Açık Kaynak Kodlu Veri Madenciliği Yazılımlarının Karşılaştırılması

Mümine KAYA[1], Selma Ayşe ÖZEL [2]

[1] Adana Bilim ve Teknoloji Üniversitesi, Bilgisayar Mühendisliği Bölümü, Adana

[2] Çukurova Üniversitesi, Bilgisayar Mühendisliği Bölümü, Adana

mkaya@adanabtu.edu.tr, saozel@cu.edu.tr

Özet: Veri Madenciliği, büyük miktarda veri içinden gizli bağıntı ve kuralların, bilgisayar yazılımları ve istatiksel yöntemler kullanılarak çıkarılması işlemidir. Veri madenciliği yöntemleri ve yazılımlarının amacı büyük miktarlardaki verileri etkin ve verimli bir şekilde işlemektir. Yapılan çalışmada; açık kaynak kodlu veri madenciliği yazılımlarından Keel, Knime, Orange, R, RapidMiner (Yale) ve Weka karşılaştırılmıştır. Böylece kullanılacak veri kümeleri için hangi yazılımın daha etkin bir şekilde çalışacağı belirlenebilmiştir.

Anahtar Sözcükler: Veri Madenciliği, Açık Kaynak, Veri Madenciliği Yazılımları.

Comparison of Open Source Data Mining Software

Abstract: Data Mining is a process of discovering hidden correlations and rules within large amounts of data using computer software and statistical methods. The aim of data mining methods and software is to process large amounts of data efficiently and effectively. In this study, open source data mining tools namely Keel, Knime, Orange, R, RapidMiner (Yale), and Weka were compared. As a result of this study, it is possible to determine which data mining software is more efficient and effective for which kind of data sets.

Keywords: Data Mining, Open Source, Data Mining Software.

1. Giriş

Günümüzde bilişim teknolojisi, veri iletişim teknolojileri ve veri toplama araçları oldukça gelişmiş ve yaygınlaşmış; bu hızlı gelişim büyük boyutlu veri kaynaklarının oluşmasına neden olmuş ve beraberinde bazı problemlere yol açmıştır [1]. Bu problemlerin başında, veritabanları içinde yer alan ancak basit SQL sorguları ile bulunamayan anlamlı ve yararlı bilginin ortaya çıkarılması gelmektedir. Bu nedenle verileri işlemek için bazı çözümleme yöntemlerine ihtiyaç duyulmuştur. Veri Madenciliği bu ihtiyacı gidermek için ortaya çıkarılmış bir yöntemdir. Veri Madenciliği daha önceden bilinmeyen, geçerli ve uygulanabilir bilgilerin geniş veri kaynaklarından elde edilmesi işlemidir [2]. Daha da özetlemek gerekirse, veri madenciliği büyük ölçekli veriler arasından yararlı ve anlaşılır olanların bulunup ortaya çıkarılması işlemidir [1]. Veri Madenciliği ile veriler arasındaki ilişkiler ortaya koyulabilmekte ve gelecekle ilgili tahminlerde bulunulabilmektedir. Veri Madenciliğinin geleneksel veritabanı sorgularından farkı şu şekilde özetlenebilir: *i)* Geleneksel veri tabanlarında sorgu, SQL gibi iyi tanımlanmış bir sorgulama dili ile yapılırken, veri madenciliğinde ise sorgu iyi tanımlı ya da tam tanımlı olmayabilir; *ii)* Geleneksel veri tabanlarında sorgunun sonucu, veri tabanında yer alan verilerin bir alt kümesi olup, veri madenciliğinde ise çoğunlukla veri tabanının bir alt kümesi olmaz, onun yerine veri tabanındaki içeriğin bir analizi olur.

Veri Madenciliğinin amacı ham veriyi anlamlı, etkin ve yararlı olan bilgiye dönüştürebilmektir [3]. Bu amaca ulaşabilmek için de Veri Madenciliği konusunda geliştirilmiş yazılımların kullanılması veri madenciliği süreçlerini kolaylaştırmaktadır.

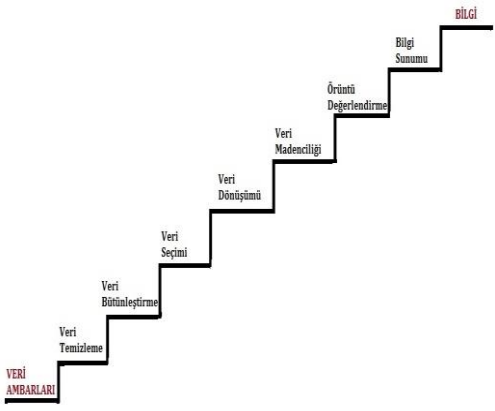
Bugüne kadar yapılan çalışmalarda; veri madenciliği yazılımlarının bir kısmının detaya girilmeden, ya tanımlamaları ya da uygulamaları yapılmıştır [4, 5 ve 6]. Bu çalışmada ise diğer çalışmalardan farklı olarak, altı adet veri madenciliği yazılımı daha detaylı karşılaştırılmıştır. Böylece ihtiyaca göre daha etkin bir şekilde kullanılacak olan yazılımlar belirlenmiştir.

Çalışmanın ikinci bölümünde veri madenciliği süreci hakkında bilgi verilmiştir. Üçüncü bölümde çalışmada kullanılan açık kaynak kodlu yazılımlar tanıtılmış olup, dördüncü bölümde bu yazılımların farklı açılardan karşılaştırılmaları yapılmıştır. Son bölümde ise sonuç ve önerilere yer verilmiştir.

2. Veri Madenciliği Süreci

Veri madenciliği bilgi keşfi işleminin en önemli adımlarındandır. Bilgi keşfi adımları: Veri Temizleme, Veri Bütünleştirme, Veri Seçme, Veri Dönüşümü, Veri Madenciliği, Örüntü Değerlendirme ve Bilgi Sunumu olmak üzere 7 basamaktan oluşmaktadır [7].

Şekil 1'de de görüldüğü üzere bu süreç, ele alınan problemin tanımlanması ile başlamakta ve sırasıyla; problemle ilgili verilerin toplanması, verilerin hazırlanması, verilere ve probleme uygun modelin tasarlanması, tasarımı yapılan modelin uygunluğunun ve yeterliliğinin değerlendirilmesi ile devam etmekte ve son olarak modelin uygulanmasıyla sonuca ulaştırılmaktadır. Bu sonuca ulaşırken de veri temizleme adımında gürültülü ve tutarsız veriler veri kümesinden çıkarılmakta; veri bütünleştirme adımında birçok veri kaynağından gelen farklı formatlardaki veri birleştirilebilmekte; veri seçme adımında yapılacak olan analiz ile ilgili olan veriler belirlenmekte; veri dönüşümü adımında verinin veri madenciliği tekniğinde kullanılabilecek hale dönüşümü gerçekleştirilmekte; veri madenciliği adımında veri örüntülerini yakalayabilmek için akıllı metotlar uygulanmakta; örüntü değerlendirme adımında bazı ölçütlere göre elde edilmiş bilgiyi temsil eden ilginç örüntüler tanımlanmakta ve bilgi sunumu adımında ise elde edilmiş bilginin kullanıcıya sunumu gerçekleştirilmektedir [7, 8].



Şekil 1. Bilgi Keşfi Süreci

2.1 Veri Madenciliğinin Kullanım Alanları

Veri Madenciliği; bankacılık, borsa, pazarlama yönetimi, perakende satış, işaret işleme, sigortacılık, telekomünikasyon, elektronik ticaret, sağlık, tıp, biyoloji, genetik, endüstri, eğitim, istihbarat, bilim ve mühendislik gibi birçok dalda uygulama alanı bulunmaktadır [1, 2].

2.2 Veri Madenciliği Yazılımları

Veri Madenciliği konusunda geliştirilmiş birçok yazılım bulunmaktadır. Bu yazılımlardan kimisi ticari iken, kimisi açık kaynak kodludur. Bu nedenle veri madenciliği yazılımları ticari ve açık kaynak kodlu olmak üzere iki gruba ayrılmaktadır. Ticari yazılımlara SPSS Clementine, Excel, SPSS, SAS, Angoss, KXEN, MS SQL

Server, MATLAB ve Oracle'ın bu amaçla geliştirdiği modülleri örnek olarak verilebilmektedir. Açık kaynak yazılımlara ise Orange, RapidMiner, WEKA, R, Keel, Knime, Tanagra, Scriptella ETL, jHepWork ve Elki örnek olarak verilebilmektedir [4, 5].

3. Açık Kaynak Kodlu Veri Madenciliği Yazılımları

Bu bölümde açık kaynak kodlu ve ücretsiz olarak kullanılabilen Keel, Knime, Orange, R, RapidMiner (Yale) ve Weka yazılımları hakkında özet bilgi verilmektedir.

3.1 Keel

Keel [9], İspanya Ulusal Bilim Projeleri Kurumunun desteği ile Granada Üniversitesi tarafından geliştirilen Java dilinde yazılmış bir yazılımdır. Keel, kümeleme ve sınıflandırma gibi klasik veri madenciliği algoritmaları açısından zengin değildir. Bunların yerine Fuzzy sınıflandırıcılar, Yapay zekâ tabanlı sınıflandırma ve Kural tabanlı kümeleme algoritmalarının birçok çeşidini içermektedir [6]. Veri görselleştirme açısından en zayıf yazılımlardan biri Keel'dir.

3.2 Knime

Konstanz Information Miner (KNIME) [10], Konstanz Üniversitesi görsel veri madenciliği araştırma grubu tarafından Eclipse Rich Client Platform üzerinde geliştirilen bir yazılımdır. Knime genişletilebilme özellikleri ile ön plana çıkmaktadır. Kullanıcılara bir yazılım geliştirme kiti sunarak kullanıcıların kendi modüllerini yazabilmelerini sağlayan tek uygulamadır [6]. Kurulum şartı olmadan çalışabilmektedir. Knime yazılımı .txt uzantılı metin dosyalarından veya .arff, .table formatından veri alabilmektedir. Knime, en zengin görselleştirme araçları sunan yazılımlarından biridir.

3.3 Orange

Orange [11], Slovenya Ljubljana Üniversitesi Bilgisayar ve Enformatik Bilimleri bölümü yapay zekâ araştırmaları ekibi tarafından geliştirilmiş bir yazılımdır [6]. Orange yazılımı C++ dili ile geliştirilmiştir. Yazılımın ara yüzleri ve grafik ortamı ise Qt3 kütüphanesi ve Python kullanılarak geliştirilmiştir [6]. Görselleştirme açısından zayıf bir yazılımdır. Yalnızca metin dosyalarından veri alma işlemini gerçekleştirmektedir.

3.4 R

Auckland Üniversitesi İstatistik Bölümü bilim adamlarından olan Robert Gentleman ve Ross Ihaka tarafından grafikler, istatistiksel hesaplamalar ve veri analizleri için geliştirilmiş bir yazılımdır [4]. R [12], Unix makinelerde yaygın olarak kullanılmaktadır. R, bir veri madenciliği yazılımı olarak çok fazla tercih edilmemektedir.

3.5 RapidMiner (Yale)

RapidMiner [13], Ralf Klinkenberg, Ingo Mierswa ve Simon Fischer tarafından Dortmund Teknoloji Üniversitesi Yapay Zeka Biriminde geliştirilmiş bir yazılımdır. Yale [14] ise Yale üniversitesi bilim adamları tarafından Java dili kullanılarak geliştirilmiş bir yazılımdır. Yale artık RapidMiner [13] adı altında yazılım olarak kullanılmaya devam etmektedir. Diğer veri madenciliği yazılımlarından farklı olarak 22 adet dosya formatındaki veriyi işleyebilmektedir. Veri Madenciliği ve Makine Öğrenme Algoritmalarını da kapsayan RapidMiner, Weka gibi oldukça fazla algoritmaya sahiptir. Veri Analizi, Önişleme, Sınıflama, Kümeleme, Birliktelik Kuralları Çıkarımı, Nitelik Seçimi işlemlerini içermektedir. Oracle, MS SQL Server, PostgreSQL, MySQL, JDBC, Sybase, Access, IBM DB2, Ingres veritabanlarını ve metin dosyalarını desteklemektedir [6]. Bu açıdan en kapsamlı yazılımlardan biridir. Excel dosyalarıyla bağlantı kurulabilmektedir. MS Windows, GNU/Linux, Mac Os X işletim sistemlerinde kolayca ve hatasız olarak çalışabilmektedir. Görselleştirme ve grafik arayüzü açısından da

en zengin yazılımlardan biridir. İçerisinden script yazılabilir.

3.6 Weka

Weka [15], Waikato Environment for Knowledge Analysis kelimelerinin kısaltılmasıdır. Waikato Üniversitesinde, Java platformu üzerinde geliştirilmiş ve GNU genel kamu lisansı altında bildirilmiş açık kodlu bir veri madenciliği yazılımıdır. Java Database Connectivity (JDBC) kullanarak SQL veri tabanlarına erişim sağlar [16]. Makine öğrenmesi algoritmalarını içermektedir. İçerdiği özelliklerle veri kümeleri üzerinde önişleme, sınıflandırma, kümeleme, birliktelik kuralı madenciliği, özellik seçimi ve görselleştirme yapabilmektedir. Weka'ya özel olarak tasarlanmış, metin yapısında tutulan .arff (Attribute Relationship File Format) dosya formatı üzerinde çalışmaktadır.

4. Açık Kaynak Kodlu Veri Madenciliği Yazılımlarının Karşılaştırılması

Bu çalışmada Keel, Knime, Orange, R, RapidMiner ve Weka yazılımları farklı açılardan karşılaştırılmıştır. Böylece kullanılacak veri kümesiyle ulaşılmak istenen hedef arasında en etkin sonucu sağlamaya yardımcı olacak yazılımlar Tablo 1'e göre belirlenmeye çalışılmıştır [17, 18, 19 ve 20]. İçerdiği Veri Madenciliği Algoritmaları açısından en kapsamlı yazılımlar Tablo 1'de de görüldüğü üzere RapidMiner ve Weka'dır. En az algoritmaya sahip olan yazılım ise R'dır. Makine Öğrenmesi paketleri açısından ise en güçlü yazılım Weka'dır. Metin Madenciliği işlemlerini Keel, Orange, RapidMiner ve Weka kendi başlarına yapabilirlerken; Knime bir modül sayesinde, R ise paket aracılığıyla yapabilmektedirler. Biyoinformatik işlemlerini ise; Keel, R ve Weka kendi başlarına yapabilirlerken; Knime ve RapidMiner modül aracılığıyla, Orange ise paket aracılığıyla yapabilmektedirler. İstatiksel hesaplama işlemlerini karşılaştırılan yazılımların hepsi yapabilmektedir. En güçlü olan yazılım ise bir istatistik yazılımı olan R'dır. Orange, RapidMiner ve Weka R'a göre istatiksel hesaplamada daha zayıf yazılımlar arasındadır. Bunda R'ın kendi istatiksel kütüphanesinin olmasının payı çok büyüktür. Veri Analizi, Sınıflama, Kümeleme, Nitelik Seçimi işlemlerinin hepsini bütün yazılımlar gerçekleştirebilmektedirler. Birliktelik Kuralları Çıkarımı işlemini de yazılımların hepsi yapabilmektedir, ancak sadece R paketler ile birlikte yapabilmektedir. Görselleştirme açısından en iyi yazılımlar Knime, R ve RapidMiner'dır. Ancak çok iyi görselleştirme sunmasının en büyük dezavantajı karmaşıklıklarını arttırmasıdır. Komut Satırı Arayüzü ile bağlantıda en iyi yazılım Weka iken, en zayıf yazılım R'dır. Kullanım ve Öğrenim kolaylığı açısından da kompleks bir yapıya sahip olmamasından dolayı en başarılı yazılım olarak Weka bulunmuştur. En fazla dosya formatı destekleyen yazılım ise RapidMiner'dır. 22 adet dosya formatını desteklemektedir. Phyton diliyle yazılmasından dolayı yazılım içerisinde script yazmada en başarılı yazılım Orange olarak tespit edilmiştir. Veri Alma/Verme işlemlerini kolayca gerçekleştirmede en başarılı yazılım ise R'dır. Çeşitli veri tabanlarıyla çalışabilmede Knime, R ve RapidMiner en başarılı yazılımlar olarak tespit edilmiştir. Excel dosyalarıyla çalışabilmede en kötü bağlantıyı Weka yazılımı yapmaktadır. Keel ise import işlemi ile gerçekleştirebilmektedir. Knime ve Orange ise hiçbir şekilde çalışamamaktadır. R ve RapidMiner bu alanda en başarılı olanlarıdır. Karşılaştırılan yazılımlar arasında kurulum şartı olmadan çalışabilen tek yazılım Keel'dir. Diğer yazılımların öncelikle bilgisayara kurulması gerekmektedir. Kurulabilecekleri işletim sistemleri Tablo 1'de gösterilmektedir. Bellek açısından bu altı yazılım incelendiğinde Keel, Orange ve R'ın limitli imkan sundukları gözlemlenmiştir. Knime'ın kullandığı bellek boyutu ayarlanabilirken, RapidMiner'da arttırma işlemi yapılabilmektedir. Weka'da da bellek boyutunun ayarlanabilme / arttırılabilme özelliği mevcuttur.

Tablo 1. Açık Kaynak Kodlu Veri Madenciliği Yazılımlarının Karşılaştırılması

	Keel	**Knime**	**Orange**	**R**	**RapidMiner (YALE)**	**WEKA**
Veri Madenciliği Algoritmaları	Var	Var	Var	Var (En Az)	Var (En Fazla)	Var (En Fazla)
Makine Öğrenmesi Paketleri	Var	Var (Güçlü)	Var (Zayıf)	Var (Zayıf)	Var	Var (Çok Güçlü)
Metin Madenciliği	Var	Var (Modül)	Var	Var (Paket)	Var	Var
Biyoinformatik	Var	Var (Modül)	Var (Paket)	Var	Var (Modül)	Var
İstatiksel Hesaplama	Var	Var	Var (Zayıf)	Var (Çok Güçlü)	Var (Zayıf)	Var (Zayıf)
Veri Analizi	Var	Var	Var	Var	Var	Var
Önişleme	Var	Var	Var	Var	Var	Var
Sınıflama	Var	Var	Var	Var	Var	Var
Kümeleme	Var	Var	Var	Var	Var	Var
Birliktelik Kuralları Çıkarımı	Var	Var	Var	Var (Paket Olarak)	Var	Var
Nitelik Seçimi	Var	Var	Var	Var	Var	Var
Görselleştirme	Var	Var (Çok İyi)	Var (6 çeşit)	Var (Çok İyi)	Var (Çok İyi)	Var (5 çeşit)
GUI	İyi	Çok İyi	İyi	Zayıf	Çok İyi	İyi
Grafik	Var	Var	Var	Var	Var	Var
Komut Satırı Arayüzü Bağlantısı	Var	Var	Var	Var (Çok Zayıf)	Var	Var (Çok İyi)
GPL Lisans	Evet (GPLv3)	Evet (GPLv3)	Evet	Evet (GPLv3)	Evet	Evet
Genişletilebilirlik	Evet	Evet	Evet	Evet	Evet	Evet
Esneklik	Evet	Evet	Evet	Evet	Evet	Evet (Az)
Kullanım Kolaylığı	Evet	Evet	Evet	Evet	Evet	Evet (En Kolay)
Öğrenim Kolaylığı	Evet	Evet	Evet	Evet	Evet	Evet (En Kolay)
Hatasız Çalışma	Evet	Evet	Evet (En Az)	Evet	Evet (En Fazla)	Evet (En Az)
Dokümantasyon	Var	Var	Var	Var	Var	Var
Script Yazma	Var	Var	Var (En İyi)	Var	Var	Var
Eklenebilir Paketler	Evet	Evet	Evet	Evet	Evet	Evet

Tablo 1. Açık Kaynak Kodlu Veri Madenciliği Yazılımlarının Karşılaştırılması (devamı)

	Keel	**Knime**	**Orange**	**R**	**RapidMiner (YALE)**	**WEKA**
Veri Alma/Verme	Var	Var	Var	Var (Çok Kolay)	Var	Var
Desteklenen Dosya Formatları	.dat, .arff, .csv, .xml, .txt, .prn, .xls, .dif, .html	.arff, .csv	.tab, .basket, .names, .data, .txt, .xls (.arff ve .csv sadece okuyabiliyor)	.r, .txt, .ods, .csv, .xml	.sml, .srff, .stt, .bib, .clm, .cms, .cri, .csv, .dat, .ioc, .log, .matte, .mode, .obf, a bar, one pair, .res, .sim, .thr, .wgt, .wls, .xrff, .arff	.arff, .csv
Veritabanlarıyla Çalışabilme	Var (SQL Verita-banları)	Var (Oracle, MS SQL Server, PostgreSQL, MySQL, Access, ODBC, JDBC)	Var (MySQL)	Var (Informix, Oracle, Sybase, DB2, MS SQL Server, MySQL, PostgreSQL, MS Access, ODBC)	Var (Oracle, MS SQL Server, PostgreSQL, MySQL, JDBC, Sybase, Access, IBM DB2, Ingres, Metin Dosyaları)	Var (JDBC, JDBC aracılığıy-la SQL Verita-banları)
Excel Dosyalarıyla Çalışabilme	Evet (import ile)	Hayır	Hayır	Evet	Evet	Evet (Kötü Bağlantı)
Bellek Kullanımı	Limitli	Ayarlana-bilir	Limitli	Limitli	Arttrılabilir	Arttrılabilir/ Ayarlanabilir
Yazıldığı Dil	Java	Java	Phyton, C++	C, R, C++, Fortran	Java	Java
Kurulum Şartı	Yok	Var	Var	Var	Var	Var
Gerekli Minimum İşletim Sistemi	MS Windows, GNU/ Linux, Mac Os X	MS Windows, GNU/ Linux, Mac Os X	MS Windows, GNU/ Linux, Mac Os X	MS Windows, GNU/Linux, Unix, Mac Os X	MS Windows, GNU/Linux, Mac Os X	MS Windows, GNU/ Linux, Mac Os X

Sınıflandırma Algoritmaları açısından hemen hemen tüm yazılımlar birçok sınıflandırma algoritmasını içerisinde barındırmaktadır. KNN algoritması her yazılımda bulunurken sadece R'da RWeka paketinde bulunmaktadır. Aynı şekilde Lazy sınıflandırıcılar da Knime ve R dışındaki tüm yazılımların içerisinde mevcutken; Knime'de Weka içerisinden, R'da ise RWeka içerisinden çalıştırılabilmektedir.

Kümeleme Algoritmaları açısından yazılımlar karşılaştırıldığında; en popüler kümeleme algoritması olan K-Means Algoritması karşılaştırdığımız bütün yazılımlarda bulunmaktadır. Hiyerarşik Kümeleme

algoritmaları ise Knime, Orange, R ve Weka'da bulunurken, Keel'de bulunmamaktadır. RapidMiner'da ise modül olarak bulunmaktadır.

Birliktelik Kuralları açısından karşılaştırma yapıldığında; en popüler birliktelik kuralı algoritması olan Apriori tüm yazılımlarda bulunurken, FP-Growth Algoritması Sadece Keel, RapidMiner ve Weka'da bulunmaktadır.

Nitelik Seçiminde Kazanç Bilgisi, Kazanç Oranı, Ki-Kare, Gini İndeks ve Genetik Algoritma gibi bir çok yöntem bulunmaktadır. Bunlardan en çok kullanılanları Kazanç Bilgisi, Kazanç Oranı ve Ki-Kare'dir. Bu üçünü aynı anda bulunduran yazılımların başında Weka ve RapidMiner gelmektedir.

Veri Ön İşleme için yapılması gereken işlemlerden; Keel, RapidMiner ve Weka yazılımları, eksik değer işlemi, kesikleştirme işlemi, gürültülü veri filtreleme işlemi, normalizasyon işlemi, nominal değerden ikili değere dönüştürme işlemi, çapraz doğrulama işleminin hepsini gerçekleştirebilmektedir

5. Sonuç ve Öneriler

Artan veri miktarından dolayı bilgiye ulaşmak zorlaştıkça, bilgiye ulaşmak için birçok araç geliştirilmektedir. Bu araçların en başında veri madenciliği olarak nitelendirilen büyük miktardaki veriden kullanılabilir bilgiyi üretme kavramı gelmektedir. Veri Madenciliği uygulamaları yapmak için bilgisayar yazılımlarına ihtiyaç duyulmaktadır. Bu yazılımlar birçok veri sınıflandırma, kümeleme, kural çıkarma yöntemi gibi birçok algoritmayı içermektedir. Bu yazılımların kullandıkları algoritmalar sayesinde işlenen ham verilerden, istenilen ve amaçlanan bilginin çıkarımı yapılabilmektedir.

Bu çalışmada açık kaynak kodlu ve popüler olan 6 adet veri madenciliği yazılımı incelenmiştir. Kullanıcı dostluğu, desteklediği dosya formatları, içerdikleri algoritmalar ve makine öğrenmesi paketleri gibi birçok açıdan incelendiğinde tarafımızca en kullanışlı bulunan yazılımlar Weka, RapidMiner (Yale) ve Keel olmuştur. Bu 3 yazılım arasından da öğrenim ve kullanım kolaylığı açısından en başarılı yazılım tarafımızca Weka yazılımı olarak belirlenmiştir.

6. Kaynaklar

[1] Özkan, Y., "Veri Madenciliği Yöntemleri", **Papatya Yayıncılık Eğitim**, İstanbul, (2008).

[2] Silahtaroğlu, G., "Kavram ve Algoritmalarıyla Temel Veri Madenciliği", **Papatya Yayıncılık Eğitim,** İstanbul, (2008).

[3] Akgöbek, Ö. ve Çakır, F., "Veri Madenciliğinde Bir Uzman Sistem Tasarımı", **Akademik Bilişim'09 - XI. Akademik Bilişim Konferansı Bildirileri**, Şanlıurfa, 801-806 (2009).

[4] Tekerek, A., "Veri Madenciliği Süreçleri ve Açık Kaynak Kodlu Veri Madenciliği Araçları", **Akademik Bilişim'11 - XIII. Akademik Bilişim Konferansı Bildirileri**, 2-4 Şubat, İnönü Üniversitesi, Malatya, 161-169 (2011).

[5] Dener, M., Dörterler, M., Orman, A., "Açık Kaynak Kodlu Veri Madenciliği Programları: Weka'da Örnek Uygulama", **Akademik Bilişim'09 - XI. Akademik Bilişim Konferansı Bildirileri**, 11-13 Şubat Harran Üniversitesi, Şanlıurfa, 787-796 (2009).

[6] Bilgin, T.T., "Veri Akışı Diyagramları Tabanlı Veri Madenciliği Araçları ve Yazılım Geliştirme Ortamları", **Akademik Bilişim'09 - XI. Akademik Bilişim Konferansı Bildirileri**, Şanlıurfa, 807-814 (2009).

[7] Han, J., Kamber, M., "Data Mining Concepts and Techniques", **Morgan Kaufmann Publishers**, (2001).

[8] Delen, D., Walker, G., Kadam, A., "Predicting breast cancer survivability: a comparison of three data mining methods", **Artificial Intelligence in Medicine**, vol 34, pp113-127 (2005).

[9] KEEL, http://www.keel.es/, (Erişim Tarihi: 2013).

[10] KNIME, http://www.knime.org/, (Erişim Tarihi: 2013).

[11] ORANGE, http://orange.biolab.si/, (Erişim Tarihi: 2013).

[12] R, http://www.r-project.org/, (Erişim Tarihi: 2013).

[13] RAPIDMINER, http://rapidminer.com/, (Erişim Tarihi: 2013).

[14] YALE, http://yale.sourceforge.net/, (Erişim Tarihi: 2013).

[15] WEKA, http://www.cs.waikato.ac.nz/ml/weka/, (Erişim Tarihi: 2013).

[16] Witten, I. H., Frank, E., "Datamining Practical Machine Learning Tools and Techniques," **Morgan Kaufmann**, Second Edition, San Fransisco, (2005).

[17] Chen X., Ye Y., Williams G. , Xu X., "A Survey of Open Source Data Mining Systems", **Proceeding PAKDD'07 Proceedings of the 2007 international conference on Emerging technologies in knowledge discovery and data mining**, Pages 3-14 (2007).

[18] Zupan B., "Demsar J., Open-source tools for data mining", **Clinics in Laboratory Medicine**, 28(1):37-54, (2008).

[19] Konjevoda P., Štambuk N., "Open-Source Tools for Data Mining in Social Science", **Theoretical and Methodological Approaches to Social Sciences and Knowledge Management**, Asunción López-Varela (Ed.), (2012).

[20] Alcalá-Fdez J., Sánchez L., García S., del Jesus M. J., Ventura S., Garrell J. M., Otero J., Romero C., Bacardit J., Rivas V. M., Fernández J. C., Herrera F.. "KEEL: A Software Tool to Assess Evolutionary Algorithms to Data Mining Problems", **Soft Computing**, 13(3):307-318 (2009).