

Veri Madenciliğinin Eğitim Alanında Kullanımına İlişkin Araştırmaların Bibliyometrik Analizi

Fatih Balaman*, Sevil Hanbay Tiryaki**

Makale Geliş Tarihi:21/12/2024

Makale Kabul Tarihi:06/05/2025

DOI: 10.35675/befdergi.1605165

Öz

Bu çalışmanın amacı veri madenciliğinin eğitim alanında kullanımına ilişkin yapılmış çalışmaların bibliyometrik analiz yöntemi ile incelenmesidir. Analiz sonuçlarına göre eğitim bilimleri alanında veri madenciliği ile ilgili en fazla yayın yapan dergi Education and Information Technologies'dir. En fazla sayıda yayın yapan yazar Xing iken en çok atıf alan yazar Romero'dur. En çok yayın sayısına ve atıfa sahip ülke ABD'dir. Alandaki eğilimi belirlemek için yapılan trend konu analizi sonuçlarına göre ise 2013 sonrasında çalışılan konuların arttığı ve kavramların değiştiği, 2015 yılı ve sonrasında e-öğrenme, öğrenme yönetim sistemleri, makine öğrenimi, davranış analizi, gibi farklılaşan konuların çalışıldığı anlaşılmıştır. 2024 yılının alandaki trend konularının ise yapay sinir ağları, cart karar ağacı ve iş birliğine dayalı filtreleme olduğu sonucuna ulaşılmıştır. Ayrıca yapılan tematik alan analizi ile alan için önemli olan "e-öğrenme", "bulut bilişim", "büyük veri", "eğitimsel veri madenciliği", "öğrenme analitikleri", "kümeleme" ve "sınıflandırma" gibi kavramların yeterince gelişmemiş olduğu tespit edilmiştir. Araştırmacılara bu alanlarda çalışma yapmaları önerilmektedir.

Anahtar Kelimeler: Eğitimsel veri madenciliği, büyük veri, örüntü, bibliyometrik analiz

Bibliometric Analysis of Research on the Use of Data Mining in the Field of Education

Abstract

The aim of this study is to examine the studies on the use of data mining in the field of education by bibliometric analysis method. According to the results of the analysis, the journal with the highest number of publications on data mining in the field of educational sciences was Education and Information Technologies. While Xing is the author with the highest number of publications, Romero is the author with the highest number of citations. The country with the highest number of publications and citations is the USA. According to the results of the trend subject analysis conducted to determine the trend in the field, it was concluded that the subjects studied increased and the concepts changed after 2013, and differentiated subjects such as e-

* Hatay Mustafa Kemal Üniversitesi, Eğitim Fakültesi, BÖTE Bölümü, Hatay, Türkiye,

fatihbalaman2010@gmail.com, ORCID: [0000-0003-2175-0778](https://orcid.org/0000-0003-2175-0778) 

** Milli Eğitim Bakanlığı, İl Genel Meclisi Anadolu Lisesi, Gaziantep, Türkiye,

sevilhanbay90@gmail.com ORCID: [0000-0003-4780-9715](https://orcid.org/0000-0003-4780-9715) 

Kaynak Gösterme: Balaman, F., & Hanbay Tiryaki, S. (2025). Veri madenciliğinin eğitim alanında kullanımına ilişkin araştırmaların bibliyometrik analizi, *Bayburt Eğitim Fakültesi Dergisi*, 20(46), 1088-1115.

learning, learning management systems, machine learning, behavior analysis were studied in 2015 and later. In 2024, it was concluded that the trending topics in the field are artificial neural networks, cart decision tree and collaborative filtering. Nevertheless, with the thematic domain analysis, it was concluded that concepts such as “e-learning”, “cloud computing”, “big data”, “educational data mining”, “learning analytics”, “clustering”, “classification”, which are important for the field, are not sufficiently developed and can be further studied.

Keywords: Educational data mining, big data, pattern, bibliometric analysis

Giriş

Bibliyometrik analiz; herhangi bir konudaki eğilimleri, literatürdeki mevcut durumu, yapılmış araştırmalar arasındaki ilişkileri irdeleyen bir analiz tekniğidir. Bu teknik ile kodlama bilgisine sahip olmadan programın görsel arayüzü sayesinde analizler yürütülebilir (Dereli, 2024). Doğası gereği nicel verileri analiz eden bu yöntem ile genellikle dergi yayınları analiz edilmektedir (Broadus, 1987; Hall, 2011).

Bibliyometrik analiz; bir konu alanının güncel durumu hakkında belli görselleştirme yazılımları kullanılarak konu alanı hakkındaki trendleri izlemeyi kolaylaştırmaktadır. Bibliyometrik analizler ile belirlenen konudaki araştırmalara yönelik nicel veriler ve ölçüm göstergeleri elde etmek amaçlanır. Bibliyometrik analizler, araştırmacılara ilham vererek yeni çalışmalar için fikir sunabilir ve ilgili alanda daha fazla araştırmaya ihtiyaç olup olmadığını değerlendirmeye yardımcı olabilir. Bu analizler, çalışılan alandaki eğilimlerin anlamlı şekilde okuyucuya sunulmasını mümkün kılmaktadır (Mokhtari vd., 2019).

Bulut sistemleri veya farklı uygulamalar sayesinde uzak mesafelerdeki araştırmacıların iş birlikçi çalışabilme imkânları oldukça kolaylaşmıştır. Bu sayede küresel düzeyde yazar bağlantıları, uluslararası ortak yazarlı çalışma sayısı, araştırmaların niteliği de artmıştır (Arruda vd., 2022). Ortak yazarlı araştırmalar, bibliyometrik analiz yoluyla incelenerek yazarlar arasındaki bağlantılar, atıf ilişkileri ve ortak yazarlığa dair birçok parametre net bir şekilde ortaya konulabilir.

Günümüzde kurumlar, şirketler ve benzeri yapılar giderek artan bir şekilde bilgi ve işlemlerini dijital ortama taşımaktadır. Ticari faaliyetlerden eğitime, sağlıktan güvenliğe kadar birçok alanın işleyişi dijital platformlar üzerinden gerçekleştirilmektedir. Bulut bilişim sistemlerinin de dijital verilerdeki artışta önemli etkisi olmuştur (Çelik 2021; Çelik vd., 2018). Dolayısıyla günlük olarak internette petabayt düzeyinde veri transferi sağlanmakta, devasa boyutlarda veri akışı veri hacimlerindeki patlayıcı büyümeye neden olmaktadır. Bu olağanüstü miktardaki veri yükünden değerli bilgileri süzmek, bu bilgileri organize etmek için güçlü ve çok yönlü yazılımlara ihtiyaç duyulmuştur. Bu ihtiyaç veri madenciliği bilimini ortaya çıkarmıştır (Han vd., 2006).

Veri Madenciliği mevcut verilerden anlamlı örüntüler bulmak için büyük hacimli verilerin analiz edilmesidir. Veri Madenciliği, iş süreçlerinin daha verimli ve hızlı yürütülmesini sağlayan, hemen her iş kolunda kullanılabilen, gelişmekte olan bir teknolojidir (Diwani & Sam, 2014). Büyük hacimli ve organize veri depolarından çıkarım yapma, tahminde bulunma işi olan veri madenciliği, bilginin keşfi sürecinde bir adımdır ve verilerden örüntüler çıkarmak için uygulanan algoritmaları ifade eder. Bilginin keşfi sürecinde veri madenciliği ile bilgiye dönüşen veriler tahmin veya sınıflandırma modeli oluşturma, eğilimleri ve ilişkileri belirleme, mevcut bir modeli iyileştirme veya bilginin özetini sunma amaçları için kullanılmaktadır (Şimşek Gürsoy, 2009).

Kullanılacak veri madenciliği teknikleri genel amaca göre değişebilir. Veri madenciliği modelleri tahmin etmeye ve tanımlamaya yönelik modeller olmak üzere iki grupta ele alınmaktadır. Tahmin edici modeller eldeki veriler kullanılarak bir model elde etmeyi, bu model ile de başka verilerin sonuçlarını tahmin etmeyi öngörür. Tanımlamaya yönelik modeller ise eldeki veriler arasındaki ilişkileri analiz ederek kullanıcıya karar verme aşamasında rehberlik etmektedir (Özbay, 2015).

Büyük ölçekli veri tabanlarındaki verilerin anlamlı bilgilere dönüştürülmesi ve verilerin analiz edilerek yorumlanması insan yeteneklerini aşmaktadır. Her türlü verinin artık dijital ortamlarda tutulması veri tabanlarının boyutlarında ve sayılarında artışa neden olmaktadır. Bu ortamlardaki veriler ile yürütülecek işlemler zaman içerisinde veri madenciliğinin ortaya çıkmasını ve gelişmesini sağlamıştır (Savaş vd., 2012). Veri madenciliği bilimsel verilere dayanarak ileriye dönük tahminlerde bulunabilmektedir (Savaş vd., 2012). Örneğin bir eğitim kurumu mevcut verilerinden veri madenciliği ile elde edeceği tahminlerle geleceğe yönelik eğitim politikalarını şekillendirebilir. Bir ticarî kuruluş veri madenciliğini kullanarak ileriye dönük satış politikalarını belirleyebilir.

E-öğrenme platformları, öğrenme yönetim sistemleri gibi dijital öğrenme ortamları veri tabanlarında öğrenci hakkında önemli düzeyde veri kaydetmektedir. Bu veriler öğrenci aktiviteleri, akademik sonuçlar, etkileşimler hakkındaki veriler olabilir. Mevcut veriler, veri madenciliği teknikleri kullanılarak öğrencinin öğrenme süreçlerine ilişkin mevcut durumunu ortaya koyabilir, mevcut durum dikkate alınarak öğrenme süreçleri, eğitim programları öğrenci lehine şekillendirilebilir (Magdin & Turcâni, 2015; Mostow vd., 2005). Bu noktada ise öğrencinin öğrenme yöntemlerini tanımlayarak nasıl öğreneceği hakkında tahminler yürüten ve günümüz eğitim sisteminde gelişmekte olan bir bilim alanı olarak eğitimsel veri madenciliği ile karşılaşılmaktadır. Eğitimsel veri madenciliği, eğitime ilişkin verileri keşfederek anlamlandırmaya, öğrencinin nasıl öğrendiğini keşfetmeye çalışır. Eğitimsel veri madenciliği için önemli bir veri kaynağı olan e-öğrenme platformlarının bazıları eğitimde veri madenciliği tekniklerini kullanarak kendi sistemlerini geliştirmektedir (Boz & Şimşek, 2022; Romero vd., 2008).

Eğitimsel veri madenciliği; öğrenme çıktılarından elde edilen verileri analiz ederek öğrenmeyi anlamaya ve eğitim çıktılarını iyileştirmeye yardımcı olacak tahminlerde bulunan bir alandır (Feng, 2021). Veri madenciliğinin bir türü olan eğitimsel veri madenciliği, veri madenciliği tekniklerinin eğitim ortamlarından gelen verilere uygulanmasıdır. Romero ve Ventura (2013) eğitim ortamından gelen bazı verileri şu şekilde sıralamaktadır:

- Etkileşimde bulunan öğrencilerin yazışma metinleri,
- Öğrenci, öğretmen ve diğer personellerin idari verileri,
- Demografik veriler (cinsiyet, yaş, sınıf, adres, aile verileri vs.),
- Öğrencilerin duyuşsal özellikleri (ilgi alanları, yetenekleri, hobileri, fobileri vs.),
- Ölçme-değerlendirme verileri (notlar, puanlamalar, değerlendirmeye ilişkin metin verileri).

Romero & Ventura (2013) eğitimsel veri madenciliğinin bilgisayar bilimi, eğitim ve istatistik olmak üzere üç temel alan ile doğrudan ilişkili olduğunu savunmaktadır. Bu görüşe ilişkin görsel aşağıda sunulmuştur:



Şekil 1. Eğitimsel veri madenciliği ile ilişkili temel alanlar (Romero & Ventura, 2013)

Romero & Ventura (2013) Şekil 1’de gösterilen tüm alanlar arasında eğitimsel veri madenciliği ile en çok ilişkili olan alanın öğrenme analitiği olduğunu belirtir. Öğrenmeyi ve öğrenmenin gerçekleştiği ortamları anlama ve uygun şartlara getirme amacıyla öğrenciler ve bağlamlar hakkında verilerin ölçülmesi, toplanması, analizi ve

raporlanması işlemlerini yerine getiren öğrenme analitikleri, veri odaklı karar almaya ve öğrenme analitiklerinin teknik ve sosyal/pedagojik boyutlarını entegre etmeye odaklanır. Bu açılardan değerlendirildiğinde eğitimsel veri madenciliği ve öğrenme analitiği birçok özelliği paylaşabilmekte ve bazı benzer hedeflere ve ilgi alanlarına sahip olabilmektedir (Romero & Ventura, 2013).

Eğitimsel veri madenciliği “belli öğrenci kitlesine ait parametrelerden elde edilen bilgiler başka öğrenci popülasyonları için geçerli olabilir mi?” sorusunu akla getirmektedir. Bu bağlamda genel sonuçlar üzerinden hareket etmek ve sonuçları sık sık farklı veriler üzerinden test etmenin daha yararlı olacağı düşünülmektedir (Romero & Ventura, 2013). Geleneksel veri analizi hipoteze, literatüre veya sezgilere dayalı olarak yapıldığından bu analiz bütün öğrenenlerin benzer nitelikte olduklarını varsayar. Geleneksel analize alternatif olan veri madenciliği analizi ise daha anlamlı örüntü ve eğilimleri otomatik olarak keşfeder (Gaudio & Talavera, 2006). Eğitimde politika belirleyiciler, öğrenciler, öğretmenler, okul yöneticileri ve veliler veri madenciliğinin sağladığı avantajlardan yararlanabilirler. Özdemir vd. (2018) araştırmasında eğitim paydaşlarının eğitimde veri madenciliği kullanımını genel olarak benimsedikleri fakat eğitimde yeterince veri madenciliğinden yararlanılmadığını ifade etmiştir.

Eğitimsel veri madenciliği çalışmalarda farklı amaçlarla kullanılmıştır. E-öğrenme sistemlerinde öğrencilerin kullanım verileri görselleştirilerek benzer özellikteki öğrenciler gruplanmıştır. Bu doğrultuda, ilgili öğrenme sistemi için bir eklenti geliştirilmiş ve bu eklenti, isteyen öğretmenlerin öğrencileri tarafından kullanıma sunulmuştur (Romero vd., 2008). Sanal öğrenme ortamlarında ise öğrencilerin tartışma verileri, veri madenciliği ile analiz edilerek sonuçlar değerlendirilmiştir (Li vd., 2014). Lisansüstü eğitimde yapılan çalışmalarda, öğrencilerin özellikleri ve beklentileri ortaya çıkarılmış; bu veriler doğrultusunda ders materyalleri yeniden tasarlanmıştır (Yang vd., 2021).

Uzaktan ve çevrimiçi eğitim süreçlerinde, öğrencilerin sisteme kaydedilen verileri kullanılarak akademik başarıları ve öğrenme davranışları tahmin edilmiştir (Bütüner & Calp, 2022; Wang, 2021). Ayrıca, eğitimde belirli konularla ilgili mevcut durum ve problemlerin tespiti amacıyla eğitimsel veri madenciliğinden yararlanılmıştır (Deng, 2022). Benzer şekilde, öğrencilerin öğrenme sürecinde kütüphane kaynaklarını kullanırken karşılaştıkları sorunlar belirlenmiş ve bu kaynakların geliştirilmesi hedeflenmiştir (Yuan, 2022). Eğitimsel veri madenciliği çalışmalarından elde edilen sonuçların gelişmeleri takip edebilmek açısından politika yapıcılara, eğitimde gelecekteki eğilimleri şekillendirmek için ise araştırmacılara ve eğitimcilere de önemli katkılar sağlamaktadır (Boztaş vd., 2024).

Bu çalışmada eğitimsel veri madenciliği ile ilgili araştırmaların bibliyometrik analizi gerçekleştirilmiştir. Veri madenciliğinin eğitim alanında kullanımına ilişkin çalışmaların varlığına literatürde rastlanmaktadır. Az sayıda da olsa literatürde eğitimsel veri madenciliği ile ilgili bibliyometrik çalışmalara da rastlanmıştır. Boztaş

vd. (2024) 2010-2021 yılları arasındaki “eğitimsel veri madenciliği” anahtar kelimesini içeren WOS veri tabanında yer alan 628 makaleyi incelemişlerdir. Koklu & Sulak (2024) araştırmalarında “eğitim” ve “veri madenciliği” anahtar kelimelerini içeren 2021-2024 yılları arasındaki WOS veri tabanındaki SSCI, SCI-Expanded ve Web of Science Index indeksli 1170 makaleyi incelemiştir. Baek & Doleck (2022) WOS veri tabanında “eğitimsel veri madenciliği” anahtar kelimesini içeren 194 makaleyi incelemiştir. Wang vd. (2022) ise 2010-2022 yılları arasındaki WOS veri tabanındaki “eğitimsel büyük veri” ve “eğitimsel veri madenciliği” anahtar kelimelerini içeren makaleyi birlikte oluşturma ve birlikte atıf analizleri ile incelemişlerdir. Bu çalışmada ise bahsedilen çalışmalardan farklı olarak eğitimsel veri madenciliğinin yıllar içerisindeki durumunu daha net biçimde ortaya koyabilmek adına çalışmaların yayınlandığı yıl aralığı daha fazla tutulmuş (1980-2025), veri madenciliğinin farklı alanlardaki kullanımı da göz önünde bulundurularak eğitim alanına yansımalarını daha net tespit edebilmek için veri tabanında “veri madenciliği” anahtar kelimesi kullanılarak sadece eğitim ile ilgili kategorilerdeki çalışmalar seçilmiş ve sadece makaleler değil makalelere ek olarak kitap ve kitap bölümü türündeki 1165 belge incelenmiştir. Bunlara ek olarak çalışmalarla ilgili daha fazla analiz (üç alan analizi, trend konu analizi, tematik harita analizi, betimsel analizler gibi) gerçekleştirilerek daha detaylı veriye ulaşılmış ve veri analizini daha detaylı gerçekleştirebilmek adına VOSviewer yazılımına ek olarak R Bibliometrix yazılımı da kullanılmıştır. Bu yönüyle bu çalışmanın diğer çalışmalardan daha detaylı veriler sağlayarak daha detaylı sonuçlara ulaşacağı ve eğitimsel veri madenciliği konusunda daha geniş ve net bir görüş sağlayacağı düşünülmektedir. Böyle bir çalışma ile, eğitimsel veri madenciliği alanında mevcut durumun istatistiksel verilerle açıklanması, konu ile ilgili araştırmacılara rehberlik edilmesi amaçlanmıştır. Bu amaç doğrultusunda aşağıdaki sorulara cevaplar aranmıştır:

1980-2025 yılları arasında Web of Science veri tabanında eğitimsel veri madenciliği ile ilgili yayınlanan kitap, kitap bölümü ve makale türündeki yayınların;

- Yıllara göre sayısı nasıl değişmektedir?
- Trend olan konuları nelerdir?
- En fazla yayınlandığı dergiler hangileridir?
- Yazar, yayın yapılan dergi ve anahtar kelimeler ögeleri arasındaki üçlü alan ilişkisi nasıldır?
- Eğitimsel veri madenciliği ilgili çalışılan konuların eğitimsel veri madenciliği alanına etkisi (tematik haritası) nasıldır?
- En fazla iş birliği yapan ve bağlantılı olan yazarlar kimlerdir?
- Yazarlarının atıf ağları nasıldır?
- Ait olduğu ülkelere göre aldıkları atıfların durumu nedir?
- Ait olduğu kurumlara göre aldıkları atıfların durumu nedir?
- Anahtar kelimeleri arasındaki ilişki nasıldır?
- Bibliyografik eşleşmeleri nasıldır?
- Yazarlarının bibliyografik eşleşmeleri nasıldır?

m) Yazarlarının ortak atıf durumu nedir?

Yöntem

Bu araştırmada veri madenciliğinin eğitim alanında kullanımına ilişkin yapılmış çalışmalar bibliyometrik analiz yöntemi ile incelenmiştir. Bibliyometrik analiz, literatürdeki nicel verileri analiz ettiğinden sistematik analizden farklı özelliktedir. Bibliyometrik analizler ile söz konusu araştırma alanında mevcut eğilimin ortaya konulması sağlanmakta, literatür hakkında okuyuculara ve belli bir konu ile ilgili araştırma yapma düşüncesinde olanlara yol gösterilmektedir (Mokhtari vd., 2019).

Verilerin Toplanması

09.11.2024 tarihinde “data mining” anahtar kelimelerinin yan yana ve tırnak içinde yazılarak Web of Science veri tabanında tüm bilim alanlarında aranmasıyla 283.404 tane araştırma sonucuna ulaşılmıştır. Yıllara göre en eski araştırmanın 1980 yılında, en yenisi 2025 yılında olmak üzere makale, kitap, kitap bölümü, bildiri, tez araştırma sonuçlarına ulaşılmıştır. 2025 yılına ait araştırmaların yer alması erken görünümde yayın yapan dergilerden kaynaklanmaktadır. Elde edilen veriler anahtar kelime, özet, dergi, atıf, yazar, ülke, kurum parametreleri açısından incelenmiştir.

Verilerin elde edilme sürecinde Joshi (2016) tarafından önerilen aşağıdaki aşamalar takip edilmiştir.

Dahil edilme kriterleri:

Araştırmada veriler en sık kullanılan ve en nitelikli indeksleri tarayan veri tabanlarından olan Web of Science veritabanı (Joshi, 2016) üzerinden elde edilmiştir. Araştırmanın amacına uygun çalışmalara ulaşmak için sonuçlar eğitim bilimleri ile ilgili olan Education Educational Research (971), Education Scientific Disciplines (195) ve Education Special (15) kategorileri ile filtrelenmiştir. Toplamda 1181 çalışmaya ulaşılmıştır.

Dışlanma kriterleri:

Görece nitelikli yayın verileri elde edebilmek için bildiri, editöryal yayın, erken görünüm, makale incelemesi yayın türleri araştırmaya dahil edilmemiştir.

Arama stratejisi:

“data mining” anahtar kelimesi ile 09.11.2024 tarihinde Web of Science veri tabanında arama yapılmıştır. Literatür türü makale, kitap ve kitap bölümleri olarak belirlenmiş, bu kriterlerde 1165 çalışma listelenmiştir.

Çalışma seçimi:

Veri tabanlarındaki sistematik atıf güncellemeleri nedeniyle verilerin elde edilmesi aşamasında veri tabanlarında tarama işlemi mümkün olduğunca geç tarihte yapılmıştır. Verilerin analize hazır hale gelmesine kadar yapılan işlemler iki

araştırmacı tarafından bağımsız şekilde gerçekleştirilmiş ve görüş birliğine varılmıştır.

Arama sonuçları:

Arama sonucunda Web of Science veri tabanında 1165 araştırmaya ulaşılmıştır.

Veri soyutlama:

Araştırmaya katılan yayınlar; yayınlanma zamanı, kurum bilgileri, bölgesel bilgiler, yayımlayan dergi/yayınevi bilgileri, atıf bilgileri meta verileri analizlere katılmıştır.

Veri Analizi

Araştırma bibliyometrik analiz ve betimsel analiz kullanılarak yürütülmüştür. Bu analiz yöntemleri ile değişkenler arasındaki ilişkileri net bir şekilde ortaya koymak ve karşılaştırma yapmayı kolaylaştırmak, görsel sunumlar oluşturarak bulguların okunabilirliğini artırmak mümkündür. Bibliyometrik analizler için meta verileri ağ haritaları ve ağ katmanları ile görselleştiren, bilimsel haritalama yapan Vosviewer ve R Bibliometrix yazılımları kullanılmıştır. Bibliyometrik çalışmalarda haritalama için kullanılan yazılımlar bilimsel araştırma alanının kavramsal, entelektüel yapısını, gelişim sürecini ve dinamiklerini görsel çıktılarla sunmayı amaçlamaktadır (Gupta, 2021).

Bibliyometrik analiz amaçlı üretilen, ücretsiz temin edilebilen bir paket program olan VOSviewer; metin madenciliğine, haritalamaya ve görselleştirmeye imkân tanımaktadır. Dergilere, yazarlara, atıflara, anahtar sözcüklere dair bazı meta verileri belirli veri tabanlarından indirdikten sonra bu meta veriler üzerinden VOSviewer aracılığı ile analizler yapılabilir. VOSviewer, üzerinde çalışılan konuda daha önce yapılmış çalışmalar ile ilgili büyük resmin görülmesini sağlamaktadır. Araştırmacıların yaptığı çalışmanın sonuçları ise konu hakkında başka araştırmacılara önemli ipuçları vermektedir. Bibliyometrik analiz çalışmalarını inceleyen araştırmacılar, hangi yazarların çalışmalarını incelemenin avantaj olacağı, hangi kaynaklara atıfta bulunmanın öncelikli olabileceği, ilgili konudaki çalışma yoğunluğu vb. konularda ipuçları elde etmektedir. Bibliyometrik analiz bulguları araştırmacıya karar verme noktasında yardımcı olmaktadır (Arruda vd., 2022).

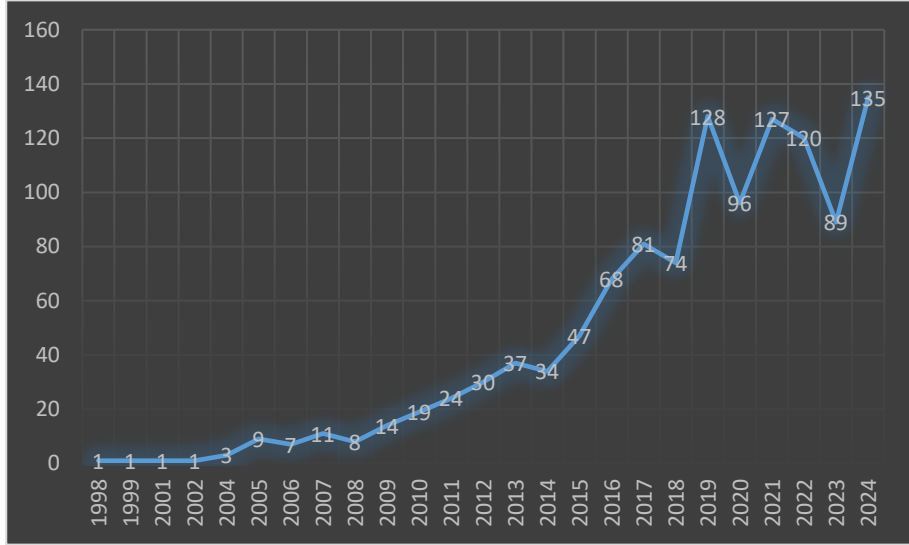
R Bibliometrix yazılımı ile trend konu analizi, üç alan analizi ve tematik harita analizi yapılmıştır. Vosviewer yazılımı ile ortak yazar analizi, yazarların atıf analizi, ülkelerin atıf analizi, kurumların atıf analizi, anahtar kelime analizi, metinlerin bibliyografik eşleşme analizi, yazarların bibliyografik eşleşme analizi ve yazarların ortak atıf analizi gerçekleştirilmiştir. Ayrıca yıllara göre yayın sayısı ve çalışmaların yayımlandığı dergiler ile ilgili betimsel istatistiklere de yer verilmiştir.

Bulgular ve Yorum

Bu bölümde yayınlar ile ilgili betimsel grafiklere yer verilmiştir. Bununla birlikte ortak yazar analizi, yazarların atıf analizi, ülkelerin atıf analizi, kurumların atıf analizi, anahtar kelime analizi, metinlerin bibliyografik eşleşme analizi, yazarların bibliyografik eşleşme analizi, yazarların ortak atıf analizi, trend konu analizi, üç alan analizi, tematik harita analizi ile yıllara göre yayın sayısı ve çalışmaların yayınlandığı dergiler ile ilgili bulgular grafik ve şekiller ile başlıklar halinde sunulmuş ve yorumlanmıştır.

Yıllara Göre Yayın Sayısı

Veri madenciliğinin eğitim alanında kullanımı ile ilgili araştırmaların yıllara göre değişimini gözlemek amacıyla yıllara göre yayın sayısı grafiği elde edilmiştir.

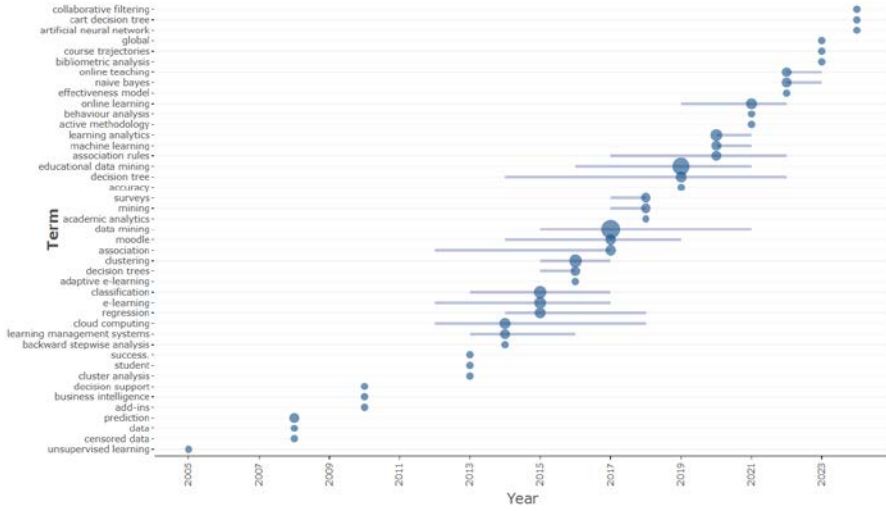


Şekil 1. Yıllara göre yayın sayıları

Web of Science veri tabanında “data mining” anahtar kelimesi ile eğitim bilimleri ile ilgili kategorilerde yer alan 1165 yayının yıllara göre sayısı Şekil 1’de görülmektedir. Buna göre eğitim bilimleri ile ilgili kategorilerde giderek artan bir trend görülmektedir. 2000’li yıllar itibariyle veri madenciliğinin eğitim bilimleri alanı ile ilgili yayınlarda daha görünür hale geldiği ve 2019 ve sonrasında ise yayın sayısının belirgin şekilde arttığı söylenebilir.

Trend Konu Analizi

Alandaki trendi, eğilimi belirlemek amacıyla R Bibliometrix yazılımı kullanılarak trend konu analizi yapılmıştır. Bu analiz yıllara göre anahtar kavramların yayınlarda nasıl yer aldığını, alanda hangi kavramların hangi yıllarda daha çok çalışıldığını göstermektedir.



Şekil 2. Trend konu analizi

Şekil 2'ye göre 2005-2011 yılları arasında çalışılan konular denetimsiz öğrenme, veri, tahmin, iş zekâsı, karar desteği gibi sınırlı sayıdaki konular iken özellikle 2013 sonrasında çalışılan konuların arttığı ve kavramların değiştiği görülmektedir. Özellikle 2015 yılı ve sonrasında veri madenciliğinin eğitim bilimleri ile ilgili alanlarında çalışılan konular dikkate alındığında e-öğrenme, öğrenme yönetim sistemleri, makine öğrenimi, davranış analizi, gibi farklılaşan konular çalışılmıştır. Burada dikkat çeken husus Şekil 1'de sunulduğu gibi 2019 ve devamında pandemi dönemine denk gelen süreç ve sonrasında (2020 ve sonrası) eğitimsel veri madenciliği kavramının trend olması ve daha çok çalışılmaya başlanmasıdır. Bununla birlikte çevrimiçi öğrenme, öğrenme analitiği ve kurs yörüngeleri kavramları da ön plana çıkmaktadır. 2024 yılının alandaki trend konuları ise yapay sinir ağları, cart karar ağacı ve iş birliğine dayalı filtreleme olduğu anlaşılmıştır

Çalışmaların Yayınlandığı Dergiler

Eğitim bilimleri kategorilerinde veri madenciliği ile ilgili en çok yayın yapan dergiler Şekil 3'te sunulmuştur.

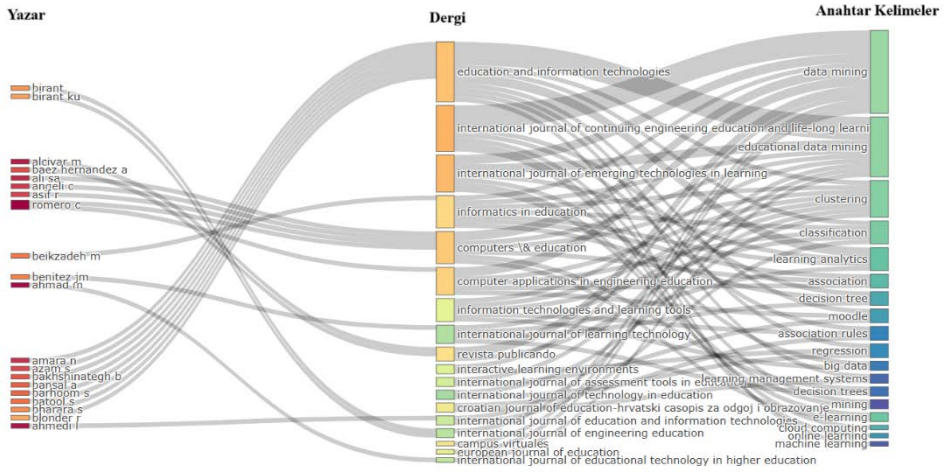


Şekil 3. Dergilere göre yayın sayıları

Buna göre eğitim bilimleri alanında veri madenciliği ile ilgili en fazla yayın yapan dergiler 88 yayın ile Education and Information Technologies, 71 yayın ile International Journal of Emerging Technologies in Learning, 59 yayın ile Journal of Surgical Education, 50 yayın ile IEEE Transactions on Learning Technologies ve 44 yayın ile Computers Education olarak sıralanmıştır.

Üç Alan Analizi

Üç alan analizi belirlenen üç alanın ana öğelerini ve aralarındaki ilişkiyi Sankey diyagramlarını kullanarak görselleştirmektedir. Sankey diyagramları farklı ağırlarda ve yöntemlerde akışı temsil eden bir görselleştirmedir (Riehmman vd., 2005). Yazar adı, atıf, kurum adı, anahtar kelime, dergi adı, ülke gibi metaveriler R Bibliometrix yazılımı ile görselleştirilebilmektedir. Bu çalışmada ise yazar, yayın yapılan dergi ve anahtar kelimeler belirlenerek analiz gerçekleştirilmiştir.

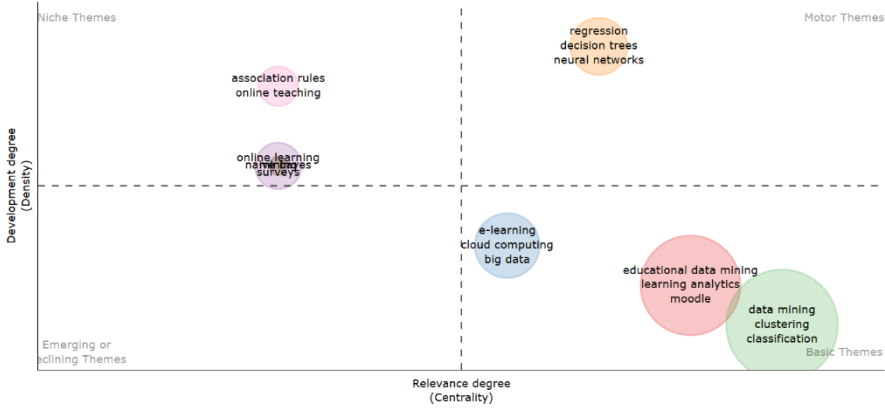


Şekil 4. Üç alan analizi

Şekil 4'te Web of Science veri tabanında eğitim ile ilgili kategorilerde veri madenciliği alanında belirlenen yayınlarda yazarlar, yayın yapılan dergiler ve kullandıkları anahtar kelimeler öğeleri arasındaki ilişki Sankey diyagramı ile gösterilmektedir. En çok kullanılan anahtar kelime olan veri madenciliği anahtar kelimesi ile en çok yayın yapılan dergi International Journal of Continuing Engineering Education and Life Long Learning olurken bu anahtar kelime ile en çok yayın yapan yazarlar olan D. Birant ve K. U. Birant'ın yayınlarının en çok yayımlandığı dergi ise International Journal of Engineering Education olmuştur.

Tematik Harita

Tematik harita, ilgili alanda çalışılan kavramların alana etkisini belirlemeye yarayan bir analiz yöntemidir. Çalışılan veri setinde anahtar kelimeler iki boyutlu sisteme aktararak sınıflama yapılabilir (Cobo vd., 2011). Eğitsel veri madenciliği konusuna ilişkin veri setinin tematik analizi Şekil 5'te sunulmuştur.

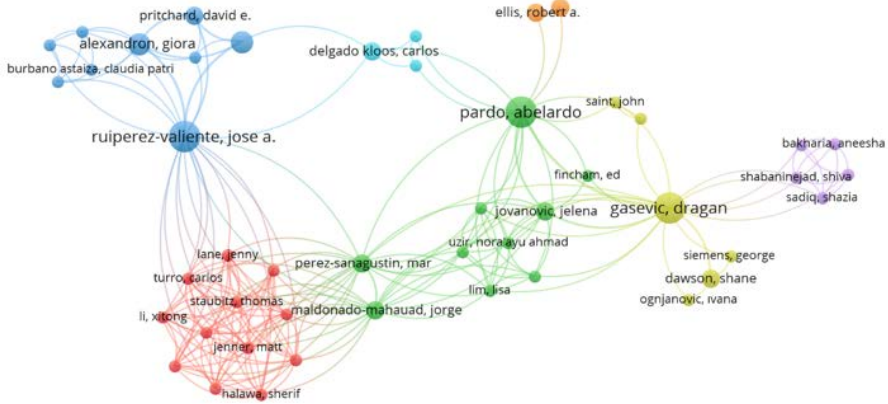


Şekil 5. Tematik harita

Elde edilen sonuçlar kartezyen bir görsele yansıtılmakta ve burada birinci kadran alandaki önemli olan ve zaman içinde geliştirilmiş kavramları yani öncül temaları (motor themes) göstermektedir. “Regresyon”, “karar ağları” ve “sinir ağları” temaları, eğitimle ilgili veri madenciliği alanında öncül temalar olarak ön plana çıkmaktadır. İkinci kadran kendi içlerinde gelişmiş ama spesifik olan, dışarıyla sınırlı ilgisi olan kavramları yani niş/yuva temaları (niche themes) göstermektedir. “Dernek kuralları”, “online öğretim”, “çevrimiçi öğrenme” ve “anketler” kavramları çalışılan alanda niş temalar olarak ön plana çıkmaktadır. Üçüncü kadran gelişmekte olan veya kaybolan temaları (emerging or disappearing themes) göstermektedir ve bu çalışmada belirgin şekilde bu kadrana ait kavram görünmemektedir. Dördüncü kadran ise alan için önemli olan fakat yeterince gelişmemiş kavramları yani temel temalar (basic themes) olarak isimlendirilen kavramları göstermektedir. “E – öğrenme”, “bulut bilişim”, “büyük veri”, “eğitimsel veri madenciliği”, “öğrenme analitikleri”, “Moodle”, “veri madenciliği”, “kümeleme” ve “sınıflandırma” kavramları ise eğitimle ilgili veri madenciliği alanında önemli olan fakat yeterince gelişmemiş kavramlardır.

Ortak Yazar Analizi

En fazla iş birliği yapan ve bağlantılı olan yazarları belirlemek amacı ile yapılan ortak yazar analizinde en az bir yayın ve en az bir atıf kriteri belirlenerek ağ haritası oluşturulmuştur.

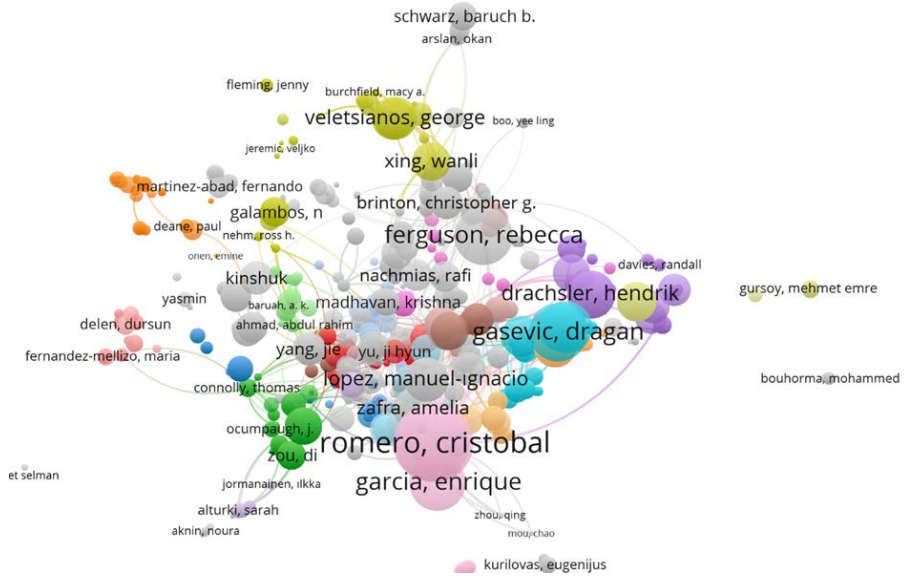


Şekil 6. Yazarlar arası iş birliğini gösteren ortak yazar ağ haritası

Aralarında bağlantı bulunan yazarlar arasında yapılan analize göre kendi aralarında ayrılmış 7 küme ve bunlara bağlı 47 öge ve 212 bağlantı tespit edilmiştir. Şekil 6'da görüldüğü gibi toplam bağlantı gücü en fazla olan yazarlar Xing (32), Ruiperez- Valiente (28), Hussain (27), Gasavic (24) ve Maldonada- Mahauad (23) olarak belirlenmiştir. Ayrıca yayın sayısına göre öne çıkan yazarlar incelendiğinde en fazla sayıda yayın yapan yazarlar şu şekilde sıralanmaktadır: Xing (12), Nuankaew (7), Doleck (7), Kimmons (7) ve Veletsianos (7). Burada en çok yayın yapan yazarın Xing (12) aldığı atıf sayısı 252 iken sayıca daha az yayın yapan Romero (7), Ventura (5) ve Gasevic (6) çok daha fazla atıf almışlardır. Romero 1317, Ventura 1247 Gasevic ise 669 atıf almıştır. Burada en çok yayın üreten ve atıf alan yazarlardan olan Xing ve Gasevic aynı zamanda en bağlantılı yazarlar arasında da yer almaktadır.

Yazarların Atıf Analizi

Yazarların atıf ağlarını belirlemek amacıyla en az 1 yayın ve en az 1 atıf kriteri ile yazar atıf analizi yapılmış ve bu analize ait ağ haritası çıkarılmıştır.

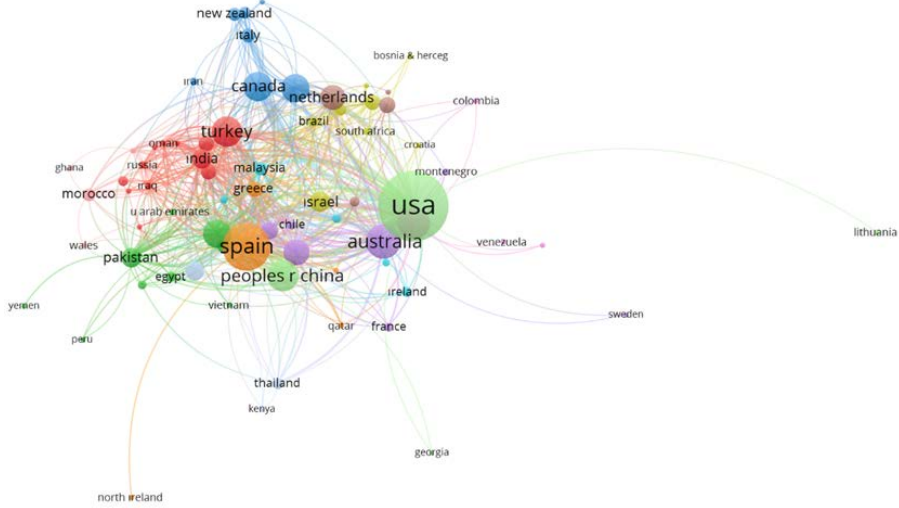


Şekil 7. Yazarların atıf bağlantıları ağ haritası

Birbiriyle bağlantılı olan 1296 birim ile yapılan analizde toplamda 39 küme, 8725 bağlantı ve toplam bağlantı gücü 9833 olarak tespit edilmiştir. Şekil 7’de görüldüğü üzere en fazla atıf alan yazarlar 1317 atıf ile Romero, 1247 atıf ile Ventura, 669 atıf ile Gasevic, 647 atıf ile Ferguson, ve 619 atıf ile Garcia olmuştur. Ayrıca Romero (439), Ventura (416) ve Gasavic (189)’in toplam bağlantı gücü bakımından da ilk 5 yazar arasında olduğu tespit edilmiştir.

Ülkelerin Atıf Analizi

Yayınların ait olduğu ülkelere göre aldıkları atıfların analizi amacıyla bir ülke tarafından en az 1 eser yayınlanması ve 1 atıf alınması kriteri dikkate alınarak ülkelerin atıf analizi ağ haritası oluşturulmuştur.

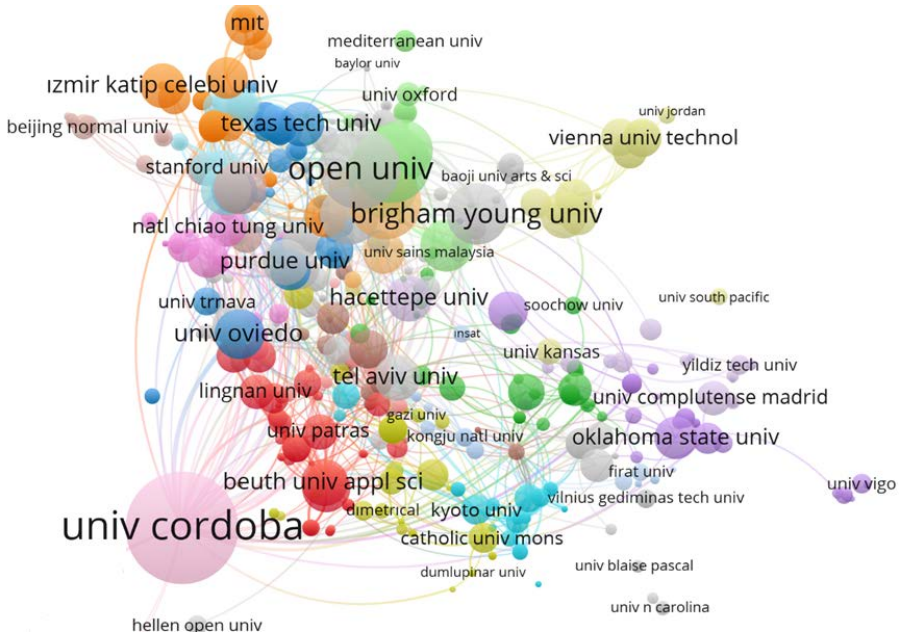


Şekil 8. Ülkelerin atıf bağlantıları ağ haritası

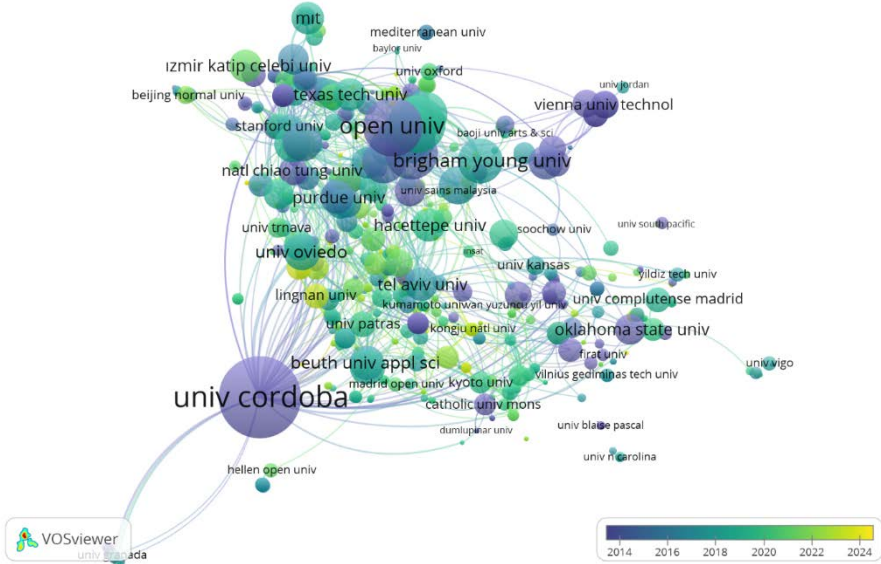
Bu kapsamda aralarında ilişki bulunan 74 gözlem birimi ile analiz yürütülmüş ve analiz sonucunda 12 küme, 578 bağlantı ve 1623 toplam bağlantı gücü tespit edilmiştir. En çok yayın sayısına sahip ülkeler Amerika Birleşik Devletleri (291), Çin (151), İspanya (97), Türkiye (63) ve Hindistan (59) olarak sıralanmaktadır. Yayınların en çok atıf alma durumuna göre ise ülkeler Şekil 8.'de de görüldüğü üzere ABD (5617), İspanya (2564), Avusturalya (1400), İngiltere (1224) ve Çin (1112) olarak sıralanmaktadır. Ayrıca bu beş ülkenin toplam bağlantı gücü en yüksek beş ülke olduğu da tespit edilmiştir; ABD (423), İspanya (266), Avusturalya (167), Çin (154), Türkiye (146).

Kurumların Atıf Analizi

Yayınların ait olduğu kurumlara göre aldıkları atıfların analizi amacıyla bir kurum tarafından en az bir eser yayınlanması ve bir atıf alınması kriteri dikkate alınarak kurumların atıf analizi ağ haritası oluşturulmuştur.



Şekil 9. Kurumların atıf bağlantıları ağ haritası

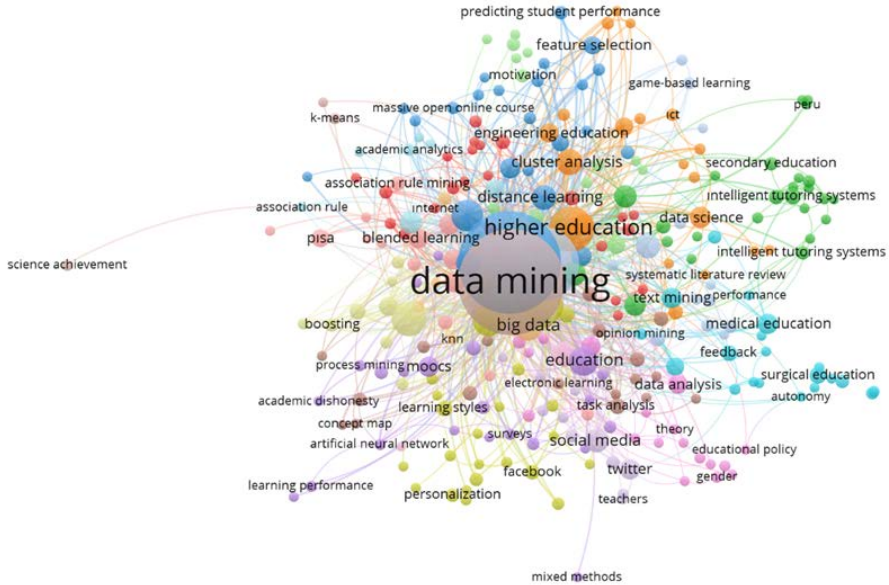


Şekil 10. Kurumların atıf bağlantıları katman haritası

Bu kapsamda aralarında ilişki bulunan 635 gözlem birimi ile analiz yürütülmüş ve analiz sonucunda 31 küme, 2782 bağlantı ve 3293 toplam bağlantı gücü tespit edilmiştir. En çok yayın yapan kurumlar Florida Üniversitesi (13), Hacettepe Üniversitesi (13), Monash Üniversitesi (11), Penn Üniversitesi (11), Michigan Üniversitesi (11)'dir. Kurumlar aldıkları atıf sayısına göre ise Şekil 9'da da görüldüğü gibi University Cordoba (1355), Open University (708), University Edinburg (651), Open University Netherlands (561) ve University South Australia (531) olarak sıralanmaktadır. Burada Open University'nin 3 yayın ile 708 atıf alarak atıf sırasında 2. sırada yer alması dikkat çekicidir. Bu durum ise Open University'nin yayınlarının daha eski olduğu, uzun süre yayında kalması sebebiyle zaman içinde daha çok atıf alması ile açıklanabilir. Nitekim Şekil 10'da da Open University'ye ait yayınların tarihinin daha eski olduğu, 2014-2016 yılları aralığına denk gelen mavi renklerden tespit edilebilmektedir.

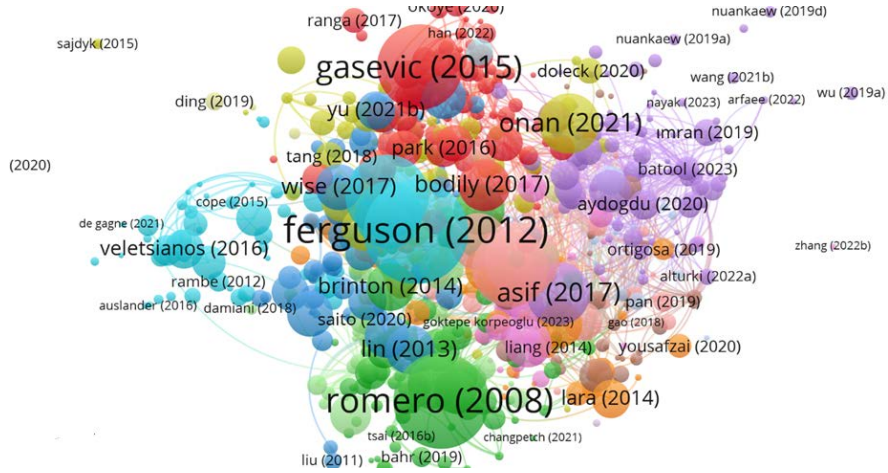
Anahtar Kelime Analizi

Anahtar kelime analizinde en az üç defa görülen ve aralarında ilişki bulunan 286 gözlem birimi ile analiz gerçekleştirilmiştir.



Şekil 11. En fazla kullanılan anahtar kelime bağlantıları ağ haritası

Analiz sonucunda toplam 18 küme, 2016 bağlantı ve 3321 toplam bağlantı gücü tespit edilmiştir. En fazla kullanılan anahtar kelimeler Şekil 11.'de görüldüğü üzere 321 tekrar ile data mining (veri madenciliği), 305 tekrar ile educational data mining (eğitimsel veri madenciliği), 162 tekrar ile learning analytics (öğrenme analitikleri),

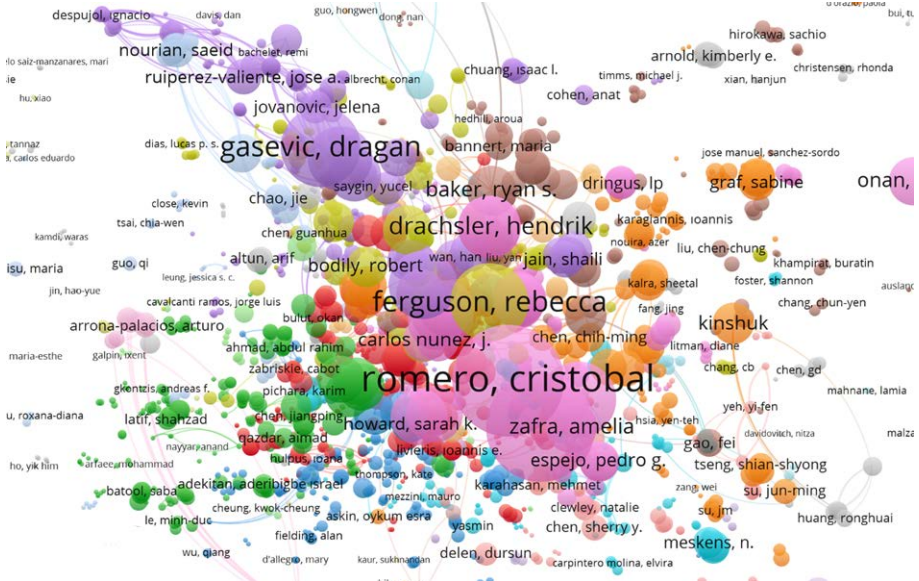


Şekil 13. Yayınların bibliyografik eşleşme bağlantıları ağ haritası

Analiz sonucunda 15 küme ve 31262 bağlantı elde edilmiştir. En fazla bibliyografik eşleşme olan yayınlar Şekil 13'te görüldüğü üzere 647 alıntı ile Ferguson (2012), 550 alıntı ile Romero (2008), 437 alıntı ile Gasevic (2015), 416 alıntı ile Greller (2015) ve 347 alıntı ile Romero (2013)'dür.

Yazarların Bibliyografik Eşleşme Analizi

En az bir eser yayınlamış ve bir atıf almış olmak kriteri ile seçilen ve aralarında bağlantı bulunan 2275 birim ile analiz gerçekleştirilmiştir.

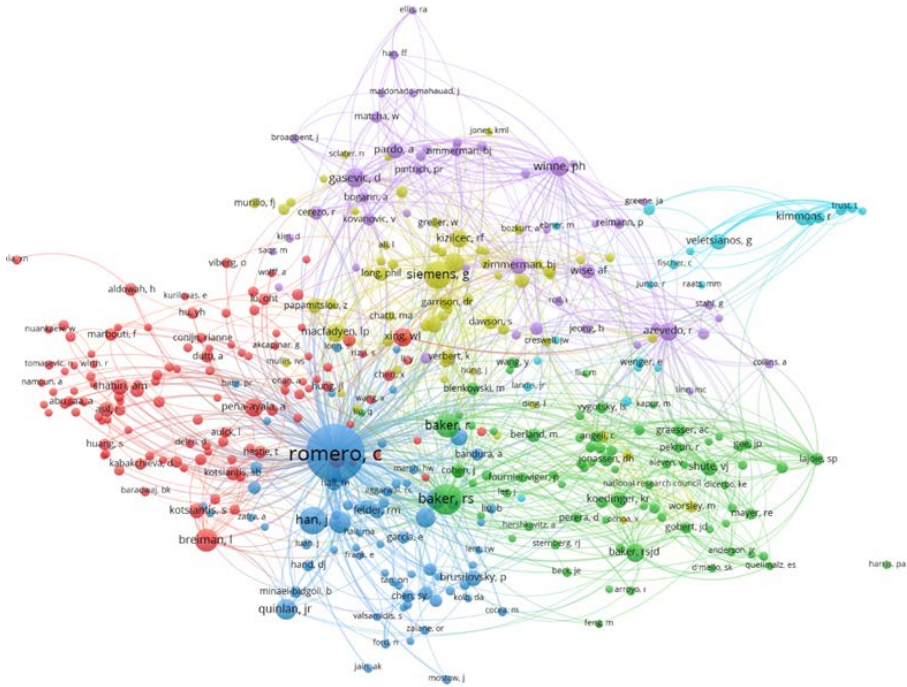


Şekil 14. Yazarların bibliyografik eşleşme bağlantıları ağ haritası

Analiz sonucuna göre 42 küme, 296269 bağlantı ve 707240 toplam bağlantı gücü elde edilmiştir. En fazla bibliyografik eşleşme olan yazarlar Şekil 14'te görüldüğü gibi 1317 alıntı ile Romero, Cristobal (6984 bağlantı gücü), 1247 alıntı ile Ventura, Sebastian (5325 bağlantı gücü), 669 alıntı ile Gasevic, Dragan (5514 bağlantı gücü), 647 alıntı ile Ferguson, Rebecca (1179 bağlantı gücü) 619 alıntı ile Garcia, Enrique (2123 bağlantı gücü) olduğu görülmektedir.

Yazarların Ortak Atıf Analizi

Bir yayında atıf yapılan farklı kaynaklar ortak atıf olarak isimlendirilmektedir. Atıf sayısı kriteri en az 10 belirlenerek 392 birim üzerinden

**Şekil 15.** Ortak atıf yapılan yazarlar arası bağlantıların ağ haritası

Analiz sonuçlarına göre toplamda 6 küme, 18063 bağlantı ve 50793 toplam bağlantı gücü tespit edilmiştir. En fazla ortak atıf yapılan yazarlar Şekil 15'te görüldüğü gibi Romero, C. 525 atıf (24223 bağlantı gücü), Baker, R.S. 143 atıf (7211 bağlantı gücü), Han, J. 117 atıf (4896 bağlantı gücü), Siemens, G. 116 atıf (5387 bağlantı gücü), Baker, R. 89 atıf (4303 bağlantı gücü) olarak tespit edilmiştir.

Tartışma ve Sonuç

Bu çalışmanın amacı Web of Science veri tabanındaki veri madenciliğinin eğitim alanında kullanımına ilişkin yapılmış çalışmaların bibliyometrik analiz yöntemi ile incelenmesidir. Araştırma bulguları konu ile ilgili araştırma yapacak veya bilgi sahibi olacak başka araştırmacılara rehberlik edilebilir. Gerçekleştirilen analiz sonuçlarına göre eğitim bilimleri alanında veri madenciliği ile ilgili en fazla yayın yapan dergi Education and Information Technologies olmuştur. En fazla sayıda yayın yapan yazar Xing iken en çok atıf alan yazar Romero'dur. En çok yayın sayısına ve atıfa sahip ülke Amerika Birleşik Devletleri'dir. En çok yayın yapan kurumlar Florida Üniversitesi iken en çok atıf alan kurum University Cordoba'dır.

Yıllara göre yayın sayıları ile ilgili gerçekleştirilen analiz sonuçlarına göre yıllara göre yayın sayıları incelendiğinde eğitim bilimleri ile ilgili kategorilerde giderek artan bir trend görülmektedir. Özellikle 2000'li yıllarla birlikte yüksek hızlı Grafik İşlem Birimleri (GPU) ve diğer donanımsal gelişmeler sayesinde veri madenciliği alanında da yaşanan gelişmelerin etkisiyle (Kızrak & Bolat, 2018; Schmidhuber, 2015) 2000'li yıllar itibarıyla veri madenciliğinin eğitim bilimleri alanı ile ilgili yayınlarda da daha görünür hale geldiği düşünülmektedir. Benzer şekilde Gülleroğlu ve Coşkun (2024) da bilgisayar teknolojilerinin eğitim sistemi üzerindeki etkisinin artışıyla birlikte eğitim alanında da büyük veri patlaması yaşandığını ve dolayısıyla veri madenciliğinin ön plana çıktığını belirtmektedir. Ayrıca 2019 yılı ve sonrasında yayın sayısının belirgin şekilde artışı, özellikle Covid-19 pandemisi döneminde gerçekleştirilen uzaktan eğitim sebebi ile farklı öğrenme yönetim sistemleri, farklı uzaktan eğitim sistemleri ve bu sistemlerdeki öğrenci başarısının araştırılması ile ilgili durumlardan dolayı eğitim bilimlerinde veri madenciliğinin yoğun şekilde çalışılması ile açıklanabilir. Tosunoğlu vd. (2021)'nin de belirttiği gibi bu dönemde uzaktan eğitim sistemlerinde kullanılan bilgi iletişim teknolojileri sebebiyle eğitim alanında kaydedilen verinin hacminde ve çeşitliliğinde müthiş bir artış yaşanarak büyük verinin oluşmasını sağlamıştır. Nitekim Yörük (2024) de özellikle Covid-19 salgını döneminde uzaktan eğitime geçiş ile birlikte küresel ölçekte yaygınlaşan çevrimiçi öğrenme ile veri madenciliğinin doğrudan ilişkili olduğunu belirtmektedir.

Alandaki eğilimi belirlemek için yapılan trend konu analizi sonuçlarına göre ise 2013 yılı sonrasında çalışılan konuların arttığı ve kavramların değiştiği, 2015 yılı ve sonrasında e-öğrenme, öğrenme yönetim sistemleri, makine öğrenimi, davranış analizi, gibi farklılaşan konuların çalışıldığı sonucuna ulaşılmıştır. 2019 yılı ve sonrasında yani Covid-19 pandemisi ve sonrasında eğitimsel veri madenciliği kavramının trend olduğu, bununla birlikte çevrimiçi öğrenme, öğrenme analitiği ve kurs yörüngeleri kavramlarının da ön plana çıktığı belirlenmiştir. Nitekim literatürde de bu süreçlerde çevrimiçi öğrenme ortamlarında kişiselleştirilmiş öğrenme ortamlarının önem kazandığı (Pai vd., 2021), zenginleştirilen öğrenme analitikleri (Hamal vd., 2022), eğitimsel veri madenciliği ve makine öğrenmesi kavramlarının da ön plana çıktığını desteklemektedir (Yörük, 2024). 2024 yılının alandaki trend konularının ise yapay sinir ağları, cart karar ağacı ve iş birliğine dayalı filtreleme olduğu sonucuna ulaşılmıştır. Yine de yapılan tematik alan analizi ile alan için önemli

olan “e – öğrenme”, “bulut bilişim”, “büyük veri”, “eğitimsel veri madenciliği”, “öğrenme analitikleri”, “kümeleme” ve “sınıflandırma” gibi kavramların yeterince gelişmemiş olduğu, üzerinde daha fazla çalışılabileceği sonucuna ulaşılmıştır. Bozkurt (2016) da eğitim alanında büyük verinin kullanılmasıyla birlikte bireyselleştirilmiş eğitim sistemleri ve öğrenme analitiği gibi alanların ortaya çıkışını, eğitim sistemine önemli katkılar sağladığını belirtirken Hotaman (2020) bu verilerin analizi için eğitsel veri madenciliği tekniklerinin kullanılması gerekliliğinden bahsetmekte, dolayısı ile bu alanların birbiri ile olan ilişkisi daha net ortaya konulmaktadır. Bu durum ise söz konusu eğilimler ve kavramlar noktasında bu çalışma için ulaşılan sonuçları desteklemektedir.

Alandaki eğilimi ortaya çıkaran göstergelerden biri sayılabilecek anahtar kelime analizi sonuçlarına göre ise en fazla kullanılan anahtar kelimeler ise veri madenciliği, eğitimsel veri madenciliği ve öğrenme analitikleri olarak belirlenmiştir. Nitekim 2020 yılı civarında ve sonrasında veri madenciliği anahtar kelimesinin eğitim, eğitimsel veri madenciliği ve koronavirüs, Covid-19 gibi anahtar kelimelerle birlikte kullandığı tespit edilmiş olup salgın hastalıklar, doğal afetler ve dolayısıyla bu zamanlarda özellikle eğitimin uzaktan gerçekleştirilmesi sonucu veri madenciliğinin bu alanlardaki incelemelerinin arttığı sonucuna ulaşılmıştır.

Öneriler

Veri madenciliğini eğitim alanında kullanımı ile ilgili çalışmalar yapanlar literatür ile ilgili incelemelere en fazla atıf alan Romero’nun çalışmalarından başlayabilirler. Ayrıca Romero, C. ile birlikte en fazla ortak atıf yapılan yazarlar olan Baker, R.S. ve Han, J.’nin de çalışmaları incelenebilir.

2024 yılının alandaki trend konularından olan yapay sinir ağları, cart karar ağacı ve iş birliğine dayalı filtreleme konuları ile alan için önemli olan ve yeterince gelişmemiş olduğu tespit edilen “e-öğrenme”, “bulut bilişim”, “büyük veri”, “eğitimsel veri madenciliği”, “öğrenme analitikleri”, “kümeleme” ve “sınıflandırma” gibi kavramlar üzerinde daha fazla çalışılabilir.

Sınırlılıklar

Bu araştırma Web of Science veri tabanında taranan, 1980 yılı ile 09.11.2024 tarihine kadar olan konu ile ilgili yapılmış makale, kitap, kitap bölümü, bildiri, tez çalışmaları ile sınırlıdır.

Çıkar Çatışması ve Etik Bildirimi: Bu çalışmada yazarlar arasında herhangi bir kişisel ve finansal çıkar çatışması bulunmamakla birlikte araştırma ve yayın etiği kurallarına uyulmuştur.

Kaynakça

- Arruda, H., Silva, E. R., Lessa, M., Proença Jr, D., & Bartholo, R. (2022). VOSviewer and bibliometrix. *Journal of the Medical Library Association: JMLA*, 110(3), 392. <https://doi.org/10.5195/jmla.2022.1434>
- Baek, C., & Doleck, T. (2022). Educational data mining: A bibliometric analysis of an emerging field. in *IEEE Access*, vol. 10, pp. 31289-31296, <https://doi.org/10.1109/ACCESS.2022.3160457>
- Boz, M., & Şimşek, İ. (2022). Milli Eğitim Bakanlığı eğitim yönetim bilgi sistemlerinin birlikte çalışabilirlik açısından incelenmesi. *Journal of Qualitative Research in Education*, 32, 213-240. <https://doi.org/10.14689/enad.32.1702>
- Bozkurt, A. (2016). Öğrenme analitiği: e-öğrenme, büyük veri ve bireyselleştirilmiş öğrenme. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 2(4), 55-81. <https://dergipark.org.tr/en/download/article-file/402115>
- Boztaş, G. D., Berigel, M., & Altınay, F. (2024). A bibliometric analysis of Educational Data Mining studies in global perspective. *Education and Information Technologies*, 29, 8961–8985. <https://doi.org/10.1007/s10639-023-12170-0>
- Broadus, R. N. (1987). Toward a definition of “bibliometrics”. *Scientometrics*, 12, 373-379.
- Bütün, R., & Calp, M. H. (2022). Estimation of the academic performance of students in distance education using data mining methods. *International Journal of Assessment Tools in Education*, 9(2), 410-429. <https://doi.org/10.21449/ijate.904456>
- Cobo, M.J., López Herrera, A.G., Herrera-Viedma, E. & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146-166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Çelik, K. (2021). Bulut bilişim teknolojileri. *Bartın Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 12(24), 436-450. <https://doi.org/10.47129/bartiniibf.1019898>
- Çelik, K., Güler, S., & Özköse, H. (2018). 4. Endüstri devrimine kuramsal bakış. *Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi (ASEAD)*, 5(9), 86-95.
- Deng, L. (2022). Developing countermeasures of integrating entrepreneurship education with professional education in colleges and universities using data mining. *Journal of Electrical and Computer Engineering*, 2023(1), 1-10. <https://doi.org/10.1155/2022/1723712>
- Dereli, A. B. (2024). Vosviewer ile Bibliyometrik Analiz. *Communicata*, 28, 1-7. <https://doi.org/10.32952/communicata.1517725>
- Diwani, S. A., & Sam, A. (2014). Data mining awareness and readiness in healthcare sector: A case of Tanzania. *Advances in Computer Science: an International Journal*, 3(1), 37-43.
- Feng, L. (2021). Research on higher education evaluation and decision-making based on data mining. *Scientific Programming*, 2021(1), 1-9. <https://doi.org/10.1155/2021/6195067>
- Gaudio, E., & Talavera, L. (2006). *Data mining to support tutoring in virtual learning communities: Experiences and challenges*. Data mining in e-learning/Wit Press.
- Gupta, N. (2021). Trends in IoT research: A bibliometric and science mapping analysis of internet of things. *Library Philosophy and Practice*, 1-20. <https://digitalcommons.unl.edu/libphilprac/5269>
- Gülleroğlu, H. D., & Coşkun, F. (2024). Büyük verinin yapısı ve eğitimde kullanılması. *Millî Eğitim*, 53(242), 1139-1152. <https://doi.org/10.37669/milliegitim.1222977>
- Hall, C. M. (2011). Publish and perish? Bibliometric analysis, journal ranking and the assessment of research quality in tourism. *Tourism management*, 32(1), 16-27. <https://doi.org/10.1016/j.tourman.2010.07.001>
- Hamal, O., El Faddouli, N. E., Harouni, M. H. A., & Lu, J. (2022). Artificial intelligent in education. *Sustainability*, 14(5), 1-11. <https://doi.org/10.3390/su14052862>

- Han, J., Kamber, M., & Pei, J. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Hotaman, D. (2020). Öğrenci başarısının değerlendirilmesinde eğitsel veri madenciliğinin kullanımı. *Ulakbilge Sosyal Bilimler Dergisi*, 8(48), 577-587. <https://doi.org/10.7816/ulakbilge-08-48-08>
- Joshi, A. (2016). Comparison between Scopus & ISI Web of science. *Journal Global Values*, 7(1), 1-11.
- Kızrak, M. A., & Bolat, B. (2018). Derin öğrenme ile kalabalık analizi üzerine detaylı bir araştırma. *Bilişim Teknolojileri Dergisi*, 11(3), 263-286. <https://doi.org/10.17671/gazibtd.419205>
- Köklü, N., & Sulak, S. A. (2024). Recent developments in educational data mining: A four-year bibliometric analysis. *Advances in Education Sciences*, M. Dalkılıç and O. Soslu (Ed.) *Platanus Publishing*, 5-29.
- Li, X., Nie, Y., Chen, M., Liu, X., & Liu, X. (2014). Members' behavior in virtual learning community: A study using data mining approach. *Computer and Information Science*, 7(4), 1-8. <https://doi.org/10.5539/cis.v7n4p1>
- Magdin, M., & Turcáni, M. (2015). Personalization of student in course management systems on the basis using method of data mining. *Turkish Online Journal of Educational Technology-TOJET*, 14(1), 58-67. <https://doi.org/10.1186/s41239-024-00486-x>
- Mokhtari, H., Mirezati, S. Z., Saberi, M. K., Fazli, F., & Kharabati-Neshin, M. (2019). A bibliometric analysis and visualization of the scientific publications of universities: A study of hamadan university of medical sciences during 1992-2018. *Webology*, 16(2), 187-211. <https://doi.org/10.31219/osf.io/92u3m>
- Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., & Heiner, C. (2005). An educational data mining tool to browse tutor-student interactions: Time will tell. *Proceedings of the Workshop on Educational Data Mining, National Conference on Artificial Intelligence, USA*. <https://www.cs.cmu.edu/~listen/pdfs/AAAI2005-WS205MostowJ.pdf>
- Özbay, Ö. (2015). Veri madenciliği kavramı ve eğitimde veri Madenciliği uygulamaları. *Uluslararası Eğitim Bilimleri Dergisi*, 2(5), 262-272. <https://doi.org/10.16991/INESJOURNAL.162>
- Özdemir, A., Saylam, R., & Bilen, B. B. (2018). Eğitim sisteminde veri madenciliği uygulamaları ve farkındalık üzerine bir durum çalışması. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 22(2), 2159-2172.
- Pai, K. C., Kuo, B. C., Liao, C. H., & Liu, Y. M. (2021). An application of Chinese dialogue-based intelligent tutoring system in remedial instruction for mathematics learning. *Educational Psychology*, 41(2), 137-152. <https://www.tandfonline.com/doi/abs/10.1080/01443410.2020.1731427>
- Riehmann, P., Hanfler, M., & Froehlich, B. (2005, 23-25 October). *Interactive sankey diagrams*. IEEE Symposium on Information Visualization - INFOVIS. (pp. 233-240). Minneapolis, USA. <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1532152>
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, 3(1), 12-27. <https://doi.org/10.1002/widm.1075>
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384. <https://doi.org/10.1016/j.compedu.2007.05.016>
- Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23. <https://dergipark.org.tr/tr/download/article-file/199592>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>

- Şimşek Gürsoy, U. T. (2009). *Veri madenciliği ve bilgi keşfi*. Pegem Akademi
- Tosunoğlu, E., Yılmaz, R., Özeren, E., & Sağlam, Z. (2021). Eğitimde makine öğrenmesi: araştırmalardaki güncel eğilimler üzerine inceleme. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 3(2), 178-199. <https://dergipark.org.tr/en/download/article-file/1873550>
- Wang, C. (2021). Analysis of students' behavior in english online education based on data mining. *Mobile Information Systems*, 2021(11), 1-10. <https://doi.org/10.1155/2021/1856690>
- Wang, C., Dai, J., & Xu, L. (2022). Big data and data mining in education: A Bibliometrics study from 2010 to 2022. *2022 7th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*. Chengdu, China, pp. 507-512. <https://doi.org/10.1109/ICCCBDA55098.2022.9778874>
- Yang, L., Zheng, S., Xu, X., Sun, Y., Wang, X., & Li, J. (2021). Medical data mining course development in postgraduate medical education: web-based survey and case study. *JMIR Medical Education*, 7(4), 1-18. <https://doi.org/10.2196/24027>
- Yörük, T. (2024). *Eğitimde yapay zeka ve kişiselleştirilmiş öğrenme*. S. Karataş (Ed.), Eğitim Bilimleri Alanında Uluslararası Araştırmalar XXIII, Eğitim Yayınevi.
- Yuan, X. (2022). Network education resource information sharing system based on data mining. *Mathematical Problems in Engineering*, 2022(1), 1-8. <https://doi.org/10.1155/2022/4080049>

Extended Abstract

Bibliometric analysis is an analysis technique that examines the trends in any subject, the current situation in the literature, and the relationships between the researches. With this technique, we can perform operations through a visual interface without coding knowledge (Dereli, 2024). With bibliometric analysis, publications are analyzed quantitatively (Broadus, 1987). Although it is a quantitative method by nature, this analysis, which includes the examination of qualitative features, can be applied to all kinds of publications, but the focus of the analysis is journals (Hall, 2011).

VOSviewer, a free package program produced for bibliometric analysis purposes, enables text mining, mapping and visualization. After downloading some metadata about journals, authors, citations and keywords from certain databases, analyses can be performed on this metadata via VOSviewer. VOSviewer allows the big picture to be seen regarding previous studies on the subject being studied. The results of the study conducted by the researchers provide important clues to other researchers on the subject. Researchers who examine bibliometric analysis studies obtain clues about which authors' studies would be advantageous to examine, which sources would be a priority to cite, the density of studies on the relevant subject, etc. Bibliometric analysis findings help the researcher in making decisions (Arruda et al., 2022).

Data Mining is the analysis of large volumes of data to find meaningful patterns from existing data. Data Mining is an emerging technology that enables business processes to be carried out more efficiently and quickly and can be used in almost every line of business (Diwani & Sam, 2014). Data mining, which is the work of

making inferences and predictions from large volumes and organized data stores, is a step in the process of knowledge discovery and refers to algorithms applied to extract patterns from data. In the process of knowledge discovery, the data transformed into knowledge by data mining are used to create a prediction or classification model, identify trends and relationships, improve an existing model, or present a summary of knowledge (Fayyad et al., 1996).

E-learning environments are frequently used in extraordinary situations where face-to-face education cannot be provided such as Covis-19, natural disasters or for different purposes to benefit from distance education. However, the effectiveness of e-learning environments has been a subject of ongoing debate and has been constantly compared with face-to-face education. It has been understood that the use of data mining methods in the evaluations of e-learning environments has significant advantages and that Educational Data Mining studies have contributed to education, although they have become widespread in recent years. In addition, bibliometric analyses reveal the current trend in the researches conducted and guide the readers about the literature and those who are thinking of doing research on a certain subject.

The existence of studies on the use of data mining in the field of education is found in the literature, but it has been observed that there is no bibliometric study on these studies. With such a study, it is aimed to explain the current situation on this subject with statistical data and to guide researchers on the subject.

This research is limited to articles, books, book chapters, proceedings, theses and dissertations scanned in the Web of Science database between 1980 and 09.11.2024. In this study, studies on the use of data mining in the field of education were examined by bibliometric analysis method. Bibliometric analysis is different from systematic analysis as it analyzes quantitative data in the literature.

On 09.11.2024, 283404 research results were obtained by searching the Web of Science database in all fields of science by typing the keywords “data mining” side by side and in quotes. According to the years, articles, books, book chapters, proceedings, thesis research results were reached, with the oldest research in 1980 and the newest in 2025. The inclusion of 2025 researches is due to the journals publishing in early view. The data obtained were analyzed in terms of keyword, abstract, journal, citation, author, country and institution parameters. In the process of obtaining the data, 1165 studies were reached in line with the application of the criteria recommended by Joshi (2016).

The research was conducted using bibliometric analysis and descriptive analysis. For bibliometric analyses, Vosviewer and R Bibliometrix software, which visualize metadata with network maps and network layers and perform scientific mapping, were used. R Bibliometrix software was used for trending topic analysis, three-domain analysis and thematic map analysis. With Vosviewer software, co-author analysis, citation analysis of authors, citation analysis of countries, citation analysis of

institutions, keyword analysis, bibliographic match analysis of texts, bibliographic match analysis of authors and co-citation analysis of authors were performed.

According to the findings, there is an increasing trend in the categories related to educational sciences. It can be said that data mining has become more visible in publications related to the field of educational sciences since the 2000s, and the number of publications has increased significantly in 2019 and beyond. While the topics studied between 2005-2011 were limited to unsupervised learning, data, prediction, business intelligence, decision support, it is seen that the topics studied have increased and the concepts have changed especially after 2013. Considering the topics studied in the fields of data mining related to educational sciences, especially in 2015 and after, differentiated topics such as e-learning, learning management systems, machine learning, behavior analysis, etc. have been studied. The journal with the highest number of publications on data mining in the field of educational sciences is Education and Information Technologies with 88 publications and Computers Education with the least number of publications is Computers Education with 44 publications.