Pamukkale Univ Muh Bilim Derg, 24(2), 336-343, 2018

Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi

Pamukkale University Journal of Engineering Sciences

Bulut bilişimde veri madenciliği tekniklerinin uygulanması: Bir literatür taraması

Application of data mining techniques in cloud computing: A literature review

Pelin YILDIRIM[1], *Derya BİRANT*[2*]

[1]Fen Bilimleri Enstitüsü, Dokuz Eylül Üniversitesi, İzmir, Türkiye.
pelin@cs.deu.edu.tr
[2]Bilgisayar Mühendisliği Bölümü, Mühendislik Fakültesi, Dokuz Eylül Üniversitesi, İzmir, Türkiye.
derya@cs.deu.edu.tr

Geliş Tarihi/Received: 12.11.2016, Kabul Tarihi/Accepted: 15.03.2017
doi: 10.5505/pajes.2017.65642
* Yazışılan yazar/Corresponding author
Derleme Makalesi/Review Article

Öz

Son yıllarda, bilgi ve iletişim teknolojilerindeki yenilikler ve gelişmeler, analiz edilmesi gereken veri miktarını önemli derecede artırmıştır. Büyük ölçekli verilerin saklanması, yüksek hacimli verilerden veri madenciliği teknikleri ile yararlı bilgilerin çıkartılması ve geleceğin tahminlenmesi maliyetli ve zorlu işlemlerdir. Bu zorlukların üstesinden gelebilmek için, bilgi keşfi süreci; bulut bilişim, paralel ve dağıtık hesaplama kullanılarak etkin bir şekilde gerçekleştirilebilmektedir. Bu makale, bulut bilişimin ölçeklenebilirliği sayesinde veri madenciliği algoritmalarının performanslarının arttırılabileceğini, ayrıca her yerden ulaşılabilirlik, düşük maliyet ve kolay yönetilebilirlik avantajlarının sağlanabileceğini göstermektedir. Makalede, bulut platformunda gerçekleştirilen veri madenciliği uygulamaları; kullanılan metotlar, veriler ve elde edilen sonuçlar çerçevesinde sunulmaktadır. Literatürde, bu konuda önerilen çözüm yaklaşımları; sınıflandırma, kümeleme ve birliktelik kuralları analizi olmak üzere üç ana kategori altında ele alınmaktadır.

Anahtar kelimeler: Veri madenciliği, Sınıflandırma, Kümeleme, Birliktelik kuralları analizi, Bulut bilişim

Abstract

In recent years, new innovations and developments in information and communication technologies have hugely increased the quantity of data required to analyze. Storing large-scale datasets, extracting useful knowledge from the huge volumes of data by applying data mining techniques and predicting the future are costly and difficult processes. To overcome these challenges, the knowledge discovery process is performed efficiently by using cloud, parallel and distributed computing. This article shows that the performance algorithms in data mining can be increased by the scalability of cloud computing with the advantages in terms of accessibility from anywhere, low cost and maintainability. In this article, data mining applications that have been implemented on cloud platforms are presented, including methods, data and obtained results. Solution approaches that have been proposed related to this topic in the literature are handled in three main categories: classification, clustering and association rule mining.

Keywords: Data mining, Classification, Clustering, Association rule mining, Cloud computing

1 Giriş

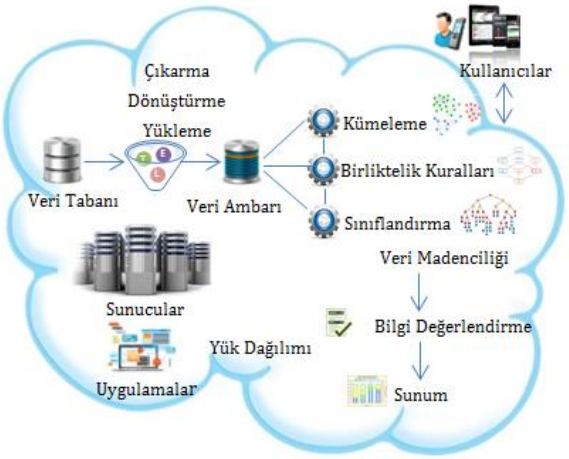
Son senelerde bilgi teknolojileri alanında yaşanan hızlı gelişmeler, veri miktarındaki artışı da beraberinde getirmiştir. Her gün eğitim, satış ve pazarlama, sağlık, telekomünikasyon, finans ve sosyal ağ gibi birçok alanda büyük miktarda veri üretilmektedir. Günümüzde, bu büyük miktardaki ham verinin işlenerek yararlı bilgiye dönüştürülmesi için veri madenciliği tekniklerinden faydalanılmaktadır.

Veri miktarındaki önlenemez artış, veri madenciliği tekniklerinin büyük hacimli veriler üzerinde etkin ve yüksek performanslı bir şekilde uygulanabilirliğini sağlamak amacıyla *bulut bilişim* [1],[2], *dağıtık hesaplama* [3], *paralel işleme* [4] gibi stratejilerin kullanılmasını zorunlu hale getirmiştir. Bulut bilişim kavramı, genel olarak, bir bilginin veya kaynağın internet yardımı ile her yerden erişilebilir bir şekilde paylaşılması olarak özetlenebilir. Söz konusu kaynaklar; uygulamalar, veri tabanları, hizmetler, sunucular, bilgisayar ağları ve benzeri yönetimsel çaba gerektiren yapılardır. Bunun yanı sıra sağlamış olduğu depolama imkânı, ölçeklenebilirlik ve paralel işleme özelliği ile yüksek hacimli veriler üzerinde analiz gerçekleştirme olanağı sunmaktadır. Bu sebepledir ki, büyük miktardaki veriden bilgi keşfedilmesi sürecinde bulut bilişim sistemlerinden yararlanmak mümkündür. Şekil 1'de veri madenciliği yöntemlerinin bulut platformunda kullanımına ilişkin mimari yapı gösterilmektedir.

Veri madenciliğinin bulut ortamında gerçekleştirilmesinin sağlayacağı başlıca avantajlar şunlardır [5],[6]:

- Ölçeklenebilirlik, esnek ve paralel işlem gücü sayesinde veri madenciliği algoritmalarının büyük verilerde daha hızlı sonuç üretebilmesi,
- Veri madenciliği uygulamasının her yerden ulaşılabilir olması,
- Büyük veri saklama ve analiz maliyetinin düşmesi,
- Kurulum, bakım gibi gereksinimlerin olmaması sayesinde kolay yönetilebilirlik,
- Müşterinin sadece kullandığı veri madenciliği hizmeti kadar ödeme yapabilmesi.

Bu makalede, yukarıda sayılan tüm avantajlardan yararlanabilmek için bulut platformunda gerçekleştirilen veri madenciliği uygulamaları sunulmaktadır. Veri madenciliğinin bulut platformunda; eğitim[7], sağlık [8], endüstri [9], ulaşım [10], sosyal ağ [11] gibi çok farklı alanlarda kullanıldığı görülmektedir. Makalenin bundan sonraki bölümlerinde, bulut bilişimde sınıflandırma [12],[13], kümeleme [14] ve birliktelik kuralları analizini [15] gerçekleştiren çalışmalar üç ayrı başlık altında ele alınmaktadır.



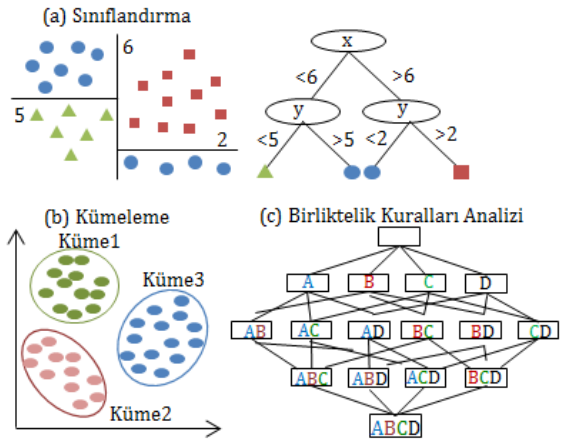
Şekil 1: Bulut platformunda veri madenciliği yöntemlerinin kullanılması.

2 Temel bilgiler

2.1 Veri madenciliği ve yöntemleri

Verilerin toplanarak önişlemden geçirilmesi, anlamlı bilgilerin çıkartılması ve sonuçların doğrulanarak sunulması sürecine "bilgi keşfi" denilmektedir. Veri madenciliği, bu süreç içerisinde yer alan önemli bir adımdır.

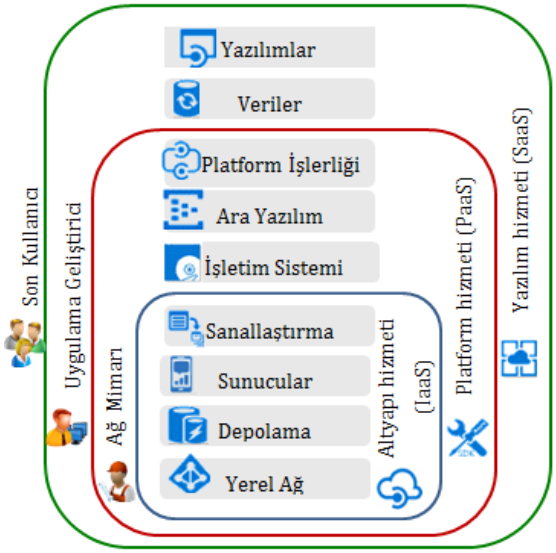
Veri madenciliği, veri içerisinden gizli, anlamlı ve potansiyel olarak değerli örüntülerin ortaya çıkartılmasını sağlayan bir dizi yöntemler bütünüdür. Veri madenciliği yöntemleri; sınıflandırma, kümeleme ve birliktelik kuralları analizi olmak üzere üç ana başlık altında toplanmaktadır. *Sınıflandırma*, sonuçları bilinen veriler ile bir sınıflandırma modeli oluşturulması ve bu model kullanılarak yeni gelen örneklem verinin daha önceden belirlenmiş sınıflardan birine dahil edilmesi işlemidir. *Kümeleme*, bir veri seti içerisindeki birbirine benzer nesneleri bir araya getirerek gruplayan bir veri madenciliği tekniğidir. *Birliktelik kuralları analizi* ise, verilerde yer alan öğeler arasındaki ilginç bağıntıların ve ilişkilerin kurallar halinde bulunması işlemidir. Veri madenciliği yöntemlerine ilişkin örnekler Şekil 2'de gösterilmektedir.



Şekil 2: Veri madenciliği yöntemleri.

2.2 Bulut bilişim

Bulut bilişim teknolojisi kullanıcıya altyapı, platform ve yazılım hizmeti olmak üzere üç farklı servis modeli sağlamaktadır (Şekil 3). *Altyapı hizmeti* (IaaS), müşteriye fiziksel ya da sanal olarak sunucu, depolama alanı ve ağ bağlantısı sağlanmaktadır. Bu servis modelinden yararlanan müşteri, uygulamasını çalıştırmak için gerekli olan işletim sistemi, ara yazılım gibi gereçlerin kurulumu ve bakımından kendisi sorumludur. *Platform hizmeti* (PaaS) modelinde ise, müşteriye gerekli olan tüm donanımsal, yönetimsel araçlar ve bunların bakımları bulut firması tarafından sağlanmaktadır. Kullanıcının tek görevi veritabanını oluşturup ve uygulamasını geliştirip, bulut üzerinde barındırılmasını sağlamaktır. *Yazılım hizmeti* (SaaS) modelinde ise her türlü donanım ve yazılım firmalar tarafından gerçekleştirilmiş olup, kullanıcı hazır olarak sunulan bir yazılıma tarayıcı yardımı ile ulaşıp kullanmaktadır.



Şekil 3: Bulut bilişim servis modelleri ve kullanıcı türleri.

3 Araştırma metodu

Bu makale kapsamında gerçekleştirilen "Sistematik Literatür Taraması", bulut bilişim ve veri madenciliği alanındaki anahtar kelimeler birlikte kullanılarak ve üç grup altında toplanarak gerçekleştirilmiştir:

(i) Bulut Bilişim ve Sınıflandırma,

(ii) Bulut Bilişim ve Kümeleme,

(iii) Bulut Bilişim ve Birliktelik Kuralları Analizi.

Bu anahtar kelimeler, akademik alanda sıklıkla kullanılan elektronik bilgi kaynaklarında (örneğin; Springer, IEEE, Elsevier, ACM) İngilizce dilinde aratılmıştır. Ayrıca, Scopus, Web of Science ve Google Scholar gibi popüler dijital veri tabanlarından da çalışmalar elde edilmiştir. Literatür tarama ve inceleme çalışmaları yaklaşık iki aylık bir süre içerisinde gerçekleştirilmiştir. İncelenen çalışmaların çok büyük bir kısmı hakemli dergilerde yayınlanan makalelerden oluşmaktadır. Bulut bilişim kavramı yeni bir teknoloji olduğundan dolayı incelenen tüm yayınlar 2010 ile 2016 yılları arasındadır. Taranan yayınların; kullandıkları yöntemler ve algoritmalar, mimari yapıları, araç seçimleri, uygulama alanları, elde ettikleri sonuçlar incelenmiş ve bu makalede sunulmuştur.

4 Bulut bilişimde sınıflandırma yöntemi

Sınıflandırma, örneklem verinin daha önceden belirlenmiş sınıflardan birine atanmasını sağlayan modeli oluşturan bir yöntemdir. Veri madenciliğinin en sık kullanılan yöntemlerinden biridir. Başlıca sınıflandırma teknikleri; karar ağacı, k-en yakın komşu, Bayesian, karar destek makinaları ve

yapay sinir ağlarıdır. Tüm bu sınıflandırma tekniklerinin bulut platformunda belirli bir veriseti üzerinde uygulanarak karşılaştırıldığı çalışmalar bulunmaktadır [16]. Elde edilen sonuçlar, karar ağacı ve Bayesian metotlarının en hızlı, yapay sinir ağlarının ise en yavaş çalışan teknikler olduğunu, ancak algoritmaların çalışma sürelerinin bulut bilişim sayesinde iyileştirilebildiğini göstermektedir.

Bazı çalışmalar [16], sınıflandırma algoritmalarını orijinal halleri ile bulut bilişim altyapısında kullanırken, bazı çalışmalar [17]-[19] daha iyi sonuçlar elde edebilmek amacıyla algoritmalarda eklentiler/değişiklikler gerçekleştirmiştir. Örneğin; Rahulamathavan ve diğ. [17] karar destek makinalarını, bulut bilişim altyapısı üzerinde gerçekleştirirken çoklu sınıflandırma özelliği eklemiş, Lvshuhong [18] ise ceza faktörü kullanarak sonuçların daha da iyileştirilebildiğini göstermiştir. Benzer şekilde, Dai ve Sun [19] yayınlamış oldukları çalışmada, Naive Bayes sınıflandırma algoritmasını kaba küme teorisi ile birleştirerek yeni bir entegre algoritma oluşturmuşlar ve bulut platformu üzerinde uygulayarak test etmişlerdir. Metin sınıflandırma üzerine yapılan deneysel çalışmalar sonucunda önerilen algoritmanın yüksek çalışma performansı gösterdiği ve doğruluk oranının %74.9'dan %77.6'ya çıktığı görülmektedir.

Temeli Bayes teoremine dayanan Bayesian sınıflandırıcı, diğer sınıflandırma algoritmaları arasında en çok bilineni ve uygulaması en kolay olanıdır. Bu sebeple, bulut bilişim tabanlı birçok sınıflandırma çalışmalarında [19]-[21] Naive Bayes algoritmasının kullanıldığı görülmektedir. Zhou ve diğ. [20] Naive Bayes sınıflandırma algoritmasını MapReduce modeli ile bulut platformunda paralel bir şekilde çalıştırmayı önermektedir. Paralel tasarımın, hem algoritmaların etkin çalışmasını hem de algoritmanın büyük veriler üzerinde çalışabilmesine olanak sağladığı görülmektedir.

Bulut bilişimde karar ağacı sınıflandırma algoritmasının da kullanıldığı çalışmalar [3],[16],[22] mevcuttur. Tüm bu çalışmalar, karar ağacı algoritmasının bulut platformunda dağıtık işleme, çoklu işleme ve gerçek zamanlı uygulama geliştirebilme örneklerini içermektedir. Bulut tabanlı sınıflandırma işleminde birden fazla karar ağacı kullanılarak doğruluk oranını yükseltmek de hedeflenebilmektedir. [23] Rastgele orman algoritması, tek bir karar ağacı üretmek yerine, farklı eğitim verileriyle çok sayıda ağacın oluşturulmasını ve ortaya çıkan kararların oylanarak birleştirilmesini önerir. Farklı eğitim kümeleri oluştururken ön yükleme ve rastgele özellik seçimi kullanılır.

Veri madenciliği algoritmalarının bulut bilişim ile entegre olarak çalıştırıldığı alanların başında metin sınıflandırma gelmektedir [19],[21],[22],[24]. Sarnovsky ve Kacur [22] doküman kategorize etmede dağıtık karar ağacı algoritmasını önermiş ve bulut platformunda GridGain sistemi ile kullanmıştır. Verilerin dağıtık bir şekilde çalıştırılmasının sağlandığı, düğüm sayısındaki artışın işlem performansını arttırdığı, böylece sonuç alma süresinde gittikçe azalan bir iyileşme olduğu görülmektedir. Bir diğer çalışmada ise, Tan ve diğ. [21] Naive Bayes algoritması ile MapReduce modeli üzerinde metin sınıflandırma işlemi gerçekleştirmiştir. Deneysel çalışmalarda, üç farklı yaklaşım (Hadoop, Topluluk (Ensemble) ve Örnekleme) ile oluşturulan sınıflandırıcılar karşılaştırılmıştır. Elde edilen sonuçlara göre en yüksek doğruluk oranı ortalamasına topluluk yaklaşımı ile ulaşıldığı görülmektedir.

Bulut bilişimde sınıflandırma yöntemlerinin kullanılmasında genellikle Hadoop ve MapReduce [19],[20],[21],[25],[26] sistemlerinin kullanıldığı görülmektedir. *Hadoop*, sunuculardan oluşturulan yapılar üzerinde, büyük miktarlardaki verilerin dağıtık olarak çalıştırılmasını sağlayan platformdur. *MapReduce*, dağıtılan iş parçacıklarının aynı anda işlenmesi ve bu işlemler sonrası oluşan verilerin tekrar bir araya getirilmesi modelidir. Bu teknolojilerin kullanılmasıyla büyük verilerin daha az maliyetle daha hızlı işlenmesi sağlanabilmektedir.

Bulut bilişimde verilerin, geniş bir ağ şeklinde konumlanan dağıtık makineler ve kaynaklardan oluşan bir ortama yüklenmesi söz konusudur. Bulut kavramı, depolanan verilere her an her yerden ulaşma imkânı sağlamaktadır. Verilerin gizliliğini ve güvenliğini sağlamak bulut sağlayıcısının sorumlulukları arasındadır. Ancak veri madenciliğinde kullanılacak verilerin buluttaki güvenliğini arttırmak amacıyla şifreleme gibi çeşitli yöntemler öneren çalışmalar bulunmaktadır [27]. Bilgi güvenliği konusunda dikkat edilmesi gereken bir diğer konu da mahremiyet korumadır (privacy preserving). Veri madenciliği sonucunda mahrem olabilecek kişisel veya kurumsal bilgiler ortaya çıkabilmekte, bu bilgiler kötü niyetli kişiler tarafından kullanılabilmektedir. Gerek doğru veri analizi, gerekse mahremiyetin korunması için bulutta gerçekleştirilen sınıflandırma çalışmalarında [17],[27] önerilen başlıca yöntemlerin şunlar olduğu görülmektedir: veri karartma tekniği, veri değiştirme yöntemi, şifreleme metodu, k-anonymity ve l-diversity modelleri.

Tablo 1'de bulut bilişimde sınıflandırmanın uygulandığı çalışmalar listelenmektedir. Tablo içerisinde, çalışmaların gerçekleştirildiği yıl, kullanılan algoritmalar, sistem yapısı, uygulandığı alan, değerlendirme ölçütü ve deneysel çalışmalarda elde edilen doğruluk oranları sunulmaktadır.

5 Bulut bilişimde kümeleme yöntemi

Kümeleme işlemi, veri seti içerisinde birbirine benzer objeleri aynı grup içerisinde toplama yöntemidir. Büyük miktarlardaki veriler üzerinde kümeleme işlemi yapmak ciddi bir zaman gerektirmektedir. Yüksek performanslı ve ekonomik bir şekilde kümeleme işlemini gerçekleştirmek amacıyla bulut tabanlı veri madenciliği çalışmaları önerilmektedir [28]. Bulut üzerindeki kümeleme çalışmalarının farklı amaçlar için yapıldığı görülmektedir. Örneğin; sağlık hizmetleri [29], görüntü analizi [30], nesneleri kalitelerine göre gruplama [28], telekomünikasyon servisleri [31] ve doküman kümeleme [32]. Farklı kümeleme yaklaşımlarının bulut üzerinde uygulandığı çalışmalar mevcuttur, örneğin *bölünmeli* [33], *hiyerarşik* [34] ve *yoğunluk tabanlı* [35].

Tablo 2'de gösterildiği gibi, bulut bilişim tabanlı kümeleme yönteminde sıklıkla kullanılan algoritma K-Means algoritmasıdır. Bazı çalışmalar [36],[29] K-Means algoritmasını bulut ortamına uyarlayarak taşırken, bazı çalışmalar [37] ise paralel versiyonunu geliştirmişlerdir. Zhong ve diğ. [37] mevcut K-Means algoritmasını iyileştirerek, paralel çalışacak hale getirip buluttaki büyük veriler üzerinde uygulamıştır. Deneysel çalışmalar sonucunda, önerdikleri algoritmanın hem normal K-Means algoritmasına, hem de CLARA ve RICA gibi diğer kümeleme algoritmalarına kıyasla daha hızlı çalıştığını göstermişlerdir. Ayrıca, daha önce sonuç alınamayan bazı büyük verilerde iyileştirilmiş paralel K-Means algoritması ile artık sonuç alınabildiği de görülmektedir.

Tablo 1: Bulut platformunda gerçekleştirilen sınıflandırma çalışmaları.

Referans	Yıl	Algoritmalar	Sistem, Mimari Yapı, Araçlar	Kullanım Alanı	Değerlendirme Ölçütü	Doğruluk Oranları
Quirita ve diğ. [23]	2016	Rastgele orman, Karar destek makinaları	Hadoop, MapReduce, Weka	Uzaktan Algılama	Çapraz doğrulama	%78.26 %64.41
Yuechao ve Shihe [26]	2016	Karar ağacı, K-en yakın komşu, Bayesian, Karar destek makinaları, Yapay sinir ağları	Hadoop ve MapReduce	Büyük veri analizi	---	---
Lvshuhong [18]	2015	Karar destek makinaları	MapReduce	Büyük veri analizi	Basit doğrulama (Holdout)	%80 - %90
Samanthula ve diğ. [27]	2015	K-en yakın komşu	Genel bulut	Gizlilik koruma ve veri güvenliği	---	---
Ayma ve diğ. [13]	2015	Karar destek makinaları	Hadoop ve MapReduce	Görüntü sınıflandırma	Basit doğrulama (Holdout)	%78.26
Yuan ve Herbert [16]	2014	Karar ağacı, K-en yakın komşu, Bayesian ağları, Karar destek makinaları, Yapay sinir ağları	Windows Azure, Weka	Mobil uygulama ile aktivite tanıma	Çapraz doğrulama	%95 ve üstü
Rahulamathavan ve diğ. [17]	2014	Karar destek makinaları	Genel bulut	Gizlilik koruyucu ve çoklu sınıflandırma	Çapraz doğrulama	mak. %98.24 min. %86.98
Dai ve Sun [19]	2014	Naive Bayes , Rough set	Hadoop ve MapReduce	Metin sınıflandırma	Doğruluk oranı, Çağırma (Recall)	%77.6 %76.1
Pranckevičius [24]	2013	Naive Bayes	Hadoop, MapReduce, Mahout	Metin sınıflandırma	---	---
Zhou ve diğ. [20]	2012	Paralel Naive Bayes	Hadoop ve MapReduce	Büyük veri analizi	---	---
Sarnovsky ve Kacur [22]	2012	Karar ağacı	GridGain, Jbowl	Metin sınıflandırma	---	---
Ding and Yang [25]	2012	Genetik Algorithm, Sprint	MapReduce	Göğüs kanseri	Basit doğrulama (Holdout)	%90 ve üzeri
Tan ve diğ. [21]	2010	Naive Bayes	Hadoop ve MapReduce	Metin sınıflandırma	Basit doğrulama (Holdout)	mak. %82.88 ort. %69.14

Tablo 2: Bulut platformunda gerçekleştirilen kümeleme çalışmaları.

Referans	Yıl	Algoritmalar	Sistem, Mimari Yapı, Araçlar	Kullanım Alanı
Rallapalli ve diğ. [29]	2016	K-Means	Hadoop, Mahout, NoSQL	Sağlık
Haut ve diğ. [30]	2016	Paralel K-Means++	Spark, Scikit-learn, OpenStack	Görüntü analizi
Zhang ve Shou [33]	2016	K-Means ve Canopy	Hadoop ve MapReduce	Büyük veri analizi
Kumari ve Babu [38]	2015	CLARA, CLARANS, K-Means, K-Medoids	Hadoop ve MapReduce	Büyük veri analizi
Li ve diğ. [28]	2015	K-Means	Hadoop ve MapReduce	Kömür kalitesi belirleme
Chen ve Chen [35]	2015	DBSCAN	MapReduce, CUDA, Hadoop, Spark	Büyük veri analizi
Zhong ve diğ. [37]	2014	Paralel K-Means, CLARA, RICA	Hadoop, MapReduce, MATLAB	Tünel gözleme
Masih ve Tanwani [36]	2014	K-Means	Hadoop, MapReduce, Özel Bulut	Büyük veri analizi
Golghate ve Shende [32]	2014	Paralel K-Means	Hadoop, MapReduce, Toplu Senkron Paralellik, Hama	Doküman kümeleme
Srivastava ve diğ. [34]	2013	Hiyerarşik	StarCluster, Amazon EC2, MySQL, Sun Grid Engine	Sağlık

Diğer veri madenciliği yöntemlerinin bulut uygulamalarında olduğu gibi, kümelemede de MapReduce sistemi oldukça sık kullanılmaktadır. Kumari ve Babu [38] gerçekleştirdikleri çalışmalarında yüksek hacimli veri üzerinde kümeleme işleminin uygulanmasında *haritalama* (map) ve *indirgeme* (reduce) modelinden yararlanmışlardır. Kümeleme işlemi için K-Means, K-Medoids, CLARA ve CLARANS algoritmalarını kullanmışlardır. Tüm bu algoritmalar arasında en karmaşık yapıya sahip olan CLARA algoritmasının bile büyük veri setleri üzerinde etkin bir şekilde uygulanabileceğini belirtmişlerdir.

Çalışmaların çoğunda [28],[36],[33], büyük miktarda veri üzerinde tek başına yetersiz kalan K-Means algoritmasının MapReduce modeli ile performanslı çalıştırılması önerilmektedir. Ancak, alternatif yöntem olarak toplu senkron

paralellik (bulk synchronous parallelism (BSP)) modelini kullanan ve MapReduce modeline göre daha etkin çalıştığını belirten makaleler de bulunmaktadır [32]. Buna sebep olarak, BSP'de her bir hesaplama adımında yeni bir iş gönderilmemesi ve BSP adımlarının MapReduce'taki iş birimlerine göre daha hızlı çalışıyor olması gösterilmektedir.

6 Bulut bilişimde birliktelik kuralları analizi

Birliktelik kuralları analizi, veriler arasındaki ilginç birlikteliklerin, ilişkilerin ve bağıntıların kurallar halinde bulunması işlemidir. Bu yöntemin en çok kullanıldığı alan pazarlama, özellikle de sepet analizidir. Sepet analizi ile büyük veriler incelenerek müşterilerin alışveriş davranışları keşfedilmeye çalışılmaktadır. Elde edilecek olan bilgi; ürün tavsiye, raf yerleşimi, promosyon stratejileri, reklam kataloğu tasarımı gibi farklı amaçlar için kullanılabilmektedir. Büyük miktarlardaki veriler üzerinde sepet analizini yüksek performanslı yapabilmek için bulut teknolojisinden yararlanan çalışmalar [39],[40] mevcuttur. Birliktelik kuralları analizinin bulut tabanlı olarak yapıldığı başka alanlar da bulunmaktadır. Örneğin; sağlık [41], sosyal ağ [42], web madenciliği [43] ve ulaşım [44]. Chuchra ve diğ. [45] çalışmalarında, bileşen tabanlı sistemler için veri madenciliğinin uygulanabilirliğinden, ürünler arasındaki ilişkiyi belirlemek için birliktelik kuralları analizi yapılabileceğinden, bu yaklaşıma *ürün madenciliği* isminin verildiğinden ve bulut bilişimin buna katkısından bahsetmişlerdir. Ayrıca, bulut platformunda günlük (log) analizi gerçekleştirerek birliktelik kurallarını çıkartan çalışmalar [46],[47] da mevcuttur.

Birliktelik kuralında, öğeler arasındaki birliktelik, destek (support) ve güven (confidence) ölçütleri kullanılarak hesaplanır. *Destek* kriteri, veride öğeler arasındaki bağıntının ne kadar sık olduğunu, *güven* kriteri ise kuralın kuvvetini belirtir. Bulut tabanlı gerçekleştirilen çoğu çalışma [40],[43],[47] bu iki kriteri kullanırken, bazı çalışmalar ise önem değeri (lift) [41],[48], çoklu destek [49] gibi daha fazla kriterden yararlanmışlardır.

Tablo 3'te gösterildiği üzere, bulut tabanlı gerçekleştirilen birliktelik kuralları analizlerinde; Apriori [50], Fp-Growth [48], Eclat [44] gibi farklı algoritmaların kullanıldığı çalışmalar mevcuttur. Ama en çok kullanılan algoritma Apriori algoritmasıdır. Bazı çalışmalar [40],[45] Apriori algoritmasını

bulut ortamına direk taşırken, bazı çalışmalar [42],[51] iyileştirme yaparak kullanmış, bazı çalışmalar [46],[47] ise paralel çalışabilir versiyonunu geliştirmiştir. Yu ve Wang [46] tarafından gerçekleştiren deneysel çalışmalar paralel Apriori algoritmasının, normal Apriori algoritmasına göre bulut üzerinde ortalama %45 daha hızlı sonuç ürettiğini göstermektedir.

Eğer veriseti içerisinde farklı öğe sayısı fazlaysa ve aralarında sık tekrarlanan ilişkiler varsa, öğelerin kombinasyonu sonucu oldukça fazla sayıda yaygın örüntü oluşabilmektedir. Bütün yaygın öğe kümelerinin tek tek yazılması yerine, örüntü sayısının azaltılması amacıyla, birden fazla öğe kümesini temsil eden sıkıştırılmış gösterimler yapılabilmektedir. Yaygın öğe kümelerinin verimli bir şekilde saklanmasında ve temsil edilmesinde kullanılan yöntemler: *kapalı yaygın öğe kümesi* (closed frequent itemset) ve *maksimal yaygın öğe kümesi*dir (maximal frequent itemset). Eğer bir öğe kümesi yaygınsa ve aynı destek seviyesinde bu öğe kümesini kapsayan başka bir yaygın öğe kümesi yoksa, *kapalı* olarak adlandırılır. Eğer bir öğe kümesi yaygınsa ve bu öğe kümesinin yaygın bir üst kümesi yoksa, *maksimal* olarak isimlendirilir. Bulut ortamında da kapalı örüntü [52] ve maksimal örüntü [53] kavramları uygulayan çalışmalar bulunmaktadır.

Diğer veri madenciliği yöntemlerinin bulut üzerinde uygulanmasında olduğu gibi, birliktelik kuralları analizinde de Hadoop ve MapReduce sistemleri oldukça sık kullanılmaktadır [42],[46]-[48],[51]-[53]. Ayrıca, Wu ve diğ. [54] gerçekleştirdikleri çalışmada, ardışık düzen (pipelined) MapReduce modelini önermiş, Apriori algoritması ile denemiş, klasik MapReduce modeli ile karşılaştırmış ve sonuç olarak önerdikleri yaklaşımın daha başarılı olduğunu belirtmişlerdir. Alternatif olarak sector/sphere platformunu tercih eden çalışmalar [43] da mevcuttur. Lal ve Mahanti [43], birliktelik kuralı analizinin sector/sphere bulut ortamında nasıl gerçekleştirilebileceğini anlatmış ve çok sayıda öğe arasındaki ilişkilerin hızlı ve etkili bir şekilde ortaya çıkarabildiğini belirtmiştir. Bulut ortamında kişi, kurum ve işletmelerin sahip oldukları veriler, sahtekarlık, bilgi hırsızlığı, bilgisayar korsanlığı, elektronik saldırı, bilgi sızdırma gibi tehdit ve tehlikelerle karşı karşıyadır. Bulut hizmeti sağlayıcının aldığı güvenlik önlemlerini daha da arttırmak, aynı zamanda Veri madenciliği sonuçlarının gizliliğini ve kişisel mahremiyeti korumak amacıyla gerçekleştirilen çalışmalar bulunmaktadır. Yi ve diğ. [50] birliktelik kuralları analizinde anonimleştirme, şifreleme gibi yöntemler önermekte ve öğe, işlem, veritabanı olmak üzere üç farklı seviyede gizliliği sağlamaya yönelik algoritmalar sunmaktadır.

Birliktelik kuralları analizi için bulut ortamında geliştirilmiş olan servisler bulunmaktadır. Apiletti ve diğ. [48] çalışmasında önermiş olduğu SEARUM isimli servis, bulut üzerinde çalışan dağıtık MapReduce iş parçalarından oluşmaktadır. Her bir iş parçası, birliktelik kuralları analizi tekniğinin bir işlemini gerçekleştirmektedir. Deneysel çalışmalarda, geliştirilen servis büyük ağ verileri üzerinde denenmiş ve sonuçların daha hızlı elde edilebildiği gösterilmiştir. Birliktelik kurallarının yanı sıra ardışık örüntüleri bulut platformunda ortaya çıkartan çalışmalar [49] da mevcuttur. Birliktelik kuralları eş zamanlı olarak gerçekleşen ilişkilerin tanımlanmasında kullanılır. *Ardışık zamanlı örüntüler* ise birbiri ile ilişkisi olan ancak belirli bir periyotta, birbirini izleyen dönemlerde ve belirli bir sıklıkla gerçekleşen olayların tanımlanmasını sağlar. Başka bir deyişle, olayların belirli bir zaman ekseninde tekrarlarını inceleme sonucunda bağıntıları ortaya çıkartır. Chen ve diğ. [49] fiyat dalgalanmaları üzerine bulut platformu yardımıyla paralel ardışık örüntü analizi gerçekleştirmişlerdir. Çalışmada, veri ön işlem aşaması, MapReduce sistemi ile Hadoop platformunda uygulanması, uygun maliyet ile sonuçların elde edilebildiği anlatılmaktadır.

7 Sonuçlar ve gelecek çalışmalar

Veriler içerisinden daha önce keşfedilmemiş yararlı bilgilerin çıkarılması ve bu bilgiler doğrultusunda daha doğru kararlar alınabilmesi, özellikle iş sektöründe oldukça büyük bir ihtiyaçtır. Bu amaç doğrultusunda, iş sektörünün yanı sıra sağlık, bankacılık ve pazarlama gibi birçok sektörde de veri madenciliği tekniklerine sıklıkla başvurulmaktadır. Veri madenciliğinin temel yöntemleri olan sınıflandırma, kümeleme ve birliktelik kuralları analizi kendi içerisinde farklı algoritmalar sunmaktadır. Yapılacak çalışma doğrultusunda uygun yöntem ve algoritma belirlenerek, eldeki veri seti üzerine uygulanmaktadır.

Tablo 3: Bulut platformunda gerçekleştirilen birliktelik kuralları analizleri.

Referans	Yıl	Algoritmalar	Sistem, Mimari Yapı, Araçlar	Kullanım Alanı
Liang ve diğ. [41]	2016	Apriori	Microsoft Azure	Sağlık
Sahay ve diğ. [53]	2016	Apriori, Pincer-Search	Hadoop ve MapReduce	Büyük veri analizi
Yi ve diğ. [50]	2015	Apriori	Genel bulut	Gizlilik koruma ve veri güvenliği
Zheng and Wang [44]	2014	Eclat	MapReduce	Ulaşım
Farzanyar ve Cercone [42]	2013	İyileştirilmiş MapReduce Apriori	Hadoop ve MapReduce	Sosyal ağ
Wang ve Li [47]	2013	Paralel Apriori	Hadoop ve MapReduce	Günlük (log) analizi
Chuchra ve diğ. [45]	2013	Apriori	Genel bulut	Bileşen tabanlı sistemler
Nappina ve Revathi [39]	2013	Birliktelik kuralları analizi	Hadoop	Pazarlama (sepet analizi)
Apiletti ve diğ. [48]	2013	Paralel FP-Growth	Hadoop ve MapReduce	Ağ trafiği analizi
Wang ve diğ. [52]	2012	AFOPT-closed	Hadoop ve MapReduce	Büyük veri analizi
Yu ve Wang [46]	2012	Apriori ve Paralel Apriori	Hadoop, MapReduce, HDFS	Günlük (log) analizi
Wu ve diğ. [54]	2012	Apriori	Hadoop ve MapReduce	Büyük veri analizi (örnek çalışma sepet analizi)
Li ve Zhang [51]	2011	Apriori	Hadoop ve MapReduce	Büyük veri analizi
Lal ve Mahanti [43]	2010	Apriori	Sector/Sphere	Anlamsal web tabanlı veri analizi

Fakat üretilen ve paylaşılan veri miktarındaki artış, yüksek hacime sahip veri içerisinden bilgi keşfini zorlaştırmakta, etkin araç, yöntem ve stratejilerin geliştirilmesini zorunlu kılmaktadır. Bu sebeple, son yıllarda veri madenciliği tekniklerinin büyük veri setleri üzerinde uygulanabilmesi için bulut bilişim teknolojisinden faydalanıldığı ve oldukça başarılı sonuçlar elde edildiği görülmektedir.

Bulut bilişim teknolojisi kullanıcıya yüksek boyutta veri saklama alanı, paralel işlem yeteneği, her yerden ulaşılabilirlik gibi birçok özelliği uygun maliyetle sunmaktadır. Bu makale, bulut bilişimde sınıflandırma, kümeleme ve birliktelik kuralları analizi tekniklerinin uygulandığı çalışmaları derlemektedir. Makalenin amacı, bulut platformundaki veri madenciliği uygulama alanlarını, geliştirilen çözümleri, algoritmalarda yapılan iyileştirmeleri ve elde edilen sonuçları bir bütün olarak sunmaktır. Sonuç olarak gerçekleştirilen bu çalışmalar göstermiştir ki, algoritmaların büyük veriler üzerinde bulut bilişim teknolojisi sayesinde paralel bir şekilde işlenmesi, yük dağılımının yapılması ve böylece çalışma süreleri kısaltılarak performansın büyük ölçüde arttırılması mümkündür.

Gelecek çalışma olarak; literatürdeki çalışmalarda kullanılan algoritmalar geliştirilerek hem çalışma performansı hem de doğruluk oranı daha yüksek çözüm yaklaşımları elde edilebilir. Özellikle son yıllarda kullanılan, toplu öğrenme (ensemble learning) yaklaşımı, yani birden fazla model oluşturularak daha doğru sonuçlar elde etmeyi sağlayan çözümler bulut tabanlı olarak gerçekleştirilebilir. Bulut tabanlı kümeleme çalışmalarda çoğunlukla K-Means algoritmasının kullanıldığı görülmektedir. Ancak daha yüksek doğruluk ve çalışma hızı sunan K-Means++ algoritmasının gelecek çalışmalarda kullanılması daha iyi sonuçlar elde edilmesini sağlayacaktır.

Diğer bir gelecek çalışma olarak, son yıllarda popüler olan nesnelerin interneti, arttırılmış gerçeklik gibi konularla bulutta veri madenciliği çalışmaları birleştirilebilir. Bulut tabanlı veri madenciliği çalışmalarının eğitim, tıp, ulaşım, sosyal medya gibi alanlarda uygulandığı görülmektedir. Tekstil, tarım, denizcilik gibi farklı konularda çalışılarak uygulama alanı daha da genişletilebilir. Gelecek çalışma olarak, veriler arasındaki anlamsal ilişkileri çıkarmak ve özellikle sektörel çalışmalarda doğruluğu yüksek karar destek sistemleri sunmak adına ontoloji tabanlı veri madenciliği içeren bulut bilişim sistemleri geliştirilebilir.

8 Kaynaklar

[1] Petre R-Ş. "Data mining in cloud computing". *Database Systems Journal*, 3(3), 67-71, 2012.

[2] Low Y, Gonzalez J, Kyrola A, Bickson D, Guestrin C, Hellerstein JM. "Distributed GraphLab: A framework for machine learning and data mining in the cloud". *Proceedings of the VLDB Endowment*, 5(8), 716-727, 2012.

[3] Talia D, Trunfio P. "How distributed data mining tasks can thrive as knowledge services". *Communications of the ACM*, 53(7), 132-137, 2010.

[4] Kholod I, Kuprianov M, Petukhov I. *Parallel and Distributed Data Mining in Cloud Advances in Data Mining*. Editor: Perner P. Advances in Data Mining. Applications and Theoretical Aspects, 349-362, New York, NY, USA, Springer, 2016.

[5] Olaide AA. "On modeling confidentiality archetype and data mining in cloud computing". *African Journal of Computing & ICT*, 6(1), 79-86, 2013.

[6] Kamala B, "A study on integrated approach of data mining and cloud mining". *International Journal of Advances in Computer Science and Cloud Computing*, 1(2), 35-38, 2013.

[7] Fernández A, Peralta D, Benítez JM, Herrera F. "E-learning and educational data mining in cloud computing: an overview". *International Journal Learning Technology,* 9(1), 25-52, 2014.

[8] Anand A, Nithin Chandran R, Abhijith T, Varun KK. "Data mining over encrypted data for patient records". *International Journal on Applications in Engineering and Technology*, 2(4), 13-16, 2016.

[9] Dahmani D, Rahal SA, Belalem G. "Improving the performance of data mining by using big data in cloud environment". *Journal of Information & Knowledge Management,* 2016.

[10] Belcastro L, Marozzo F, Talia D, Trunfio P. "Using scalable data mining for predicting flight delays". *ACM Transactions on Intelligent Systems and Technology*, 8(1), 2016.

[11] Vrbic R. "Data mining and cloud computing", *Journal of Information Technology and Applications*, 2(2), 75-87, 2012.

[12] Kamdar AB, Jagani JM. "A survey: classification of huge cloud datasets with efficient map-reduce policy". *International Journal of Engineering Trends and Technology (IJETT)*, 18(2), 103-107, 2014.

[13] Ayma VA, Ferreira RS, Happ P, Oliveira D, Feitosa R, Costa G, Plaza A, Gamba P. "Classification algorithms for big data analysis, a Map Reduce approach". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-3(W2), 17-21, 2015.

[14] Zhengqiao X, Dewei Z. "Research on clustering algorithm for massive data based on Hadoop platform". *International Conference on Computer Science and Service System*, Nanjing, China, 11-13 August 2012.

[15] Mahendra TV, Deepika N, Rao NK. "Data mining for high performance data cloud using association rule mining". *International Journal of Advanced Research in Computer Science and Software Engineering,* 2(1), 1-6, 2012.

[16] Yuan B, Herbert J. "A cloud-based mobile data analytics framework: case study of activity recognition using smartphone". *2nd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering*, Oxford, UK, 2014.

[17] Rahulamathavan Y, Phan RCW, Veluru S, Cumanan K, Rajarajan. "Privacy-Preserving multi-class support vector machine for outsourcing the data classification in cloud". *IEEE Transactions on Dependable and Secure Computing*, 11(5), 467-479, 2014.

[18] Lvshuhong. "Improved SVM in cloud computing information mining". *International Journal of Grid Distribution Computing*, 8(1), 33-40, 2015.

[19] Dai Y, Sun H. "The Naive Bayes text classification algorithm based on rough set in the cloud platform". *Journal of Chemical and Pharmaceutical Research*, 6(7), 1636-1643, 2014.

[20] Zhou L, Wang H, Wang W. "Parallel implementation of classification algorithms based on cloud computing environment". *Indonesian Journal of Electrical Engineering*, 10(5), 1087-1092, 2012.

[21] Tan AX, Liu VL, Kantarcioglu M, Thuraisingham B. "A comparison of approaches for large-scale data mining-utilizing MapReduce in large-scale data mining". Department of Computer Science, The University of Texas at Dallas, Dallas, Texas, Technical Report, UTDCS-24-10, 2010.

[22] Sarnovsky M, Kacur T. "Cloud-based classification of text documents using the Gridgain platform". *7th IEEE International Symposium on Applied Computational Intelligence and Informatics*, Timisoara, Romania, 24-26 May 2012.

[23] Quirita VAA, Costa GAOP, Happ PN, Feitosa RQ, Ferreira RS, Oliveira DAB, Plaza A. "A new cloud computing architecture for the classification of remote sensing data". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 99, 1-8, 2016.

[24] Pranckevičius T. "Investigation of cloud computing technology on the visualisation and classification algorithms, Institute of Mathematics and Informatics, Vilnius University, Vilnius, Lithuania, Technical Report, MII-DS-07T-13-5, 2013.

[25] Ding J, Yang S. "Classification rules mining model with genetic algorithm in cloud computing". *International Journal of Computer Applications*, 48(18), 24-32, 2012.

[26] Yuechao W, Shihe C. "An approach to smart grid online data mining based on cloud computing". *International Journal of Simulation Systems, Science & Technology*, 17(2), 1-5, 2016.

[27] Samanthula BK, Elmehdwi Y, Jiang W. "K-nearest neighbor classification over semantically secure encrypted relational data". *IEEE Transactions on Knowledge and Data Engineering*, 27(5), 1261-1273, 2015.

[28] Li Z, Song X, WenhuiZhu, YanxiaChen. "K-means clustering optimization algorithm based on MapReduce". *International Symposium on Computers & Informatics (ISCI 2015)*, Beijing, China, 17-18 January 2015.

[29] Rallapalli S, Gondkar RR, Rao GVM. *Cloud Based K-Means Clustering Running as a MapReduce Job for Big Data Healthcare Analytics Using Apache Mahout*. Editors: Satapathy SC, Mandal JK, Udgata SK, Bhateja V. Advances in Intelligent Systems and Computing, 127-135, Springer, 2016.

[30] Haut JM, Paoletti M, Plaza J, Plaza A. "Cloud implementation of the k-means algorithm for hyperspectral image analysis". *The Journal of Supercomputing*, 73(1), 514-529, 2016.

[31] Kamalraj N, Malathi A. "Hadoop operations management for big data clusters in telecommunication industry". *International Journal of Computer Applications*, 105(12), 40-44, 2014.

[32] Golghate AA, Shende SW. "Parallel k-means clustering based on Hadoop and Hama". *International Journal of Computing and Technology*, 1(3), 33-37, 2014.

[33] Zhang D, Shou Y. *An Improved Parallel K-Means Algorithm Based on Cloud Computing*. Editors: Li K, Li J, Liu Y, Castiglione A. Computational Intelligence and Intelligent Systems, 312-320, Springer, 2016.

[34] Srivastava K, Shah R, Valia D, Swaminarayan H. "Data mining using hierarchical agglomerative clustering algorithm in distributed cloud computing environment". *International Journal of Computer Theory and Engineering*, 5(3), 520-522, 2013.

[35] Chen C-C, Chen M-S. "HiClus: Highly Scalable Density-based Clustering with Heterogeneous Cloud". *Procedia Computer Science,* 53, 149-157, 2015.

[36] Masih S, Tanwani S. "Distributed framework for data mining as a service on private cloud". *International Journal of Engineering Research and Applications*, 4(11), 65-70, 2014.

[37] Zhong L, Tang K, Li L, Ye J. "An improved clustering algorithm of tunnel monitoring data for cloud computing". *The Scientific World Journal,* 2014, 1-6, 2014.

[38] Kumari MC, Babu PN. "Survey on clustering on the cloud by using map reduce in large data applications". *International Journal of Engineering Trends and Technology (IJETT)*, 21(8), 392-395, 2015.

[39] Nappina V, Revathi N. "Data mining over large datasets using Hadoop in cloud environment". *International Journal of Computer Science & Communication Networks*, 3(2), 73-78, 2013.

[40] Kamalraj R, Kannan AR, Vaishnavi S, Suganya V. "A data mining based approach for introducing products in saas (software as a service)". *International Journal of Engineering Innovation & Research*, 1(2), 210-214, 2012.

[41] Liang Z, Ploderer B, Martell MAC, Nishimura T. *A Cloud-Based Intelligent Computing System for Contextual Exploration on Personal Sleep-Tracking Data Using Association Rule Mining*. Editors: Martin-Gonzalez A, Uc-Cetina V. Intelligent Computing Systems, 83-96, Mérida, México, Springer, 2016.

[42] Farzanyar Z, Cercone N. "Efficient mining of frequent itemsets in social network data based on MapReduce framework". *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Niagar Falls, ON, Canada, 25-28 August 2013.

[43] Lal K, Mahanti NC. "A novel data mining algorithm for semantic web based data cloud". *International Journal of Computer Science and Security*, 4(2), 160-175, 2010.

[44] Zheng X, Wang S, "Study on the method of road transport management information data mining based on pruning Eclat algorithm and MapReduce". *Procedia-Social and Behavioral Sciences*, 138, 757-766, 2014.

[45] Chuchra R, Jindal M, Mehta B. "Role of component based systems in data mining & cloud computing". International *Journal of Emerging Technology and Advanced Engineering*, 3(5), 513-517, 2013.

[46] Yu H, Wang D. "Mass log data processing and mining based on Hadoop and cloud computing". *7th International Conference on Computer Science & Education*, Xiamen, Fujian, China, 14-17 July 2012.

[47] Wang Z, Li H. "Research of massive web log data mining based on cloud computing". *International Conference on Computational and Information Sciences*, Shiyan, Hubai, China, 21-23 June 2013.

[48] Apiletti D, Baralis E, Cerquitelli T, Chiusano S, Grimaudo L. "SeARuM: A cloud-based service for association rule mining". *12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, Melbourne, Australia, 16-18 July 2013.

[49] Chen M, Chiang IJ, Lai CW. "Frequent pattern mining for price fluctuation based on cloud computing". *IEEE International Conference on Granular Computing*, Hangzhou, China, 11-13 August 2012.

[50] Yi X, Rao FY, Bertino E, Bouguettaya A. "Privacy-Preserving association rule mining in cloud computing". *10th ACM Symposium on Information*, New York, NY, USA, Singapore, 14-17 April 2015.

[51] Li L, Zhang M. "The strategy of mining association rule based on cloud computing". *International Conference on Business Computing and Global Informatization*, Shanghai, China, 29-31 July 2011.

[52] Wang SQ, Yang YB, Chen GP, Gao Y, Zhang Y. "MapReduce-based closed frequent itemset mining with efficient redundancy filtering". *12th International Conference on Data Mining Workshops,* Brussels, Belgium, 10-10 December 2012.

[53] Sahay S, Khetarpal S, Pradhan T. "Hybrid data mining algorithm in cloud computing using MapReduce framework". *International Conference on Advanced Communication Control & Computing (ICACCCT).* India, 25-27 May 2016.

[54] Wu Z, Cao J, Fang C, "Data cloud for distributed data mining via pipelined MapReduce". *8th International Workshop, ADMI 2012,* Valencia, Spain, 4-5 June 2012.