulusal veya uluslararası olarak ön plana çıkan dergilerinde olduğu söylenebilir. Öyle ki makale türünde “Computers & Education” dergisinde sadece 11 çalışmanın yayınlandığı ve yine “Procedia Computer Science” tarafından ise yedi tane konferans bildirisinin literatüre kazandırıldığı gözlemlenmiştir. Bununla beraber makale ve bildirilerin aksine Türkiye’de yapılmış olan doktora ve yüksek lisans düzeyindeki bütün tez çalışmalarına ise YÖKTez üzerinden ulaşılmıştır. İncelenen tez çalışmalarının ise dört tanesi Fırat Üniversitesi’nde iken onu Marmara (2) ve Hacettepe (2) Üniversitesi takip etmiştir.

3.3. Araştırmaların Yayın Türü Eğilimleri

Google Scholar (Akademik), EBSCO, Elsevier, ERIC ve YÖKTez tarama merkezi veri tabanları üzerinden araştırma kapsamına dâhil edilen 102 çalışmanın yayın türü dağılımı Şekil 5’ de gösterilmiştir.

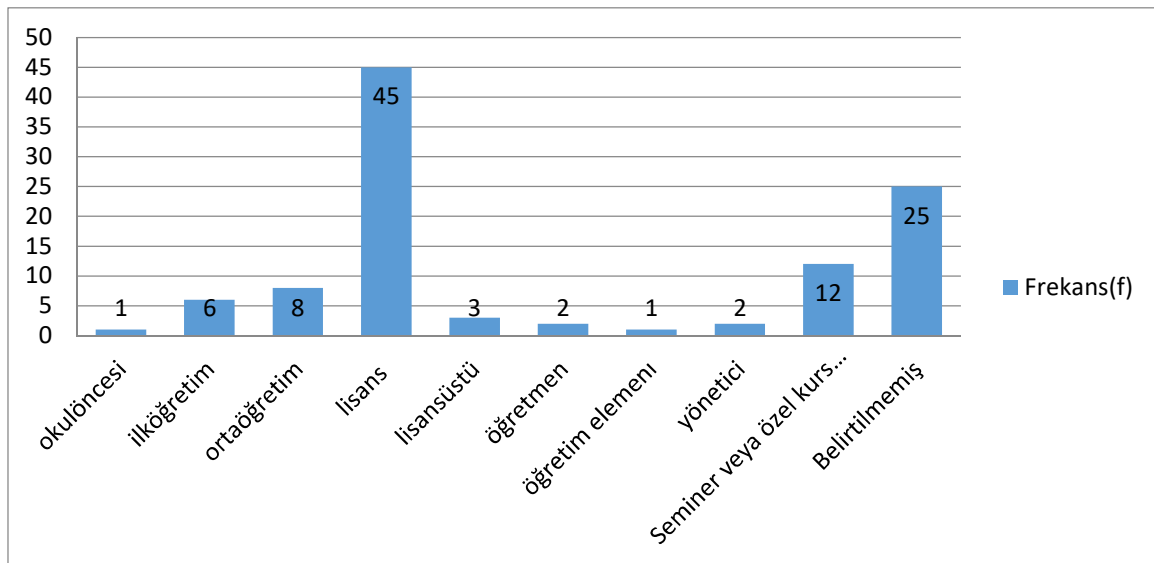


Şekil 5: Çalışmaların yayın türlerine göre dağılımı

Şekil 5’de de görüldüğü gibi incelenen 102 çalışma arasında %38.0 ($f=63$)’lük oranla makaleler baskın gelirken onu %10.8 ($f=18$)’lük oranla yüksek lisans tezleri takip etmektedir. Bununla beraber incelenen çalışmalar içerisinde en küçük dilim ise %0.6 ($f=1$)’lük oran ile kitap bölümü olmuştur.

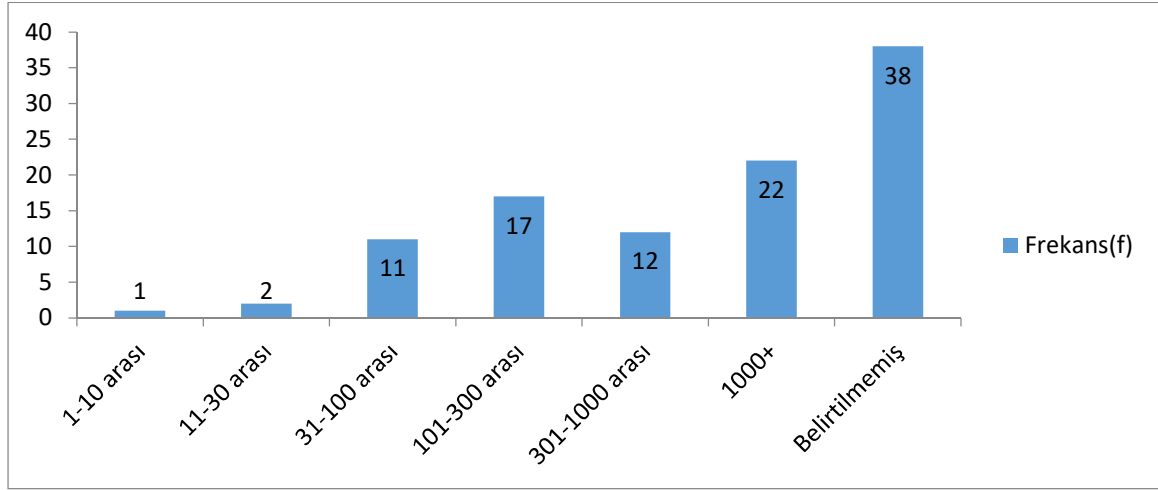
3.4. Araştırmaların Örneklem Sayısı, Düzeyi Ve Seçim Şekillerindeki Eğilimler

Araştırma kapsamında çalışmalarda ele alınarak verilerin toplama noktasında araştırmacıların tercih ettiği örneklem düzeylerinin kullanım sıklığını gösteren bulgular Şekil 6’da gösterilmiştir.



Şekil 6: Çalışmalarda örneklem düzeylerinin seçilme sıklıkları

Araştırma amacı doğrultusunda incelemeye alınan çalışmalarda araştırmacılar tarafından farklı düzeylerden örneklem seçimi yapılmış olsa da çalışmaların daha çok lisans düzeyinde öğrenciler (%27.1, $f=45$) ile yürütüldüğü görülmüştür. Bunun yanında alan yazın derlemesi veya bir yazılım geliştirmek amacıyla yapılan çalışmaların büyük bir çoğunluğunda ise herhangi bir örneklem düzeyinden bahsedilmemiştir. Bu çalışmaların tüm incelenen çalışmalar içerisindeki oranı ise %15.0 ($f=25$) 'dir. Farklı örneklem düzeyinde gerçekleştirilen çalışmaların örneklem büyüklüğü ile ilgili oranlar ise Şekil 7'de paylaşılmıştır.

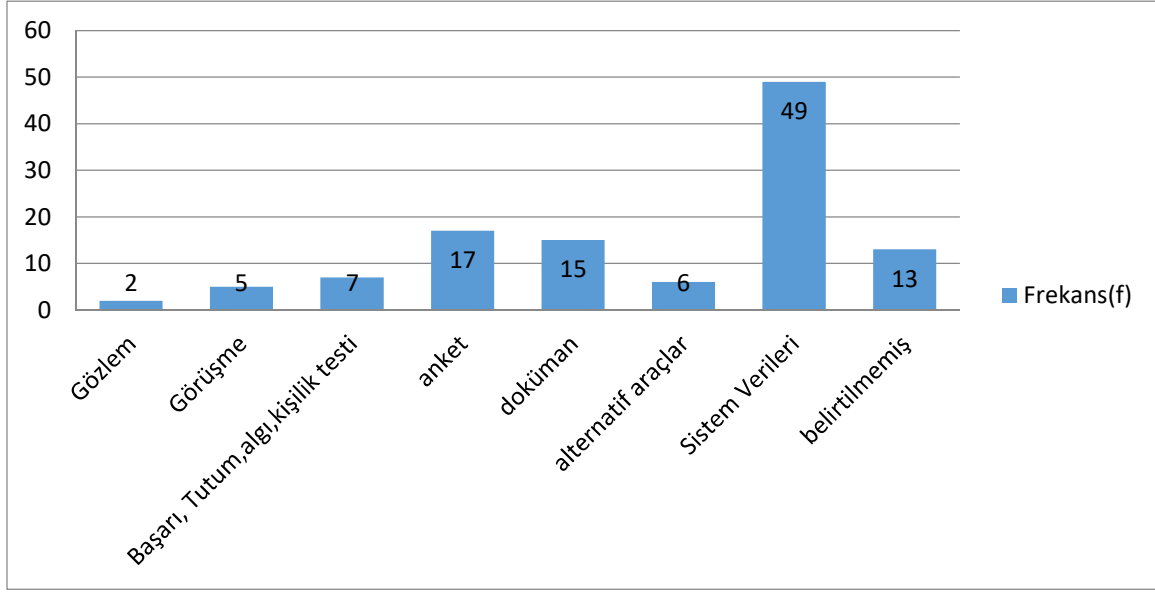


Şekil 7: Çalışmalardaki örneklem sayıları dağılımı

Seçilen örneklem boyutu ile ilgili olarak Şekil 7 incelendiğinde genel anlamda büyük sayılar dikkat çekerken (1000 üstü) en yüksek ayı ise örneklem kullanılmayan literatür derleme ve örneklem sayısının belirsiz olduğu çalışmaların birlikte ele alındığı belirtilmemiş kategorisine aittir ($f=38$).

3.5. Araştırmalarının Kullanılan Veri Toplama Araçlarındaki Eğilimler

Araştırma kapsamında incelenen 102 çalışmada kullanılan veri toplama araçlarının tercih edilme sıklığı Şekil 8'de gösterilmiştir. Birden fazla veri toplama aracı kullanılması halinde ise her bir araç ayrı ayrı ele alınarak bulgular tespit edilmiştir.

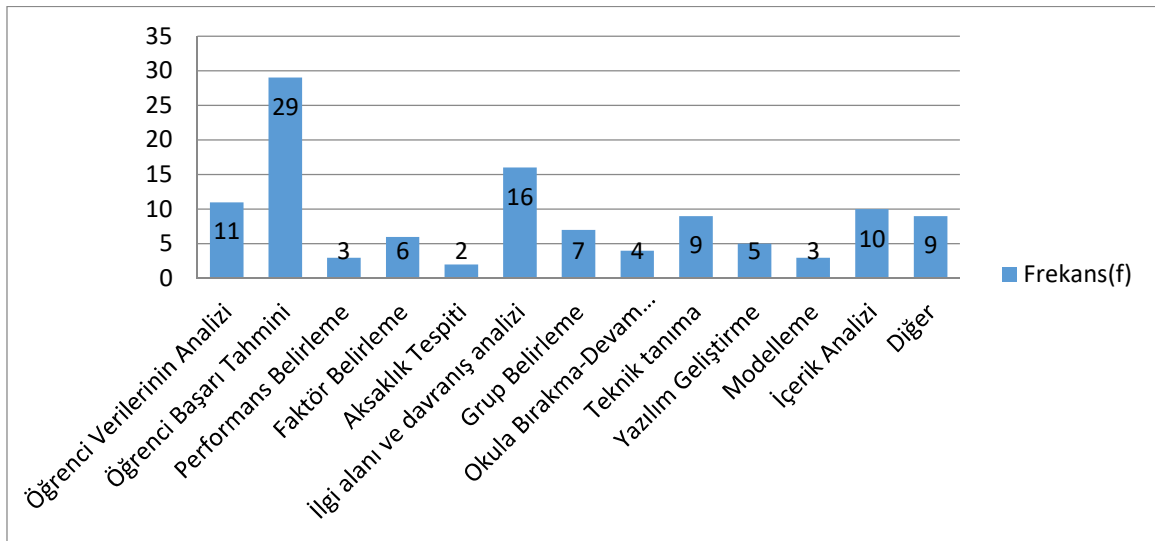


Şekil 8: Çalışmalarda kullanılan veri toplama araçları

Şekil 8’de görüldüğü gibi veri madenciliği ile ilgili incelenen çalışmalar içerisinde veri toplama aracı olarak en az kullanılan araç ise gözlem ve başarı testi olurken sistem verileri ise %29.5’lik ($f=49$) oranla en fazla kullanılırken onu %10.2’lik ($f=17$) oran ile anket takip etmektedir. Diğer taraftan incelenen çalışmaların 13 tanesinde ise kullanılan veri toplama aracı ile ilgili bilgi paylaşılmamıştır.

3.6. Araştırmaların Konu Eğilimleri

Eğitsel veri madenciliği yöntemleri kullanılarak yapılan çalışmalar ele alınan konu başlıkları ile ilgili ulaşılan bulgular aşağıda Şekil 9’da gösterilmiştir.

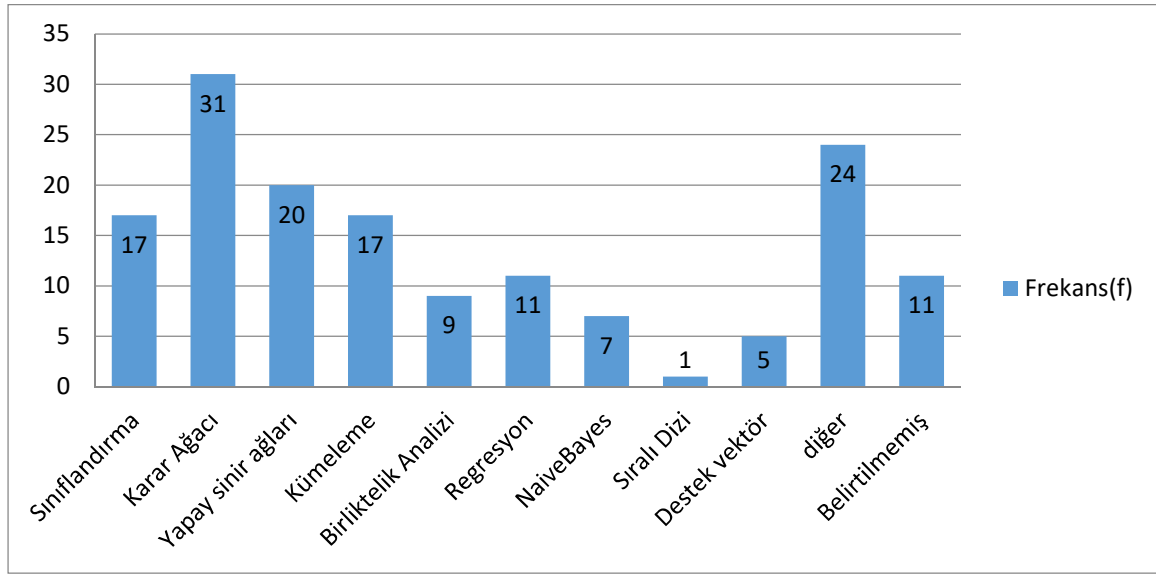


Şekil 9: Çalışmalarda incelenen konuların dağılımı

Şekil 9’da görüldüğü gibi eğitsel veri madenciliği ile ilgili incelenen 102 çalışma içerisinde en fazla işlenen konu %17,5 ‘luk ($f=29$) oranı ile başarı tahmini, amacıyla yapılan çalışmalar olurken onu %9.6 oran ile davranış analizi ve bir alana yönelik ilgi veya tutumu belirlemek amacıyla yapılan çalışmalar takip etmektedir.

3.7. Araştırmalarda Tercih Edilen Veri Madenciliği Yöntemi Eğilimleri

Eğitsel veri madenciliği yöntemleri kullanılarak yapılan çalışmalar da kullanılan veri madenciliği tekniklerin dağılımı ile ilgili ulaşılan bulgular aşağıda Şekil 10’da gösterilmiştir.

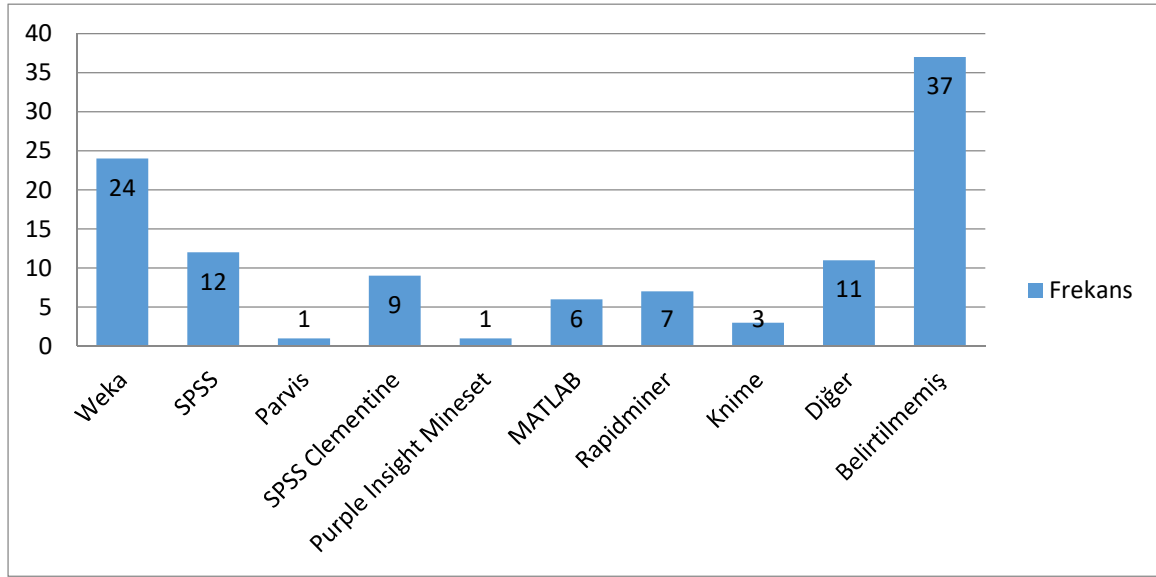


Şekil 10: Veri madenciliği tekniklerinin dağılımı

Şekil 10’da görüldüğü gibi veri madenciliği kullanılarak yapılan çalışmalarda birçok farklı yöntem ve algoritma ile karşılaşmıştır. Bununla birlikte yapılan çalışmalarda genellikle birden fazla teknik bir arada kullanılmış en sık kullanılan algoritmalar ayrı kategorilerde değerlendirilirken nadiren kullanımların görüldüğü algoritmalar ise diğer (Link analizi, Apriori algoritması, Klasik İstatistiksel Tespitler, Aykırı Değer Tespiti, Metin Madenciliği, Duygu Analizi, J4.8, Ardışık Örüntü Keşfi vs.) şeklinde kategorize edilmiş ve yüzdesel olarak grafikte %14.5’lik bir yer tutmuştur. Bununla birlikte çalışmalarda en fazla kullanılan teknik %18.7 ($f=31$) ile karar ağaçları olmuş onu ise %12.2 ($f=20$) ile yapay sinir ağı ve %10.2 ($f=17$) ile kümeleme yöntemleri takip etmiştir.

3.8. Araştırmalarda Kullanılan Veri Madenciliği Araçlarının Eğilimleri

Veri madenciliği yöntemleri uygulanarak gerçekleştirilen çalışmalar da kullanılan veri madenciliği araçlarının dağılımı ile ilgili ulaşılan bulgular aşağıda Şekil 11’de gösterilmiştir.



Şekil 11: Veri madenciliği araçlarının dağılımı

İncelenen çalışmalar içerisinde toplanan verilerin veri madenciliği yöntemleri ile analizi sürecinde araştırmacılar tarafından en sık tercih edilen araç WEKA ($f=24$) olurken onu SPSS ($f=9$) takip etmektedir. Bununla birlikte kullanılan aracın net bir şekilde belirtilmediği çalışma sayısı 37'dir. Bu çalışmalar içerisinde yer alan çalışmaların 12 tanesi ise alan yazın tarama türünde olan çalışmalardır.

4. SONUÇ VE ÖNERİLER

Teknolojide yaşanan gelişmeler ile birlikte artan veri yığınlarının hızlı bir şekilde çözümlenmesinde önemli derecede yardımcı olan veri madenciliği disiplininin ilk ortaya çıktığı andan günümüze kadar geçen tüm tarihlerde sürekli olarak bir gelişim ve değişim içerisinde olduğu görülmektedir. Bu noktada alandaki eğilimlerin belirlenmesi de sonraki araştırmalara rehber olması açısından önem kazanmaktadır. Bu çalışmada eğitsel veri madenciliği yöntemi konulu çalışmaların incelenerek alandaki eğilimlerin ortaya konulması amacıyla gerçekleştirilen bu araştırmada ilgili konunun ele alındığı 102 çalışma, araştırmacılar tarafından içerik analizi yöntemi ile incelenerek analiz edilerek mevcut durum ortaya konmuştur.

Çalışmada elde edilen bulgulara göre, eğitsel veri madenciliği araştırmalarının tarihleri 2003-2019 tarihleri arasında değişmekte olup bu aralıkta yapılan çalışma sayısının genel olarak artış eğilimi gösterdiği belirlenmiştir. Çalışmalarda kullanılan veri toplama araçlarına ilişkin olarak araştırmacı eğilimleri incelendiğinde en dikkat çekici olanlarının ise farklı veri tabanları ve öğrenme yönetim sistemleri vs. üzerinde depolanan sistem verilerinin kullanıldığı bunu ise anketlerin takip ettiği görülmüştür. Yapılan çalışmalarda bu veri toplama araçlarının bu kadar fazla tercih edilmesinin nedeni az sürede ve hızlı bir şekilde çok fazla veriye ulaşma çabası kaynaklandığı düşünülebilir. Diğer taraftan çalışmanın ana odağı olan eğitsel veri madenciliği alanında gerçekleştirilen çalışmaların konu bazında öğrenci başarısını tahmin etme üzerinde yoğunlaşmıştır. Bunu ise kişilerin bir durum karşısında sergilediği tutum veya

davranışı analiz etmek amacı ile yapılan çalışmaların takip ettiği görülmüştür. Uygulamaya yönelik çalışmalarda en sık kullanılan tekniğin karar ağaçları olduğu ve bu tekniği yapay sinir ağlarının takip ettiği görülmüştür. Ancak eğitsel veri kümelerinin her geçen gün daha da çok artmasına karşın literatürde yol gösterici nitelikteki çalışmaların sayısının oldukça eksik olduğu görülmüştür. Bu nedenle araştırmacılar tarafından farklı değişkenler üzerinden farklı bakış açıları derinleştirilerek yeni çalışmaların yapılması önerilebilir. Yine yapılan analizler sonucunda eğitsel veri madenciliği çalışmalarının genel olarak birbirinin tekrarı nitelikte olduğu tespit edilmiştir. Bu nedenle eğitim bazında araştırma yapılmamış yeni konu başlıkları üzerinden çalışma sayısı artırılarak alanda var olan eksiklik giderilebilir.

KAYNAKLAR

- Aksoy, E. (2014). Matematik alanında üstün yetenekli ve zekalı öğrencilerin bazı değişkenler açısından veri madenciliği ile belirlenmesi. (Yayınlanmamış yüksek lisans tezi). Dokuz Eylül Üniversitesi, İzmir.
- Atalay, M., ve Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ Ve Makine Öğrenmesi Uygulamaları. Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 9(22), 155-172.
- Ayık, Y. Z., Özdemir, A., ve Yavuz, U. (2007). Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkinin Veri Madenciliği Tekniği İle Analizi. Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 10(2), 441-454.
- Bırtıl, S (2011). Kız meslek lisesi öğrencilerinin akademik başarısızlık nedenlerinin veri madenciliği tekniği ile analizi. Yayınlanmamış yüksek lisans tezi. Afyon Kocatepe Üniversitesi, Afyonkarahisar.
- Can, E. (2017). Temel Eğitimden Ortaöğretime Geçiş Sınavı Kazanımlarının Veri Madenciliği Yöntemleri İle Değerlendirilmesi. (Yayınlanmamış yüksek lisans tezi). Afyon Kocatepe Üniversitesi, Afyonkarahisar.
- Çalış, A., Kayapınar, S., ve Çetinyokuş, T. (2014). Veri Madenciliğinde Karar Ağacı Algoritmaları İle Bilgisayar Ve İnternet Güvenliği Üzerine Bir Uygulama. Journal Of Industrial Engineering (Turkish Chamber Of Mechanical Engineers), 25.
- Göktaş, Y., Küçük, S., Aydemir, M., Telli, E., Arpacık, Ö., Yıldırım, G. ve Reisoğlu, İ. (2012). Türkiye’de eğitim teknolojileri araştırmalarındaki eğilimler: 2000-2009 dönemi makalelerinin içerik analizi. Kuram ve Uygulamada Eğitim Bilimleri Dergisi, 12(1), 177-199.
- İrmak, S. (2009), Veri Madenciliği Yöntemleri ile Sağlık Sektörü Veri tabanlarında Bilgi Keşfi: Tanımlayıcı ve Kestirimci Model Uygulamalar. (Yayınlanmamış Doktora Tezi). Akdeniz Üniversitesi, Antalya.
- Karabatak, M. ve Çelik, Y. (2016). Eğitsel Bir Veri Madenciliği Aracının Geliştirilmesi. 4. Uluslararası Öğretim Teknolojileri ve Öğretmen Eğitimi Konferansı Bildirileri, 6-8 Ekim, Fırat Üniversitesi, Elazığ, ss.264-269.
- Kartal, T. ve Afacan, Ö. (2011). Pedagojik formasyon eğitimi alan öğretmen adaylarının mesleğe ilişkin tutumlarının incelenmesi. Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 1(24), 76-96.

- Narlı, S., Aksoy, E., ve Ercire, Y. E. (2014). Veri madenciliği ile ilköğretim matematik öğretmen adaylarının öğrenme stillerinin ve aralarındaki ilişkilerin incelenmesi. *International Journal of Educational Studies in Mathematics*, 1(1), 37-57.
- Özbay, Ö. (2015). Öğretim Yönetim Sistemi Üzerinde Üniversite (Lisans) Düzeyindeki Öğrenci Hareketliliğinin Veri Madenciliği Yöntemleriyle Analizi. (Yayınlanmamış yüksek lisans tezi). Başkent Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Özbay, Ö. (2015). Veri madenciliği kavramı ve eğitimde veri Madenciliği uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi*, (5), 262-272.
- Rizvi, S., Rienties, B. ve Khoja, S. A. (2019). The role of demographics in online learning; A decision tree based approach. *Computers & Education*, 137, 32-47.
- Romero, C., López, M. I., Luna, J. M. ve Ventura, S. (2013). Predicting students' final performance from participation in on-line discussion forums. *Computers & Education*, 68, 458-472.
- Romero, C. ve Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12-27.
- Savaş, S., Topaloğlu, N., ve Yılmaz, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.
- Şahin, M. (2018). E-öğrenme ortamlarına yönelik öğrenme analitiklerine dayalı müdahale motoru tasarımı ve geliştirilmesi. (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Şengür, D., ve Tekin, A. (2013). Öğrencilerin Mezuniyet Notlarının Veri Madenciliği Metotları İle Tahmini. *Bilişim Teknolojileri Dergisi*, 6(3), 7-16.
- Taşdemir, M. (2012). Veri Madenciliği (Öğrenci Başarısına Etki Eden Faktörlerin Regresyon Analizi İle Tespiti). (Yayınlanmamış yüksek lisans tezi). Dicle Üniversitesi Sosyal Bilimler Enstitüsü, Diyarbakır.
- Türel, Y. K., ve Baz, E. (2016). Eğitsel Veri Madenciliği Üzerine Bir Araştırma. 4. Uluslararası Öğretim Teknolojileri ve Öğretmen Eğitimi Konferansı Bildirileri, 6-8 Ekim, Fırat Üniversitesi, Elazığ, ss.54-64.
- Yakupoğlu, Y.(2018). Eğitimsel veri madenciliği ve bir uygulaması. (Yayınlanmamış yüksek lisans tezi). İstanbul Teknik Üniversitesi, İstanbul.

YAYIN SINIFLAMA FORMU

A-MAKALENİN KÜNYESİ			
1. Makalenin Adı:.....			
2. Yazarı/ları.....			
3.Derginin Adı:.....			
.....			
4. Yıl:.....			
B-YAYIN TÜRÜ			
1. Makale () 2. Konferans () 3. Doktora Tezi () 4. Y. Lisans Tezi ()			
C-MAKALENİN YÖNTEMİ			
1. Nicel () 2. Nitel () 3. Karma () 4. Alan Yazın Derl. () 5.Diğer ()			
6.Analiz yapılmamış()			
D-VERİ TOPLAMA ARAÇLARI			
1. Gözlem () a. Katılımcı b. Katılımcı olmayan	2.Görüşme/ () a.Yapılandırılmış() b.Yarı-yapılandırılmış () c.Yapılandırılmamış () d.Online görüşme ()	3. Başarı Testleri () a.Açık uçlu () b.Çoktan seçmeli () c.Diğer ()	4. Tutum, algı, kişilik veya yetenek testleri () a. Açık uçlu () b.Çoktan seçmeli () c. Likert () d. Diğer ()
5. Anket () a. Açık uçlu b. Çoktan seçmeli c. Likert d. Diğer	6. Döküman ()	7.Alternatif Araçlar () (performans testleri, tanılayıcı testler, kavram haritaları, portfölyo)	8.Diğer (yazınız)....
E- ÖRNEKLEM			
Örneklem Düzeyi 1. Okul Öncesi 2. İlköğretim 3. Ortaöğretim 4. Lisans ve lisans mezunu 5. Lisansüstü(Master-Doktor) 6. Öğretmenler 7. Öğretim elemanları 8. Veliler 9. Yöneticiler 10. Diğer.....	Örneklem Sayısı 1. 1-10 arası () 2. 11-30 () 3. 31-100 () 4. 101-300 () 5. 301-1000 () 6. 1000'den fazla ()	Örneklem Seçim Şekli 1. Rastgele () 2. Kolay ulaşılabilir örnekleme () 3. Amaca uygun () 4. Evrenin tamamı () 5. Diğer ()	
G-VERİ MADENCİLİĞİ METODU			

1.Sınıflandırma () 2.Karar Ağacı () 3.Bayes Sınıflandırma () 4.Yapay sinir ağları () 5.Zaman Serisi Analizi () 6.Kümeleme ()	7.Birliktelik Analizi() 8.Regresyon () 9.Sıralı Dizi Analizi () 10.Diğer () 11. Analiz yapılmamış ()
F- VERİ MADENCİLİĞİ İLE İNCELENEN DURUM	
1. Öğrenci Verilerinin Analiz () 2. Öğrenci Başarı () 3. Öğrenci Başarılarının Arttırılması () 4. Eğitim-Öğretim Ortamlarındaki Aksaklıkların Tespiti () 5. Etkili Eğitim-Öğretim Ortamları Oluşturma () 6. Devam Durumu Tahmini ()	