

Bilimsel arařtırmalarda veri madencilięi kullanımı

The use of data mining in scientific research

Mehmet Albayrak¹

Gönderilme Tarihi: 04 / 11 / 2016

Kabul Tarihi: 01 / 04 / 2017

Öz

Bu çalışmada son yıllarda önemi giderek artan büyük veri, veri tabanları ve veri ambarlarında veri işleme için kullanılan yöntemler, gözle görünmeyen veri ilişkileri ve bunların veri tabanından keşfi için kullanılan yöntemler anlatılmıştır. Büyük miktarda veriyi incelemek için yapılandırılmış olan veri madencilięi, veri tabanları ile yakın bir ilişkiye sahiptir. Veri madencilięi bir çözüme ulaşmak için izlenecek olan yolu gösteren bir araçtır. Birçok disiplin tarafından farklı amaçlar için veri madencilięi kullanılmaktadır. Ülkemizde farklı seviyelerde, farklı bilim dallarında araştırma ve analiz amaçlı yapılan veri madencilięi çalışmaları özetlenmeye ve farklı uygulamalara değinilmeye çalışılmıştır. Yapılan çalışmalarda kullanılan yazılımlar ve birbirinden farklılıklarına değinilmiştir. Ayrıca veri madencilięinde sıklıkla kullanılan sınıflama ve kümeleme algoritmalarının hangi tür veriler üzerinde kullanılabileceęi, sonuçların yorumlanması ve anlamlandırılmasının önemi üzerinde durulmuştur.

Anahtar sözcükler: Veri madencilięi, bilimsel araştırma, veri ambarı

Abstract

In this study, we have discussed the large data that have become increasingly important in recent years, the methods used for data processing in databases and data warehouses, invisible data relationships and methods used for exploring them from the database. The data mining, which is structured to examine the data in large quantities, has a close relationship with the databases. Data mining is a tool that shows the way to be tracked to achieve a resolution. Data mining is used by many disciplines for different purposes. We have tried to summarize the data mining studies conducted in different branches of science in our country for research and analysis purposes and to refer to different applications. The software used in the studies and the differences between them are mentioned. In addition, the classification and clustering algorithms frequently used in data mining can be used on which kind of data, and the importance of interpretation and understanding of the results is emphasized.

Keywords: Data mining, scientific research, dataware house

1. Giriş

Bilişim Sistemleri kurum veya kuruluşlardaki kişisel kullanıcıların ve grupların veriyi (data) istenen ve faydalı bilgiye (knowledge) çeviren yöntem, uygulama, yöntembilim (methodology) ve algoritmaların bütünüdür. Bilgisayar Bilimi, Bilgisayar Mühendislięi, Yazılım Mühendislięi, Matematik, Elektronik, Fizik, Yönetim Bilimleri, Dil ve Psikoloji alanlarını kuşatmaktadır. Bilişim sadece bir mühendislik alanı olmayıp, gündelik yařantının ve bilimsel alanın bütününe girmiştir. Bilişim Sistemlerinin en büyük faydası, kullanıldığı tüm alanlarda düşük maliyetli ve kaliteli iş-akışı sağlamasıdır.

Son zamanlarda dünyadaki veri miktarının ikiye katlandığı öngörülmektedir. Teknolojik cihazların veri saklama kapasiteleri her geçen gün artmaktadır. Bunun sonucunda bilginin kaydedildięi alan sayısı çoęalmaktadır. Bu nedenle sahip olunan verilerin analizi ve sonuca gidilmesi

¹Yrd.Doç.Dr., Süleyman Demirel Üniversitesi Uzaktan Eğitim MYO, Isparta, Türkiye, mehmetalbayrak@sdu.edu.tr

bu verilerden kestirme yöntemlerinin önemi karar vericiler için giderek çoęalmaktadır. Saęlık ile ilgili veriler, uydulardan alınan veriler, ekonomik veriler, alışveriş verileri, insan kaynakları verileri, endüstriyel veriler vb. veriler elde edilmektedir. Bu alanlar kendi içlerinde de farklı türlere ayrılmaktadır. Örneęin tıbbi veriler ele alındığında metinsel veriler olarak tahlil sonuçları ve raporlar, görsel veriler olarak filmler, ultrason sonuçları, mikroskop verileri ve kameralardan gelen veriler olarak çeşitlenmektedir. Bilgisayar tarafından üretilen veriler tek başına bir anlam ifade etmezler, anlamsızdırlar. Belirli bir amaç için kullanıldığı zaman bu veriler anlam ifade ederler (Kalikov, 2006).

Veri madencilięi, alanı son yıllarda verileri analiz ederek yararlı ve mantıksal bilgileri ortaya çıkarmak için kaynağını istatistik, yapay zekâ ve makine öğrenmeden alarak meydana gelmiştir. Akademik alanda, finans ve iş çevrelerinde yoğun ilgi görmüştür (Coşkun ve Baykal, 2011). Veri madencilięi, sahip olunan verilerden, çok açık olmayan, önceden bilinmeyen ancak sonuç olarak işe yarayacak bilginin elde edilmesidir. Bunun içinde kümeleme, veri özetleme, deęişikliklerin analizi, sapmaların tespiti gibi farklı yaklaşımlar bulunmaktadır. Başka bir anlamda; veri madencilięi veriler arasındaki gizlenmiş olan bağlantı ve eğilimleri ortaya çıkarmaktır (Thuarisingham, 2003).

Bu çalışmada, veri madencilięinin günümüzde geldięi nokta ve veri madencilięinin bilimsel arařtırmalarda kullanımı ve ortaya çıkan uygulamalar arařtırılmıştır.

2. Veri madencilięi nedir?

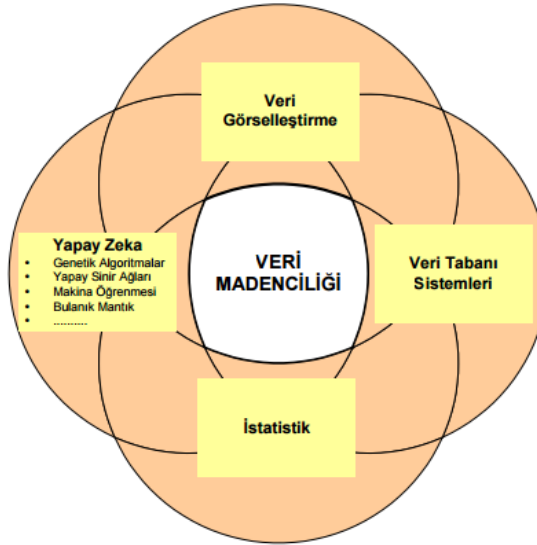
Veri madencilięi, yeni ve mantığa uygun bilgi elde etmek için büyük veri setleri içerisinde birçok adımdan meydana gelen sentezleme işlemi yapmaktır. Veri madencilięi yapmaktaki hedef, veri setleri içerisinde bulunan ilişkilerin üzerinden işlem yapmaktır. Anlamlı bağlantılar ve kurallar elde etmek için, büyük miktardaki verinin irdelenmesi ve üzerinde arařtırma yapılmasıdır. Büyük miktarda verinin irdelenmesini yapmak için ortaya çıkan güçlü yöntemlerden oluşmaktadır. Tek bir veri madencilięi modeli yoktur, bunun yerine bu tekniklerin birbiri ile oluşturdukları birleşimler tercih edilmektedir. Ayrıca çok büyük miktardaki verilerin kendi aralarındaki bağlantıları inceleyerek aralarındaki ilişkiyi elde etmeye yardımcı olan ve veri tabanı yönetim sistemleri içerisinde saklı bulunan verilerin alınmasını saęlayan veri analizi yöntemidir (Kalikov, 2006). Veri madencilięi Şekil-1'de gösterildięi birçok disiplinin birleşmesinden oluşmaktadır.

Şekil 1. Veri madencilięini oluşturan disiplinler



İstatistiksel yöntemler dizisi olarak da veri madencilięini tanımlayabiliriz. Fakat veri madencilięi, geleneksel istatistikten birkaç şekilde deęişiklik arz etmektedir. Veri madencilięinde amaç, rahat bir şekilde mantıksal kurallara veya görsel materyallere dönüştürebilecek nitel modellerin elde edilmesidir. Şekil 2'de vurgulandıęı gibi veri madencilięi kullanım alanları, istatistik, yapay zekâ, veri tabanları ve veri görselleştirme gibi alanlar ile birbiri ile bağlantılıdır (Baykasoęlu, 2005).

Şekil 2. Veri madenciliği ilgili alanları



Veri madenciliği, veritabanı yönetim sistemlerinden ve OLAP sistemlerinden farklı bir yapıdadır. Bahsedilen üç sistem arasındaki fark Tablo-1’de gösterilmektedir.

Tablo 1. VTYS, OLAP ve veri madenciliği farklılıkları

Alan	VTYS	OLAP	Veri Madenciliği
İşlev	Detaylı ve özet bilginin çıkarılması	Özetler, trendler ve tahminler	Gizli bilgilerin ve olayların ortaya çıkarılması
Sonuç Tipi	Bilgi	Analiz	Tahmin ve Sezgi
Yöntem	Genelde Özele(Soruyu sor, veriyle karşılaştır)	Çok boyutlu veri modelleme, İstatistik	Özelden Genele (Model kur, yeni veriye uygula, sonucu elde et)
Örnek Soru	Son 5 yılda kimler ev satın aldı?	Yıl ve bölge başına, ortalama ev satın alma durumu nedir?	Gelecek 2 ay içerisinde kimler ev satın alacak ve hangi sebeple?

3. Veri madenciliği (bilgi keşfi) işleyişi

Veri Madenciliği gelişen olay ve hareketler dizisi olarak tanımlanabilir. Bu işlemde belirleyici etken işleyişi oluşturan uygulamacıdır. İşleyişi oluşturan adımlar düzgün olarak yapılmadığı zaman hedeflenen çıktıya varılamaz. Veri madenciliği algoritmalarından en yüksek düzeyde verim elde etmek için çalışılan konunun ve verinin özelliklerine hâkim olmak gerekmektedir (Savaş vd, 2012). Veri madenciliği izlenen adımlar aşağıda ve Şekil-3’de gösterilmektedir(Shearer, 2000).

1. Problemin tanımlanması: Yapılacak olan işlemin hangi amaca hizmet edeceğini veya hangi problemin çözümü için kullanılacağını iyi belirlenmesi gerekmektedir.

2. Verilerin hazırlanması: Veriler birden fazla ortamda saklanmaktadır. Örneğin; Microsoft verilerini veri ambarlarında ve OLTP veri tabanında depolamaktadır. Bu kısımda önemli olan

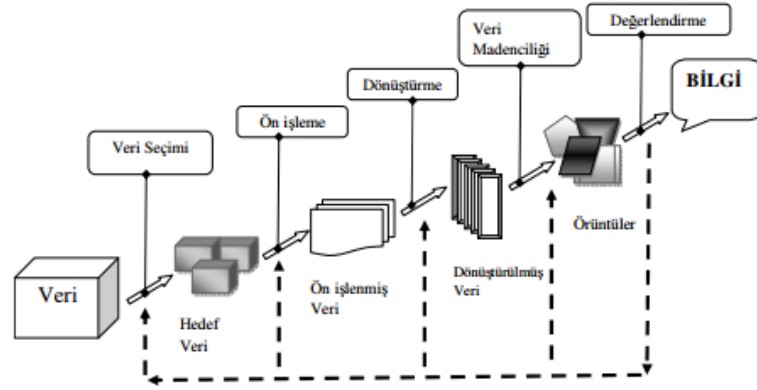
uygun verileri, veri tabanından veya veri ambarlarından alabilmektir. Veri toplama iřlemi tamamlandıktan sonra, verilerin %80'i eęitim, %20'si ise test verisi olmak üzere iki bölüme ayrılır (Tang ve Maclenan, 2005).

3. Modelin kurulması ve deęerlendirilmesi: Modelin doęru seęimi veri madencilięinde çok önemlidir. Doęru model bulununcaya kadar birçok modelde denenebilir. Probleme en uygun bulunduęunda model kurma süreci tamamlanmıř olmaktadır.

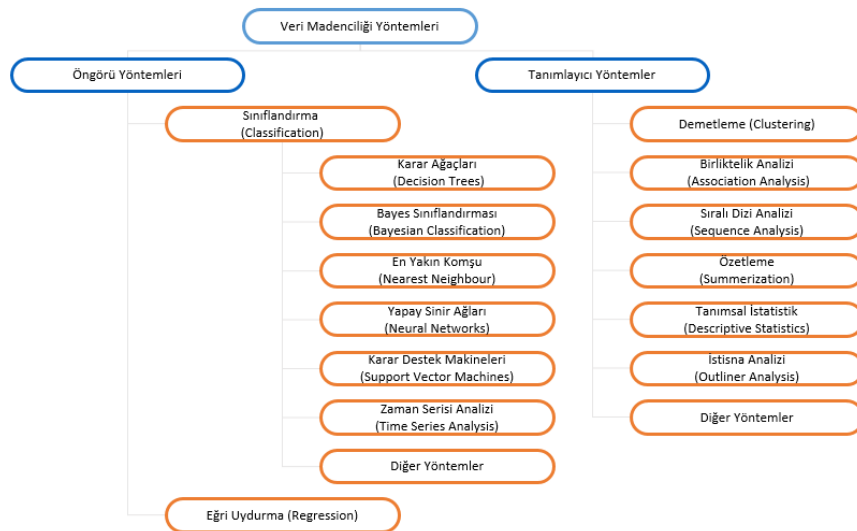
4. Modelin kullanılması: Probleme göre en uygun seęilen model bir program veya programın alt bir paęası olabilmektedir.

5. Modelin izlenmesi: Model çalıřtıķa ortaya çıkardıęı sonuçların zaman ięerisinde izlenerek gerekirse modelde güncelleme yapılabilir.

řekil 3. Veri madencilięi bilgi keřfi süreci



řekil 4. Veri Madencilięi Yöntemleri



Veri madencilięinde farklı yöntemler kullanılmaktadır. Kullanılan yöntemler problem türüne göre, elde bulunan veriye göre ve uygulama alanına göre farklılıklar gösterebilmektedir. Kullanılan yöntemlerin birçoęu yapay zekânın alanıdır. Veri madencilięinde kullanılan yöntemler řekil-4'de gösterilmektedir.

4. Veri madenciliği kullanım alanları

Veri madenciliğinin kullanım alanı son dönemde giderek artmaktadır. İşletmeler, kurum ve kuruluşlar tarafından etkili sonuçlar vermesi, kolay kullanılması nedeniyle sık kullanılmaktadır. Veri madenciliğinin kullanıldığı alanlara örnekler aşağıdaki gibi verilebilir.

- Finans Sektörü: Risk analizleri ve usulsüzlük tespiti.
- Pazarlama: Çapraz satış analizleri, müşteri segmentasyonu.
- Sigortacılık: Müşteri azalmasındaki nedenlerin tespit edilmesi, yanlış işlemlerin önlenmesi.
- İletişim: Yanıltma tespiti, sistemlerde yoğunluk öngörüsü.
- Borsa: Hisse senedi fiyatındaki hareketliliklerinin tahmini, genel piyasa değerlendirmeleri.
- Sağlık: Hastalık teşhis ve tedavi önerileri
- Bilim ve Mühendislik: Ampirik veriler üzerinde modeller kurularak bilimsel ve teknik problemlerin çözümlenmesi.
- Sanayi: Kalite kontrol, taşıma.

2016 yılında Kahveci ve Özdemir öğrenci bilgi sistemleri üzerinde veri madenciliği ile bir çalışma yapmıştır. Yapılan çalışma ile önceki yıllarda yapılmış olan sınavlarda sınıfın not ortalaması, notların toplandığı noktalar vb. seçeneklere dikkat edilerek not aralık yöntemlerinden en yakın olan seçilerek önceki çalışmalardan faydalanılmış olmaktadır. Yeni sınıflar içinde yakın bir başarı notu hesaplaması yapılması hedeflenmektedir. Bağlı değerlendirmede kullanılan uygulamalardan biri olan not aralıklandırma sistemi kullanan üniversitelerde geçmiş verilerden yararlanarak sistemin bir öneride bulunması kullanıcının "Acaba daha önce buna benzer bir sınıfta nasıl bir aralık kullanılmıştı?" sorusuna yanıt verebilmektedir. Bu şekilde hem insan eli hem de bilgisayar yardımının bir arada bulunduğu bir yardımcı sistem tasarımı ortaya çıkmaktadır. Geçmiş aralık verilerinde kullanıcıların düşünme sürecinden geçen bir emeği bulunmaktadır. Tasarlanan sistemde kullanıcıların bu deneyiminden yararlanarak yine kendilerine bir öneride bulunmaktadır. Bu şekilde geleneksel olarak kullanılan bir girdi verip işlemde geçip bir sonuç alma yerine kullanıcı deneyimi olan verileri girdi olarak verip yine çıktısının kullanımına kullanıcının karar verdiği bir sistem olarak tasarlanmaktadır. Eğer ki kullanıcı sistemin önerisini kabul etmeyip kararında farklı faktörler söz konusu ise sistem kullanıcının bu farklı kararını geçmiş deneyime kaydedip tercihi hafızaya almaktadır. Bu çalışmada kullanıcının verdiği kararlar ile uyumlu çalışan bir sistemin eğitim alanında veri madenciliği tekniklerini kullanımı görülmektedir (Kahveci ve Özdemir, 2016).

Çeşmeli ve arkadaşları öğrenci başarılarını veri madenciliği yöntemi ile incelemiştir. Yapılan çalışmada; veri madenciliği teknikleri kullanılarak Yönetim Bilişim Sistemleri bölümünde kayıtlı öğrencilerin, Yönetim Bilişim dersindeki başarı durumları araştırılmıştır. Veriler 11 öznitelikten meydana gelmektedir. Bu öznitelikler 3. ve 4. sınıf öğrencilerinin cinsiyet, yaş, öğrenim türü, hazırlık sınıfı okuma, üniversite öncesi ikametgâh yeri, ceza durumu gibi maddelerden oluşmaktadır. Çalışmada Yapay Sinir Ağları ve Uyarlamalı Sinir-Bulanık Mantık Sınıflayıcısı kullanılmıştır. YSA'nın eğitilmesi sonucunda eğitim setinde %90.4, test setinde ise %83.5 başarı elde edilirken, USPS ile yapılan sınıflandırmada ise eğitim verisi için %86.6, test verisi için %86.3 başarı elde edilmiştir. Dersin kategorisi ve öğrencinin hazırlık sınıf okumasının öğrencinin başarısını etkileyen en önemli unsurlar olarak ortaya çıkmıştır (Çeşmeli vd., 2015).

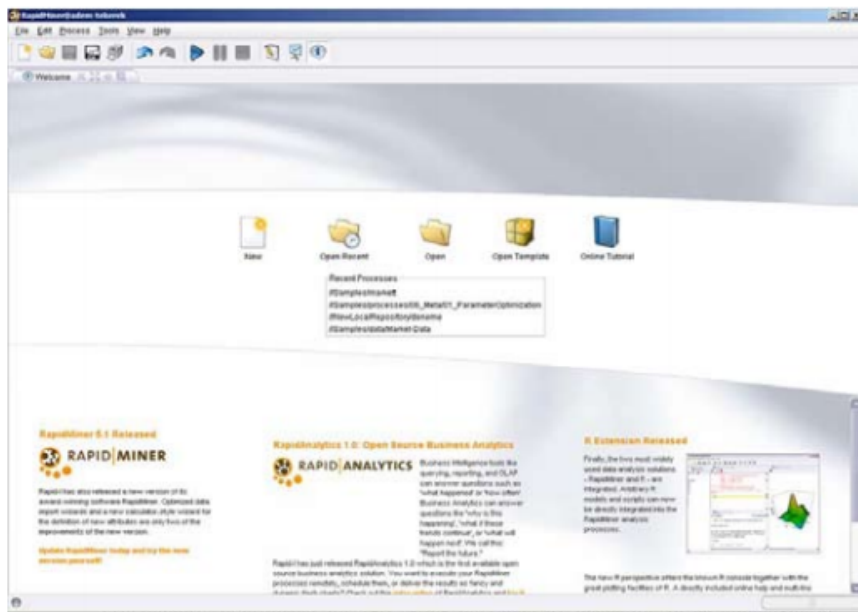
Alan (2012) yaptıęı alıřmada veri madencilięi yntemi ile lisansst ęrencilere ynelik bir alıřma yapmıřtır. Yapılan alıřmada; Cumhuriyet niversitesi Sosyal Bilimler Enstits ęrencilerinin verileri kullanılarak veri madencilięi ile modelleme yapılmıřtır. ęrencilerin Sivas ilinden veya farklı bir řehirden olması, arařtırma grevlisi kadrosunda alıřıp alıřmadıęı ve ders dnemlerindeki farklılıkların ęrencilerin başarısına etkisi arařtırılmıřtır. Ayrıca bu verileri en başarılı olarak sınıflandıran algoritmanın da bulunması amalanmıřtır. Yapılan denemeler sonucunda SimpleCART algoritması sınıflandırma iřleminde başarılı bulunmuřtur. Yksek lisans yapan ęrencilerin, doktora ęrencilerine gre başarısız olduęu ancak cinsiyetin, kadronun, ilin ve ders dneminin başarıyı etkilemedięi sonucuna varılmıřtır (Alan, 2012).

Irmak ve arkadařları hastane hazır bir hastane veri tabanını kullanarak veri madencilięini kullanmıřlardır. Bu alıřmada řu anda aktif olarak alıřan bir saęlık kurumu veri tabanında veri madencilięi yntemleri uygulanmıřtır. Uygulanan modelde hastanede ki hasta yoęunluęu tahmin edilmeye alıřılmıř ve ıkan sonular dięerleri ile kıyaslanmıřtır. stel dzgnleřtirme yntemlerinden Winters Additive modeli, ARIMA yntemlerinden ARIMA(3,1,0)(1,0,0)₁₂ modeli ve yapay sinir aęları yntemlerinden Prune yntemi ile elde edilen model en iyi sonuları vermiřtir. Winters Additive stel dzgnleřtirme modeli ise bunlar arasında en kestirimci model olarak ne ıkmıř ve gerekleřen deęerlere gre en başarılı tahminlere ulařtıęı gzlemlenmiřtir.

5. Veri madencilięi programları

1. RapidMiner (YALE) : Java dili kullanılarak Yale niversitesinde alıřan bilim adamlarınca ortaya ıkarılmıřtır. Elimizde bulunan byk miktardaki veriden anlamlı sonular ıkarmak iin RapidMiner kullanılabilir. Obf, bar, per, res, sim, thr, wgt, wls, xrff, aml, arff, att, bib, cfm, cms, cri, csv, dat, ioc, log, mat, mod, uzantılı dosyaları desteklemektedir(İnternet). Veri madencilięi uygulamaları iin 400'den fazla operatr nermektedir. Deneysel tasarımıların otomatik optimizasyonu iin Meta operatrleri kullanılır. Bu iřleminde kullanıcıdan herhangi bir parametre ayarlaması yapařması istenmez.

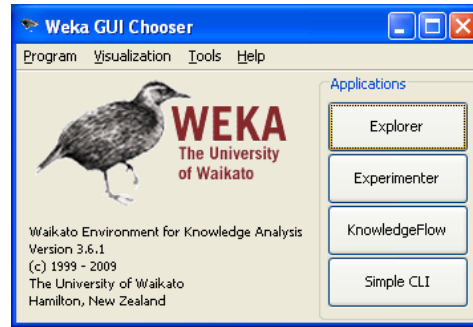
řekil 5. RapidMiner Aılıř Ekranı



Albayrak, M. (2017). The use of data mining in scientific research. *International Journal of Social Sciences and Education Research*, 3(2), 751-760.

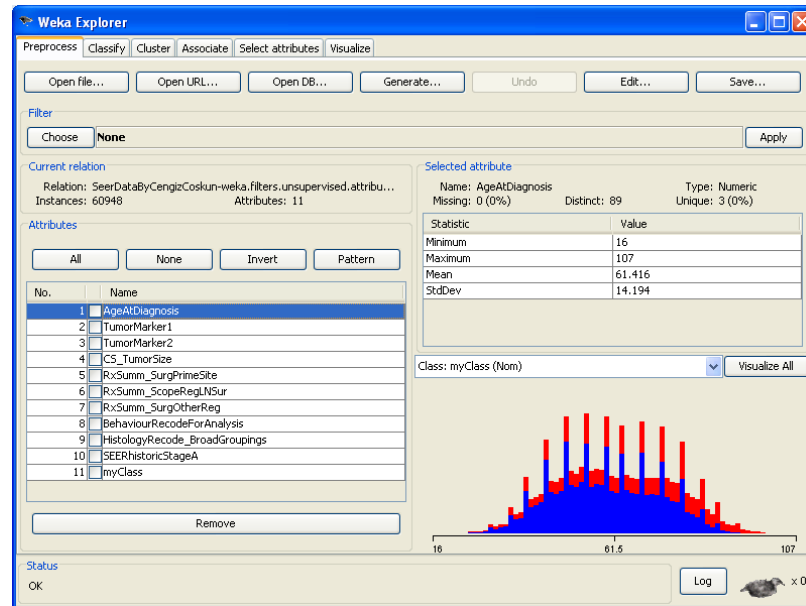
2.WEKA: Weka ücretsiz, açık kaynak kodlu, kullanıcı dostu ara yüzü olan, http protokolü kullanan dosyalara, Metin tabanlı dosyalarına, jdbc sürücüsü bulunan veri tabanlarına ulaşma izni veren yazılımdır. Weka menü ekranı Şekil-6’da gösterilmektedir.

Şekil 6. Weka menü ekranı



WEKA ile rastgele bir text dosyasında bulunan veriler işlenemez. Arff, Csv, C4.5 formatındaki dosyaların işlenmesi mümkündür. Jdbc kullanılarak veritabanı üzerinde işlem yapılabilir. Veri İşleme, Veri Sınıflandırma, Veri Kümeleme, Veri İlişkilendirme gibi özellikleri üzerinde barındırır. Weka çalışma ekranından bir görüntü de Şekil-7 de gösterilmektedir.

Şekil 7. Weka çalışma ekranı



Bu programlar haricinde veri madenciliği işlemlerinde kullanılan Orange, Konstanz Information Miner (KNIME), R, Tanagra gibi ücretsiz programlarda mevcuttur.

6. Sonuç

Ülkemizde de son dönemlerde veri madenciliği uygulamalarına olan ilgi artmaktadır. Veri madenciliğinin kullanım alanları yaygınlaşmaktadır. Bu çalışma da Veri madenciliğinin kullanım alanları, bilimsel alanda veri madenciliği çalışmaları ve veri madenciliği uygulama programları anlatılmıştır.

Veri madencilięi doęru, gncel bilgiye eriřmede, mantıklı czmlerin kullanılmasında karar destek aracı olarak grev yapmaktadır. Gelecekte sayısal karar verme aracı olarak kullanılacak ve profesyonelleri hizmetlerin daha etkin, kaynakların daha verimli bir řekilde kullanabilmesine imkn saęlayacaktır.

Gnmzde nemli bir konu haline gelen byk verinin kullanılmasında da veri madencilięi uygulamalarının nemi her geen gn artmaktadır. Bunun sonucu olarak, bu alana yapılan yatırımlarda bymektedir. Sistemler, anketler veya elde edilen bilgiler kullanılarak veri madencilięi modellerinin oluřturduęu ngrler ile řirketler, kurumlar vb. yapılar uygulama politikalarına yn vermektedir.

Kaynaka

- Alan, A. (2012). Veri Madencilięi Ve Lisansst Öğrenci Verileri zerine Bir Uygulama. *Dumlupınar niversitesi Sosyal Bilimler Dergisi*, Sayı 33, Aęustos 2012.
- Baykasoęlu, A. (2005). Veri Madencilięi Ve Cimento Sektrnde Bir Uygulama. *Akademik Biliřim '05-VII. Akademik Biliřim Konferansı*, 2005.
- eřmeli, M.ř., Bozkurt, . C., Kalkan, A., Pene, İ, (2015). Ynetim Biliřim Sistemleri Blm Öğrencilerinin Ynetim Ve Biliřim Derslerindeki Bařarılarının Veri Madencilięi Yntemleri İle İncelenmesi. *Ynetim Biliřim Sistemleri Dergisi*. Cilt: 1 Sayı:2, Yıl:2015, Sayfa:36-47. ISSN: 2148-3752.
- Cořkun, C., , Baykal, A., (2011). Veri Madencilięinde Sınıflandırma Algoritmalarının Bir rnek zerinde Karřılařtırılması. *Akademik Biliřim'11 - XIII. Akademik Biliřim Konferansı*. 2-4 řubat 2011
- İrmak, S., Kksal, D. C., Asilkan, . (2012). Hastanelerin Gelecekteki Hasta Yoęunluklarının Veri Madencilięi Yntemleri İle Tahmin Edilmesi. *Uluslararası Alanya İřletme Fakltesi Dergisi*. Yıl:2012, C:4, S:1, s.101-114
- Kahveci, F., zdemir, A. (2016). Öğrenci Bilgi Sisteminde Deęerlendirmenin Veri Madencilięi İle Yapılması. *Ynetim Biliřim Sistemleri Dergisi*. Cilt:1, Sayı:3, Yıl:2016, Sayfa 1 -10. ISSN: 2148-3752.
- Kalıkov, A., (2006). *Veri Madencilięi ve Bir E-Ticaret Uygulaması*, Yksek Lisans Tezi, Gazi niversitesi, Fen Bilimleri Enstits.
- Savař, S., , Topaloęlu, N.,, Yılmaz, M. (2012). Veri Madencilięi Ve Trkiye'deki Uygulama rnekleri. *İstanbul Ticaret niversitesi Fen Bilimleri Dergisi*. Yıl:11 Sayı: 21 Bahar 2012 s. 1-23
- Shearer, C., (2000). The Crisp-DM Model: The New Blueprint for Data Mining. *Journal of Data Warehousing*, Cilt 5 No 4, 13-23.
- Tang, Z., MacLennan, J. (2005). *Data Mining with Sql Server 2005*, USA:Wiley.
- Thuarisingham, B.M. (2003). *Web Data Mining and Applications in Business Intelligence and Counter Terrorism*, CRC Press LLC, Boca Raton, FL, USA.
- RAPİDMİNER (2010). ----- ", <http://www.aktueryabilimleri.com/index.php?option=comcontent&view=category&id=97:rapidminer&Itemid=252&layout=default>

Extended abstract in English

Information Systems is the whole of method, application, methodology and algorithms that translate desired and useful knowledge of personal users and groups in institutions or organizations. Computer Science, Computer Engineering, Software Engineering, Mathematics, Electronics, Physics, Management Science, Language and Psychology. Information is not only an engineering field, it has entered into the whole of daily life and scientific field. The greatest benefit of Information Systems is that it provides low cost and quality workflow in all areas where it is used.

In this study, we have discussed the large data that have become increasingly important in recent years, the methods used for data processing in databases and data warehouses, invisible data relationships and methods used for exploring them from the database. The data mining, which is structured to examine the data in large quantities, has a close relationship with the databases. Data mining is a tool that shows the way to be tracked to achieve a resolution. Data mining is used by many disciplines for different purposes. We have tried to summarize the data mining studies conducted in different branches of science in our country for research and analysis purposes and to refer to different applications. The software used in the studies and the differences between them are mentioned. In addition, the classification and clustering algorithms frequently used in data mining can be used on which kind of data, and the importance of interpretation and understanding of the results is emphasized.

Data mining has taken its place in recent years by analyzing data and extracting usefulness and logical information from the source without learning statistics, artificial intelligence and machinery. There was intense interest in the academic field, finance and business circles (Coskun and Baykal, 2011). Data mining is the acquisition of information that is not very clear, previously unknown, but will eventually work, from the data it contains. This includes different approaches such as clustering, data summarization, analysis of changes, detection of deviations. In another sense; Data mining is to reveal the hidden links and tendencies between the data (Thuarisingham, 2003). The use of data mining has been increasing in recent years. It is frequently used by enterprises, institutions and organizations to provide effective results and easy use. Examples of areas where data mining is used are given below.

- Finance Sector: Risk analysis and irregularity detection.
- Marketing: Cross-sell analysis, customer segmentation.
- Insurance: Identification of causes of customer decline, prevention of wrong transactions.
- Communication: Misleading detection, density predictions in systems.
- Stock Exchange: Estimation of movements in stock price, general market assessments.
- Health: Disease diagnosis and treatment recommendations
- Science and Engineering: solving scientific and technical problems by establishing models on empirical data.
- Industry: Quality control, transportation.

Data mining can be defined as a sequence of events and movements that develop. The determining factor in this process is the practitioner who creates the function. When the steps that make up the operation are not done properly, the targeted output cannot be reached. In order to

obtain the highest level of efficiency from the data mining algorithms, it is necessary to dominate the features of the subject and the data being worked on (Savař et al., 2012).

The steps followed by data mining are shown below (Shearer, 2000):

1. Identification of the problem: It is necessary to determine well for what purpose the process to be performed is to be used or to be used for the solution of the problem.

2. Preparation of data: The data are stored in multiple media. For example; It stores Microsoft data in data warehouses and the OLTP database. The important thing in this section is to get the appropriate data from the database or from the data warehouses. After the data collection is complete, 80% of the data are divided into two parts: training (20%) and test data (Tang and Maclenan, 2005).

3. Establishing and evaluating the model: The right choice of model is very important in data mining. It can be tried on many models until the correct model is found. When the probing is most appropriate, the modeling process is completed.

Using model

4. The most appropriate model according to the probing model can be a lower part of a program or program.

5. Model monitoring: The model results can be monitored over time and updated if necessary.

Different methods are used in data mining. The methods used can be different according to the problem and the application area. Many of the methods used are the domain of artificial intelligence. In Turkey, interest in applying data mining is increasing recently. The areas of use of data mining are widespread. This study describes the uses of data mining, scientific field data mining studies and data mining application programs.

Data mining is working as a decision support tool for accurate, up-to-date information access and use of logical solutions. It will be used as a digital decision tool in the future, and professionals will be able to use services more effectively and resources more efficiently.

The importance of data mining applications is increasing day by day in the use of large data which has become an important topic today. As a result, it is growing in investments made in this area. Systems, surveys or data mining models using the information provided by the predictions created by companies, institutions and so on. Structures lead to implementation policies.