

## **VERİ MADENCİLİĞİ VE TÜRKİYE'DEKİ UYGULAMA ÖRNEKLERİ**

**Serkan SAVAS<sup>1</sup>, Nurettin TOPALOĞLU<sup>2</sup>, Mithat YILMAZ<sup>3</sup>**

*Geliş: 11.01.2012 Kabul: 22.03.2012 (Tarama Makalesi)*

### **ÖZET**

Günümüz teknolojisi hızla ilerlemekte ve her geçen gün gücü de artmaktadır. Bilgisayarların bilgi saklama kapasitelerinin artmasıyla birlikte bilgi kaydı yapılan alanların sayısı da artmaktadır. Bundan dolayı elde edilen verilerin analizi ve sonucu bu verilerden kestirme yöntemlerinin önemi karar vericiler için gittikçe artmaktadır. Bilgisayar sistemleri ile üretilen veriler tek başlarına değersizdir, çünkü çıplak gözle bakıldığında bir anlam ifade etmezler. Bu veriler belli bir amaç doğrultusunda işlendiği zaman bir anlam ifade etmeye başlar. Bu yüzden büyük miktardaki verileri işleyebilen teknikleri kullanabilmek büyük önem kazanmaktadır. Bu ham veriyi bilgiye veya anlamlı hale dönüştürme işlemleri veri madenciliği ile yapılabilmektedir. Bu çalışmada veri madenciliğinin günümüz disiplinleri arasında geldiği noktaya değinilmiş ve Türkiye'de veri madenciliği üzerine yapılan çalışmalar ve gerçekleştirilen uygulamalar incelenmiştir.

**Anahtar Kelimeler:** *Veri madenciliği, Türkiye'deki uygulamaları.*

### **DATA MINING AND APPLICATION EXAMPLES IN TURKEY**

#### **ABSTRACT**

Today's technology is advancing rapidly and its power is increasing everyday. The number of the fields which are storing information is increasing with the increasing of computers' information storage capacity. Therefore the importance of analyzing data and prediction results from these data is increasing for decision-makers. Data which are produced by computers are worthless alone because they are meaningless when you look with eyes. These data become meaningful when they are processed for an aim. Because of this, using the techniques which can process large amount of data is becoming important. Changing this raw data to information and to significant state is possible with data mining. In this study data mining's position between today's disciplines is mentioned and data mining application examples in Turkey are examined.

**Keywords:** *Data mining, applications in Turkey.*

---

<sup>1</sup> *Teknik Öğretmen, Kızılcahamam Teknik ve Endüstri Meslek Lisesi.*

<sup>2</sup> *Gazi Üniversitesi Teknoloji Fakültesi.*

<sup>3</sup> *Gazi Üniversitesi Teknik Eğitim Fakültesi.*

## 1.GİRİŞ

Bilgisayar sistemleri ile üretilen veriler tek başlarına değersizdir, çünkü çıplak gözle bakıldığında bir anlam ifade etmezler. Bu veriler belli bir amaç doğrultusunda işlendiği zaman bir anlam ifade etmeye başlar (Kalikov, 2006). Bilgi bir amaca yönelik işlenmiş veridir. “Ham veri” veya yalnız geçmişte ne olduğunun bir görüntülemesi olan “enformasyon” a dayalı karar almak mümkün değildir. Geçmişte yaşanan kötü bir tecrübeden kaynaklanan kaybın engellenmesi de mümkün değildir. Önemli olan geçmişe ait olaylara dair gizli bilgilerin keşfedilmesi, ileriye yönelik durumsal öngörüler veren modeller ile önceden tedbir almamızı sağlayacak bir yönetim anlayışına geçmek ve olası kayıpları öngörebilmektir (İnan, 2003). Bu yüzden büyük miktardaki verileri işleyebilen teknikleri kullanabilmek büyük önem kazanmaktadır. Bu ham veriyi bilgiye veya anlamlı hale dönüştürme işlemleri veri madenciliği ile yapılabilmektedir (Kalikov, 2006). Veri madenciliği, bu gibi durumlarda kullanılan büyük miktardaki veri setlerinde saklı durumda bulunan örüntü ve eğilimleri keşfetme işlemidir (Thuarisingham, 2003).

Günümüzde veri madenciliği işletmeler için çok önemli hale gelmiştir. Çok büyük ölçekli veriler, farklı alanlardaki büyük ölçekli veri tabanları içerisinde değerli verileri bulunduran bir veri madeni gibi düşünülebilir. Bu büyüklükteki verilerin analizi, bu analiz sonucunda daha anlamlı bilgi elde etme ve elde edilen bilgiyi yorumlama işi, insan yeteneği ve ilişkisel veri tabanlarının yapabileceklerini aşmaktadır. Bilhassa dijital veri miktarında artış patlaması ve buna karşılık, bu veriler üzerinde araştırma ve uygulama yapan kişilerin sayısının değişmemesi, çalışmaları veri madenciliğine doğru zorlamıştır. Bu ihtiyaçların sonucunda otomatik ve akıllı veri tabanı analizi için yeni kuşak teknikler doğmuştur. Bu teknikler öyle olmalıdır ki, veriyi akıllı ve otomatikleşmiş şekilde işe yarar bilgiye dönüştürebilsin. Tüm bunların sonucunda veri madenciliği cevap olarak sunulmuş ve giderek önemini artıran bir araştırma alanı haline gelmiştir. Bu çalışmada veri madenciliğinin günümüz disiplinleri arasında geldiği nokta, Türkiye’de veri madenciliği üzerine yapılan çalışmalar ve gerçekleştirilen uygulamalar incelenmiştir.

## 2. VERİ MADENCİLİĞİ (DATA MINING)

Veri madenciliği, büyük miktarlardaki verinin içinden geleceği tahmin edilmesinde yardımcı olacak anlamlı ve yararlı bağlantı ve kuralların bilgisayar programlarının aracılığıyla aranması ve analizidir. Ayrıca veri madenciliği, çok büyük miktardaki verilerin içindeki ilişkileri inceleyerek aralarındaki bağlantıyı bulmaya yardımcı olan ve veri tabanı sistemleri içerisinde gizli kalmış bilgilerin çekilmesini sağlayan veri analizi tekniğidir (Kalikov, 2006). Bu işlemlerin uygulama alanı oldukça geniştir. Bu alanlar içerisinde Şekil 1.1’de gösterildiği gibi, veri tabanı sistemleri, Veri Görşelliği, Yapay Sinir Ağları, İstatistik, Yapay Öğrenme, vb. gibi disiplinler bulunmaktadır.



**Şekil 1.1. Veri madenciliği ve disiplinler**

Veri madenciliği araçları kullanılarak, işletmelerin daha etkin kararlar almasına yönelik karar destek sistemlerinde gerekli olan eğilimlerin ve davranış kalıplarının ortaya çıkarılması mümkün olmaktadır. Geçmişteki klasik karar destek sistemlerinin kullanıldığı araçlardan farklı olarak, veri madenciliğinde çok daha kapsamlı ve otomatize edilmiş analizler yapmaya yönelik, birçok farklı özellik bulunmaktadır (İnan, 2003).

Veri madenciliğinin işletmelere sunduğu en önemli özellik, veri grupları arasındaki benzer eğilimlerin ve davranış kalıplarının belirlenmesidir. Bu süreç aynı zamanda otomatize edilmiş bir biçimde hayata geçirilebilmektedir. Bu fonksiyon özellikle hedef pazarlara yönelik pazarlama faaliyetlerinde yoğun olarak kullanılmaktadır (İnan, 2003). Başka bir özelliği ise daha önceden bilinmeyen, veri ambarları içerisinde bulunan ancak ilk etapta görülemeyen bilgilerin ortaya çıkarılabilmesidir. Örneğin bir firma sattığı ürünleri analiz ederek, ilerideki kampanyalarını şekillendirebilir ya da sattığı ürünler arasındaki bağları keşfedebilir. Burada amaç daha önceden fark edilmeyen veri kümelerinin bulunabilmesidir.

Günümüzün ekonomik koşulları ve yaşanan hızlı değişim ortamlarında, iş deneyimi ve önseziilere dayanarak alınan kararlarda yanlış karar alma riski çok yüksektir. Riski azaltmanın tek yolu bilgiye dayalı yönetimi öngören karar destek çözümleridir. Veri madenciliği teknikleri gerçek anlamda bir karar destek sistemi oluşturmada olmazsa olmaz araçlardır. Bu noktada bilgi teknolojilerinden yararlanmak kaçınılmaz olmuştur.

## 2.1. Veri Madenciliğinin Tanımı

Bu güne kadar farklı kaynaklarda veri madenciliğinin pek çok tanımıyla karşılaşmıştır. Bu kaynaklardan bazılarında göre veri madenciliğinin tanımı şöyledir:

- Jacobs (1999), veri madenciliğini, ham datanın tek başına sunamadığı bilgiyi çıkararak, veri analizi süreci olarak tanımlamıştır (Jacobs, 1999).
- Veri madenciliği, büyük veri yığınları arasından gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağlantıların, bilgisayar programı kullanarak aranması işidir (Doğan ve Türkoğlu, 2007).
- Hand (1998), veri madenciliğini istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veritabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak tanımlamıştır (Hand, 1998).
- Kitler ve Wang (1998), veri madenciliğini oldukça tahminci anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesini sağlama yeteneği olarak tanımlamışlardır (Kitler ve Wang, 1998).

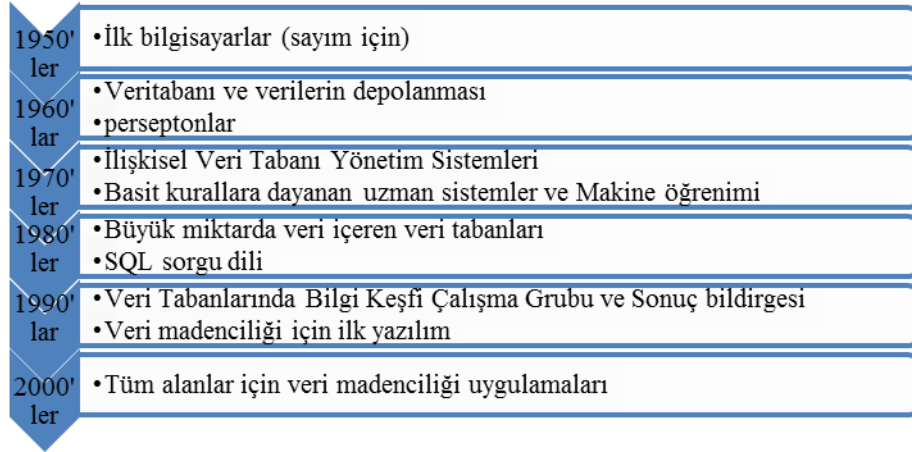
Bu tanımlardan yola çıkarak şöyle bir tanım yapmak mümkündür: Veri madenciliği, çok büyük miktarda bilginin depolandığı veri tabanlarından, amacımız doğrultusunda, gelecek ile ilgili tahminler yapmamızı sağlayacak, anlamlı olan veriye ulaşma ve veriyi kullanma işidir.

## 2.2. Veri Madenciliğinin Tarihi

Günümüzde neredeyse her eve bilgisayar girmiştir ve internet erişimi hemen hemen her yerden sağlanmaktadır. Disk kapasitelerinin artması, her yerden bilgiye ulaşma olasılığı, bilgisayarların çok büyük miktarlarda veri saklamasına ve daha kısa sürede işlem yapmasına olanak sağlamıştır. Geçmişten günümüze veriler her zaman yorumlanmış, bilgi elde etmek istenmiştir ve bunun için de donanımlar oluşturulmuştur. Bu sayede bilgi, geçmişten günümüze taşınır hale gelmiştir.

1950'li yıllarda ilk bilgisayarlar sayımlar için kullanılmaya başlamıştır. 1960'larda ise veri tabanı ve verilerin depolanması kavramı teknoloji dünyasında yerini almıştır. 1960'ların sonunda bilim adamları basit öğrenmeli bilgisayarlar geliştirebilmişlerdir. Minsky ve Papert, günümüzde sinir ağları olarak bilinen perseptron'ların sadece çok basit olan kuralları öğrenebileceğini göstermişlerdir (Adriaans ve Zantinge, 1997). 1970'lerde İlişkisel Veri Tabanı Yönetim Sistemleri uygulamaları kullanılmaya başlanmıştır. Bilgisayar uzmanları bununla beraber basit kurallara dayanan uzman sistemler geliştirmişler ve basit anlamda makine öğrenimini sağlamışlardır. 1980'lerde veri tabanı yönetim sistemleri yaygınlaşmış ve bilimsel alanlarda, mühendisliklerde vb. alanlarda uygulanmaya başlanmıştır. Bu yıllarda şirketler, müşterileri, rakipleri ve ürünleri ile ilgili verilerden oluşan veri tabanları oluşturmuşlardır. Bu veri tabanlarının içerisinde çok büyük miktarlarda

veri bulunmaktadır ve bunlara SQL veri tabanı sorgulama dili ya da benzeri diller kullanarak ulaşılabilir. 1990'larda artık içindeki veri miktarı katlanarak artan veri tabanlarından, faydalı bilgilerin nasıl bulunabileceği düşünölmeye başlanmıştır. Bunun üzerine çalışmalara ve yayınlara başlanmıştır. 1989, KDD (IJCAI)-89 Veri Tabanlarında Bilgi Keşfi Çalışma Grubu toplantısı ve 1991, KDD (IJCAI)-89'un sonuç bildirgesi sayılabilecek "Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop" makalesinin KDD (Knowledge Discovery and Data Mining) ile ilgili temel tanım ve kavramları ortaya koyması ile süreç daha da hızlanmış ve nihayet 1992 yılında veri madenciliği için ilk yazılım gerçekleştirilmiştir. 2000'li yıllarda veri madenciliği sürekli gelişmiş ve hemen hemen tüm alanlara uygulanmaya başlanmıştır. Alınan sonuçların faydaları göröldükçe, bu alana ilgi artmıştır. Veri madenciliğinin tarihsel gelişim süreci, Şekil 1.2'de gösterilmiştir.



Şekil 1.2. Veri madenciliğinin tarihsel süreci

### 2.3. Veri Madenciliğinin Kullanıldığı Alanlar

Büyük hacimde veri bulunan her yerde veri madenciliği kullanmak mümkündür. Günümüzde karar verme sürecine ihtiyaç duyulan birçok alanda veri madenciliği uygulamaları yaygın olarak kullanılmaktadır. Örneğin pazarlama, biyoloji, bankacılık, sigortacılık, borsa, perakendecilik, telekomünikasyon, genetik, sağlık, bilim ve mühendislik, kriminoloji, sağlık, endüstri, istihbarat vb. birçok dalda başarılı uygulamaları görölmektedir (İnan, 2003; Albayrak, 2008; Akgöbek ve Çakır, 2009).

Son 20 yıldır Amerika Birleşik Devletleri'nde çeşitli veri madenciliği algoritmalarının gizli dinlemeden, vergi kaçakçılıklarının ortaya çıkartılmasına kadar çeşitli uygulamalarda kullanıldığı bilinmektedir. Kaynaklar incelendiğinde veri madenciliğinin en çok kullanıldığı alan olarak tıp, biyoloji ve genetik görölmektedir.

#### 2.4. Veri Madenciliğini Etkileyen Etmenler

Veri madenciliği temel olarak 5 ana faktörden etkilenir (Akpınar, 2000):

1. *Veri*: Veri madenciliğinin bu kadar gelişmesindeki en önemli faktördür.
2. *Donanım*: Gelişen bellek ve işlem hızı kapasitesi sayesinde, birkaç yıl önce madencilik yapılamayan veriler üzerinde çalışmayı mümkün hale getirmiştir.
3. *Bilgisayar ağları*: Yeni nesil internet, çok yüksek hızları kullanmayı sağlamaktadır. Böyle bir bilgisayar ağı ortamı oluştuktan sonra, dağıtık verileri analiz etmek ve farklı algoritmaları kullanmak mümkün olacaktır.
4. *Bilimsel hesaplamalar*: Günümüz bilim adamları ve mühendisleri, simülasyonu, bilimin üçüncü yolu olarak görmekteler. Veri madenciliği ve bilgi keşfi, teori, deney ve simülasyonu birbirine bağlamada önemli bir rol almaktadır.
5. *Ticari eğilimler*: Günümüzde, işletmeler rekabet ortamında varlıklarını koruyabilmek için daha hızlı hareket etmeli, daha yüksek kalitede hizmet sunmalı, bütün bunları yaparken de minimum maliyeti ve en az insan gücünü göz önünde bulundurmalıdır.

#### 2.5. Veri Madenciliğinde Karşılaşılan Problemler

Büyük hacimli verilerin bulunduğu veri ortamlarında büyük sorunlar ortaya çıkabilir. Bu nedenle küçük veri kümelerinde, benzetim ortamlarında hazırlanmış veri madenciliği sistemleri, büyük hacimli, eksik, gürültülü, boş, atık, aykırı veya belirsiz veri kümelerinin bulunduğu ortamlarda yanlış çalışabilir. Bu nedenle veri madenciliği sistemleri hazırlanırken bu sorunların çözülmesi gerekmektedir.

Veri madenciliği uygulamalarında karşılaşılabilecek sorunlar şunlardır:

**Artık veri:** Artık veri, problemde istenilen sonucu elde etmek için kullanılan örneklem kümesindeki gereksiz niteliklerdir. Bu durum pek çok işlem sırasında karşımıza çıkabilir.

**Belirsizlik:** Yanlışlıkların şiddeti ve verideki gürültünün derecesi ile ilgilidir.

**Boş veri:** Bir veri tabanında boş değer, birincil anahtarda yer almayan herhangi bir niteliğin değeri olabilir. Boş değer, tanımı gereği kendisi de dâhil olmak üzere hiçbir değere eşit olmayan değerdir.

**Dinamik veri:** Kurumsal çevrim içi veri tabanları dinamiktir ve içeriği sürekli olarak değişir. Bu durum, bilgi keşfi metotları için önemli sakıncalar doğurmaktadır.

**Eksik veri:** Veri kümesinin büyüklüğünden ya da doğasından kaynaklanmaktadır. Eksik veriler olduğunda yapılması gerekenler şunlardır:

- Eksik veri içeren kayıt veya kayıtlar çıkarılabilir.
- Değişkenin ortalaması eksik verilerin yerine kullanılabilir.
- Var olan verilere dayalı olarak en uygun değer kullanılabilir.

Eksik veriler, yapılacak olan istatistiksel analizlerde önemli problemler yaratmaktadır. Çünkü istatistiksel analizler ve bu analizlerin yapılmasına olanak veren ilgili paket programlar, verilerin tümünün var olduğu durumlar için geliştirilmiştir (Albayrak, 2008).

**Farklı tipteki verileri ele alma:** Gerçek hayattaki uygulamalar makine öğreniminde olduğu gibi yalnızca sembolik veya kategorik veri türleri değil, fakat aynı zamanda tamsayı, kesirli sayılar, çoklu ortam verisi, coğrafi bilgi içeren veri gibi farklı tipteki veriler üzerinde işlem yapılmasını gerektirir.

**Gürültülü ve kayıp değerler:** Veri girişi veya veri toplanması esnasında oluşan sistem dışı hatalara gürültü denir. Büyük veri tabanlarında pek çok niteliğin değeri yanlış olabilir. Veri toplanması esnasında oluşan hatalara ölçümden kaynaklanan hatalar da dâhil olmaktadır. Bu hataların sonucu olarak birçok niteliğin değeri yanlış olabilir ve bu yanlışlardan dolayı veri madenciliği amacına tam olarak ulaşmayabilir.

**Sınırlı bilgi:** Veri tabanları genel olarak basit öğrenme işlerini sağlayan özellik veya nitelikleri sunmak gibi veri madenciliği dışındaki amaçlar için hazırlanmışlardır. Bu yüzden, öğrenme görevini kolaylaştıracak bazı özellikler bulunmayabilir.

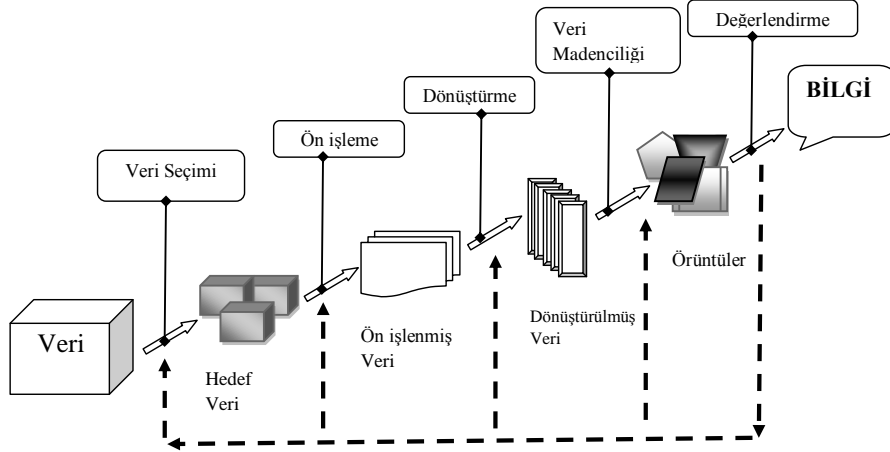
**Veri tabanı boyutu:** Veri tabanı boyutları büyük bir hızla artmaktadır. Veri tabanı algoritması çok sayıda küçük örnekleme ele alabilecek biçimde geliştirilmiştir. Aynı algoritmaların yüzlerce kat büyük örneklemlerde kullanılabilmesi için çok dikkat gerekmektedir.

## 2.6. Veri Madenciliği Süreci

Veri madenciliği, aynı zamanda bir süreçtir. Veri yığınları arasında, soyut kazılar yaparak veriyi ortaya çıkarmanın yanı sıra, bilgi keşfi sürecinde örüntüleri ayrıştırarak süzmek ve bir sonraki adıma hazır hale getirmek de bu sürecin bir parçasıdır. Bu süreç Şekil 1.3'de gösterilmiştir. Üzerinde inceleme yapılan işin ve verilerin özelliklerinin bilinmemesi durumunda ne kadar etkin olursa olsun hiç bir veri madenciliği algoritmasının fayda sağlaması mümkün değildir. Bu sebeple, veri madenciliği sürecine girilmeden önce, başarının ilk şartı, iş ve veri özelliklerinin detaylı analiz edilmesidir.

Veri madenciliği sürecinde izlenen adımlar genellikle aşağıdaki şekildedir (Shearer, 2000):

1. Problemin tanımlanması,
2. Verilerin hazırlanması,
3. Modelin kurulması ve değerlendirilmesi,
4. Modelin kullanılması,
5. Modelin izlenmesi.



**Problemin tanımlanması:** Veri madenciliği çalışmalarında başarılı olmanın en önemli şartı, projenin hangi işletme amacı için yapılacağını ve elde edilecek sonuçların başarı düzeylerinin nasıl ölçüleceğinin tanımlanmasıdır.

**Verilerin hazırlanması:** Modelin kurulması aşamasında ortaya çıkacak sorunlar, bu aşamaya sık sık geri dönülmesine ve verilerin yeniden düzenlenmesine neden olacaktır. Bu durum verilerin hazırlanması ve modelin kurulması aşamaları için, bir analistin veri keşfi sürecinin toplamı içerisinde enerji ve zamanının %50 - %85'ini harcamasına neden olmaktadır (Piramuthu,1998). Verilerin hazırlanması, “toplama”, “değer biçme”, “birleştirme ve temizleme”, “örneklem seçimi” ve “dönüştürme” aşamalarından oluşmaktadır.

**Modelin kurulması ve değerlendirilmesi:** Tanımlanan problem için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veri hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir.

**Modelin kullanılması:** Kurulan ve geçerliliği kabul edilen model doğrudan bir uygulama olabileceği gibi, bir başka uygulamanın alt parçası olarak kullanılabilir.

**Modelin izlenmesi:** Zaman içerisinde bütün sistemlerin özelliklerinde ve dolayısıyla ürettikleri verilerde ortaya çıkan değişiklikler, kurulan modellerin sürekli olarak izlenmesini ve yeniden düzenlenmesini gerektirecektir.

## 2.7. Veri Madenciliği Metotları

Veri madenciliği ile ilgili kullanılan pek çok yöntemin yanına hemen her geçen gün yeni yöntem ve algoritmalar eklenmektedir. Bunlardan bir kısmı onlarca yıldır kullanılan klasik teknikler diyebileceğimiz, ağırlıklı olarak istatistiksel yöntemlerdir.



Diğer yöntemler de genellikle istatistiği temel alan ama daha çok makine öğrenimi ve yapay zekâ destekli yeni nesil yöntemlerdir.

Veri madenciliği modelleri, gördükleri işlemlere göre temel olarak 3 grupta toplanır. Bunlar:

1. Sınıflama (*Classification*) ve Regresyon (*Regression*),
2. Kümeleme (*Clustering*),
3. Birliktelik Kuralları (*Association Rules*),

olmak üzere üç ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları modelleri tanımlayıcı modellerdir (Özkes, 2003).

### **3. TÜRKİYE'DEKİ VERİ MADENCİLİĞİ ÇALIŞMALARI ve UYGULAMALARI**

Pek çok alanda etkili bir şekilde kullanılmaya başlanan veri madenciliği, günümüzün en çok uygulanan disiplinlerinden birisi olmuştur. Her geçen sene kendisine daha da yaygın bir kullanım alanı bulmakla birlikte, kolay uygulanabilirliği ve etkili sonuçlar ortaya çıkarması sayesinde, kurum ve kuruluş yöneticileri tarafından en çok başvurulan yöntemlerden bir tanesidir. Literatür taramasıyla ulaşılan veri madenciliği ile gerçekleştirilmiş uygulamaları, eğitim, ticaret, mühendislik, bankacılık ve borsa, tıp ve telekomünikasyon başlıkları arasında sınıflandırarak şu şekilde özetleyebiliriz.

#### **3.1. Mühendislik Alanında Gerçekleştirilen Veri Madenciliği Uygulamaları**

Kıyas Kayaalp tarafından 2007 yılında yapılan bir yüksek lisans çalışmasında, veri madenciliği tekniği ile üç fazlı asenkron motordaki sargı spiralleri arasında oluşabilecek kısa devre veya yalıtım bozuklukları ve motor milinde oluşabilecek mekanik dengesizlik hatalarının tespiti gerçekleştirilmiştir (Kayaalp, 2007).

Ali İnan tarafından 2006 yılında yapılan bir çalışmada şu bulgulara ulaşılmıştır: Kişilerin konum bilgilerinin toplanması, kullanımı ve dağıtılması ile ilgili gizlilik kaygıları zaman-mekân bilgisi içeren verilerde veri madenciliği teknikleri uygulanmasının önündeki tek engeldir. Kimlik belirteçlerinin veriden temizlenmesi kişisel gizliliğin sağlanmasında tek başına yeterli olamıyor çünkü umuma açık ev ve işyeri adresleri kullanılarak kişilerin hareket yörüngeleri ile kimliklerinin eşlenmesi mümkündür. Var olan gizliliği koruyan veri madenciliği teknikleri de yeterli olmuyor çünkü bu tekniklerin zaman-mekân bilgisi içeren verilere uygulanabilmesi için ardışık konum gözlemlerinin kişilerin birbirinden bağımsız nitelikleri olduğunu varsaymak gerekmektedir. Ancak bu varsayım hatalı olacaktır. Bu nedenle konum-zaman veri tabanlarında veri madenciliğini mümkün kılmak, bu tip veriler için özel olarak tasarlanmış algoritmalar gerektirir. Bu çalışmada zaman-mekân nitelikleri olan veriler için bir gizliliği koruyan veri madenciliği tekniği ve iki ön-işleme tekniği önerilmiştir: (1) Dağıtık kümeleme, (2) Merkezi anonimleştirme ve (3)

Dağıtık anonimleştirme. Önerilen tekniklerin güvenlik ve performans analizleri de yapılmış ve sonuçta mantıklı varsayımlar altında minimum mahrem bilgi kaybıyla veri madenciliğinin mümkün olduğu gözlemlenmiştir (İnan, 2006).

Gökhan Yavaş tarafından 2003 yılında gerçekleştirilen başka bir çalışmada ise mobil kullanıcıların hareket modellerinin veri madenciliği kullanılarak çıkarılması ve bu modeller kullanılarak mobil kullanıcıların daha sonraki hareketlerinin tahmin edilmesi için yeni bir algoritma geliştirilmiştir. Üç aşamadan oluşan bu algoritmanın ilk aşamasında kullanıcı hareket modelleri, kullanıcıların önceden kaydedilmiş mobil yörüngelerinden veri madenciliği kullanılarak çıkarılmaktadır. İkinci aşamada bulunan hareket modellerinden hareket kuralları üretilmekte, son aşamada ise bu hareket kuralları kullanıcının bir sonraki hücreler arası hareketinin tahmini için kullanılmaktadır. Sunulan algoritmanın performansı simülasyonlar yardımıyla iki farklı tahmin yöntemiyle karşılaştırılmıştır. Performans sonuçları algoritmanın diğer metotlardan daha doğru tahminler yapabildiğini göstermiştir (Yavaş, 2003).

Sibel Kırmızıgül Çalışkan ve İbrahim Soğukpınar 2008 yılında, veri madenciliği yöntemlerinden “K-means” ve “K en yakın komşu” yöntemlerinin iyileştirilmesi amacıyla; nüfuz tespiti için kümelemeyi ve sınıflandırmayı, denetimli ve denetimsiz öğrenimi, k-means ve k en yakın komşu yöntemlerini bir arada kullanan hibrit bir yapı geliştirmiştir. Farklı boyutlardaki veri gruplarında düşük performans gösterebilen, fakat gerçekleştirilmesi kolay ve zaman karmaşası az olan “K-means” ile tek ve geniş bir küme için belirlenen k ve eşik değeri, küme içindeki farklı özelliklere sahip normal davranış ve saldırı verileri için zorunlu kılan ve zaman karmaşası çok olan, fakat k komşu ortalaması aldığı için gürültülü verilerden az etkilenen “k en yakın komşu” yöntemleri bir arada kullanılmıştır. Geliştirilen uygulamada en hızlı sonucu veren k-means uygulaması ile test kümesi daha küçük alt kümelere ayrılarak k en yakın komşu yönteminin zaman karmaşası ve bellek gereksinimi azaltılmıştır (Çalışkan ve Soğukpınar, 2008).

N. Duru ve M. Canbay 2007 yılında veri madenciliği ile deprem verilerinin analizi üzerine bir çalışma gerçekleştirmiştir. Bu çalışma deprem verileri kullanılarak seçilen bir bölgeye ait sismik tehlikenin diğer deyişle gerçekleşme olasılığının veri madenciliği yönünden ele alınarak incelenmesini kapsamaktadır. Çalışma sonuçları jeofizik sonuçlar ile korele edilerek doğruluk payı da araştırılmıştır. Her gelecek 10 yıl için % sismik tehlike değeri artış göstererek devam etmiş, örneğin 6 magnitudündeki bir depremin olma olasılığı 10 yıl içinde %27 iken, 30 yıl içinde %60 ve 60 yıl için de %80’leri bulmaktadır. Bu değerler daha önce çalışma bölgesinde yapılmış çalışmalarla uyum göstermektedir. Ancak burada unutulmaması gereken bu çalışmanın deprem tahmini için kullanılan tekniklerden sadece birisi olduğu ve bu çalışmanın konusu itibarıyla çalışma bölgelerinin tektonik özelliklerini hiç irdelemeden dahi olsa olumlu sonuçlara varılabilmesinin mümkün olduğunun gösterilebilmesidir. Ayrıca yapılan çalışmanın sonuçlarının büyük bölgelere göre küçük bölgelerde daha iyi sonuç verdiğinin görülmesidir. Uygulama, dünya ölçeğindeki her noktanın analizini yapacak şekilde geliştirilmiş olup, ihtiyaç halinde programa eklemeler yapmak suretiyle, başka bu tür çalışmalar yapacak şekilde tasarlanmıştır (Duru ve Canbay, 2007).

Yaşar Doğan tarafından 2004 yılında Deniz Harp Okulu'nda, su altı taktik duyurga ağlarında veri madenciliği tabanlı hedef sınıflandırması çalışması hazırlanmıştır. Bu çalışmada, açık, sığ ve çok sığ sularda denizaltı, küçük sualtı taşıma araçları, sualtı mayınları ve dalgıçları sınıflandırmada maliyeti çok az olan mikroduyurgalar kullanılmıştır. Algoritma, yüzeydeki şamandıralara bağlı ve ayarlanabilir derinliklere indirilebilen duyurgalardan oluşan taktik su altı duyurga ağları için tasarlanmıştır. Sınıflandırma veri madenciliği tekniği olarak karar ağacı algoritmaları kullanılmıştır (Doğan, 2004).

Eyüp Sıramkaya'nın 2005 yılında hazırladığı bir uygulamada internet üzerinden ulaşılabilen basın-yayın kaynaklarında yer alan görsel ve metinsel verilerin hızlı ve etkin bir şekilde erişimi ve bu kaynaklardan anlamlı ve önemli bilgilerin çıkarılması hedeflenmiştir. Çalışmalar istihbarat açısından önem taşıyan kişi ve örgütlerle ilgili haberler üzerinde yoğunlaşmıştır. Sunucu bilgisayarda internet üzerinde yer alan haber kaynaklarından toplanmış ve işlenmiş metinsel belgelerden oluşan veri-tabanı ile bu bilgileri işleyen uygulama yazılımları bulunmaktadır. Bir arayüz ile kullanıcının bu bilgileri sorgulaması sağlanmıştır. Çalışma, Birliktelik Kural Madenciliği tekniği ile uygulanmıştır. Bu teknik uygulanırken Apriori Algoritması kullanılmıştır. Yapılan veri madenciliği çalışmasında Bulanık Mantık çalışması, kişi-kişi ilişkilerini bulmakta uygulanmıştır. Bu uygulamadaki amaç kullanıcıların arama yapmak istedikleri kişilerin isimlerini yazarken yapabilecekleri yazım hatalarını elemektir. İsimlerdeki harflerin konumlarının birbirlerine göre uzaklıklarını temel alarak bulanık mantık kurallarının uygulandığı bir algoritma kullanılmıştır (Sıramkaya, 2005).

Yomi Kastro 2006 yılında, bir yazılımın yeni sürümlerindeki hata oranını eski sürümlerine göre olan değişikliklerini temel alarak tahmin eden bir model ortaya koyma amaçlı bir uygulama gerçekleştirmiştir. Bu uygulamada bahsedilen değişiklikler yazılımdaki bir yenilik, bir algoritma değişikliği ve hatta bir hata ayıklama değişikliği olabilir. Bu tür değişikliklerin türünü formel ve nesnel bir bakış açısıyla analiz ederek ve buna yazılımın hacimsel değişikliğini de katarak, yeni sürümündeki hata oranını doğru bir şekilde tahmin edebilme amaçlanmıştır. Bu araştırmada önerilen modeli kullanarak, yazılım hayat döngüsündeki test sürecini kısaltabilmek ve harcanan gücü azaltabilmek mümkün olmuştur. Buna ek olarak, yeni bir yazılım sürümünün sağlamlığını saptamak bu model sayesinde mümkündür. Bu model, aynı zamanda bir yazılım ürününe katılan yeniliklerin, hata ayıklama değişiklikleri gibi değişiklik türlerinin, hata oluşturma ihtimallerine olan katkısını ayrı ayrı anlamaya yardımcı olmaktadır (Kastro, 2006).

Seda Dağlar Toprak tarafından yeni bir melez çok ilişkili veri madenciliği tekniği 2005 yılında gerçekleştirilmiştir. Bu çalışmada kavram öğrenme, kavram ile kavramı gerçekleştirme önkoşulları arasındaki eşleştirme olarak tanımlanmış ve ilişkisel kural madenciliği alanında buluşsal yöntem olarak kullanılan Apriori kuralı örüntü uzayını küçültmek amacı ile kullanılmıştır. Önerilen sistem, kavram örneklerinden ters çözümlük operatörü kullanılarak genel kavram tanımlarını oluşturan ve bu genel örüntüleri Apriori kuralını temel alan bir operatör yardımı ile

özelleştirerek güçlü kavram tanımlamaları elde eden melez bir öğrenme sistemi olarak tanımlanmıştır. Sistemin iki farklı sürümü, üç popüler veri madenciliği problemi için test edilmiş ve sonuçlar önerilen sistemin, en gelişkin ilişkisel veri madenciliği sistemleri ile karşılaştırılabilir durumda olduğunu göstermiştir (Toprak, 2005).

Coşku Erdem, 2006 yılında, matematiksel morfoloji kullanarak yoğunluk temelli kümeleme adında bir uygulama gerçekleştirmiştir. Bu uygulamadaki algoritma veri depolarının imgelere benzerliğinden yola çıkarak bir imge işleme tekniği olan gri tonlu morfolojinin çok boyutlu veri üzerine uygulanması temeline dayanmaktadır. Önerilen bu algoritmanın gerek sentetik gerekse doğal veri üzerindeki başarımı değerlendirilmiş ve uygun parametrelerle çalıştırıldığında başarılı ve yorumlanabilir sonuçlar üretebildiği görülmüştür. Ek olarak, algoritmanın işlemsel karmaşıklığının düşük boyutlu veri için veri noktası sayısı ile doğrusal, yüksek boyutlu veri içinse temelde morfoloji işlemlerine bağlı olarak boyut sayısı ile üstel olarak artışı hesaplanmıştır (Erdem, 2006).

T. Tugay Bilgin ve A. Yılmaz Çamurcu, veri madenciliğinde güncel araştırma alanlarından biri olan çok boyutlu veri tabanları ve bunların görselleştirilmesinde kullanılan görselleştirme tekniklerini incelemiş ve bu alanda çalışmalar gerçekleştiren araştırma grupları ve bunların geliştirdikleri yeni yöntemler ve teknikleri irdelemiştir. Ayrıca başka bir çalışmada T. Tugay Bilgin, veri akış diyagramları ve veri akışı tabanlı veri madenciliği süreçleri görselleştirilmesini açıklamıştır. Üç farklı tür veri akışı tabanlı yazılımı incelemiş ve detaylı özelliklerini karşılaştırmıştır (Bilgin ve Çamurcu, 2008; Bilgin, 2009).

2004 yılında Serkan Toprak tarafından, ilişkisel veri tabanları üzerinde çoklu ilişkisel yapıdaki ortak kuralları bulmayı sağlayan bir uygulama geliştirilmiştir. Uygulama altyapısı olarak ilişkisel veri tabanlarındaki desenleri tanımlayabilen, bu desenleri eklerle geliştirebilen ve bu desenlerin çeşitli ölçmeleri için gerekli sayımları veri tabanından temel yetilerle alan bir yapı kullanılmıştır. Bu altyapı, veri tabanının tanımında yer alan bilgileri kullanarak arama alanının daraltılmasını sağlamıştır. Bu çalışma, Apriori algoritmasını arama alanını daha da küçültmek için kullanarak ve altyapı tarafından desteklenmeyen özyinelemeli desenlerin bulunmasını sağlayarak altyapıya yenilikler getirmiştir. Apriori algoritması her tablo üzerinde sık karşılaşılan desenleri bulmak için kullanılmış ve bu algoritmanın gerekli destek değerini bulma yöntemi değiştirilmiştir. Veri tabanındaki özyinelemeli ilişkileri belirlemek için bir yöntem sunulmuş ve uygulama bu durumlar için tablo kısaltmalarının kullanıldığı bir çözüm sağlamıştır. Veri tabanı alanlarında saklanan sürekli değerleri bölümleyebilmek için eşit derinlik yöntemi kullanılmıştır. Uygulama bir veri madenciliği yarışması olan KDD Cup 2001'den alınan örnek genlerde yer tahmini problemi ile test edilmiş ve ortaya çıkan sonuçlar yarışmayı kazanan yaklaşımın sonuçlarıyla karşılaştırılmıştır (Toprak, 2004).

Ulaş Baran Baloğlu tarafından 2006 yılında gerçekleştirilen uygulamada, DNA veri kümesinde bulunan biyolojik sıralar üzerinde veri madenciliği yapılarak tekrarlı

örüntüler ve potansiyel motifler çıkartılmıştır. Önerilen yöntem yukarıdan-aşağı veri madenciliği ve genetik algoritma tabanlı hibrit bir çözümdür. Bu yöntemdeki yaklaşım iki temel adımda ele alınabilir. Birinci adım, genetik algoritma kullanılarak aday motiflerin bir popülasyonunun oluşturulmasıdır. Bunu diğer nesillerin genetik operatörler ve uygunluk fonksiyonu kullanılarak oluşturulması takip eder. İkinci adımda, veri madenciliği yöntemi yukarıdan-aşağı haliyle kullanılarak aday motiflerin uygunluğunun değerlendirilmesi yapılır. E. coli bakterilerinden alınmış DNA sıralarında önerilen yöntem denenerek uygulanabilirliği ve üstün yanları gösterilmiştir (Baloğlu, 2006).

Barış Yıldız 2010 yılında, sık kümelerin bulunması için gizliliği koruyan bir yaklaşım önermiştir. Ayrıca bu çalışmada Matrix Apriori algoritması üzerinde değişiklikler yapılmış ve sık küme gizleme çerçevesi de geliştirilmiştir (Yıldız, 2010).

Yasemin Kılınç 2009 yılında hazırladığı bir çalışmada, birliktelik kuralları için bir yöntem sunmuştur. Apriori algoritmasının ürettiği kurallar elenerek bir elektronik firmasında üretim ve mal giriş kalite verileri üzerinde uygulanmıştır. Ortaya çıkarılan kurallar test verileri ile doğrulanmış ve sonuçlar analiz edilmiştir (Kılınç, 2009).

### **3.2. Tıp Alanında Gerçekleştirilen Veri Madenciliği Uygulamaları**

Barış Aksoy tarafından 2009 yılında Dekompresyon Analizinin Cluster Analizi üzerine bir veri madenciliği uygulaması gerçekleştirilmiştir. Bu çalışmada, farklı clustering algoritmaları (k-ortalama, COBWEB, EM ) ile Divers Alert Network (Dalğışların Acil Durum Ağı)'nın dalış yaralanmaları bildirim formlarından elde edilen belirti ve bulgu listeleri kullanılarak dekompresyon hastalığı sınıflandırılmış ve sonuçlar klasik sınıflandırma yöntemleri, yeni yapılan istatistiksel sınıflandırma yöntemleri ve tedavi sonuçları ile karşılaştırılmıştır. Ayrıca teşhiste yardımcı olabilecek birliktelik kuralları (association rules) elde edilmiştir. Sonuç olarak, clustering yöntemleriyle elde edilen sınıfların yeni yapılan istatistiksel sınıflandırmalarla ve klasik sınıflandırmalarla uyumlu olduğu ve hafiften şiddetli vakalara giden hiyerarşik yapıda olduğu gözlemlenmiştir (Aksoy, 2009).

Pınar Yıldırım, Mahmut Uludağ ve Abdülkadir Görür tarafından 2008 yılında yapılan çalışmada, hastane bilgi sistemlerindeki veri madenciliği uygulamalarına değinilmiştir (Yıldırım vd., 2008) .

Şengül Doğan ve İbrahim Türkoğlu tarafından 2008 yılında gerçekleştirilen bir çalışmada, kan biyokimya parametreleri ile demir eksikliği anemisi teşhisinde, hekime yardımcı olacak ve kolaylık sağlayabilecek bir karar destek sistemi oluşturulmuştur. Örüntü tanıma süreci esas alınmış olup, sistemin işleyişi veri madenciliği tekniklerinden olan karar ağaçları yapısı ile sağlanmaktadır. Sisteme giriş olarak, biyokimya parametrelerinden demir eksikliği anemisi hastalığı için temel belirleyiciler olan Serum demiri, Serum demir bağlama kapasitesi (SDBK) ve

Ferritin enzimleri kullanılarak, çıkış olarak da Anemi(+) ve Anemi(-) değerlendirmelerinde bulunulmuştur. Tasarlanan sistemde 96 hasta verisi değerlendirilmiştir. Karar destek sisteminin sonuçları, doktorun verdiği kararlarla tamamen örtüşmüştür (Doğan ve Türkoğlu, 2008).

Mustafa Danacı, Mete Çelik ve A. Erhan Akkaya tarafından 2010 yılında gerçekleştirilen çalışmada kanser çeşitlerinden biri olan ve kadınlar arasında en sık görülen meme kanseri hakkında kısa bilgi verilmiştir. Daha sonra Xcyt örüntü tanıma programı yardımı ile doku hakkında genel veriler elde edilmiş, Weka programı kullanılarak meme kanseri hücrelerinin tahmin ve teşhisi yapılmıştır (Danacı vd., 2010).

### **3.3. Bankacılık ve Borsa Alanında Gerçekleştirilen Veri Madenciliği Uygulamaları**

Nihal Ata, Eregül Özkök ve Uğur Karabey tarafından 2007 yılında gerçekleştirilen bu çalışmada, yaşam çözümlenmesi yöntemlerini veri madenciliği konusu çerçevesinde ele aldıktan sonra kredi kartı sahiplerine ait bir veri kümesi için yaşam olasılıkları, hazard olasılıkları ve regresyon modelleri incelemiştir. Buna göre çalışmada yaş, gelir ve medeni durumun, müşterilerin kredi kartı kullanmayı bırakmalarını etkileyen önemli risk faktörleri olduğu görülmüştür (Ata vd., 2008).

Ali Sait Albayrak ve Şebnem Koltan Yılmaz tarafından 2009 yılında gerçekleştirilen bir çalışmada, İMKB 100 endeksinde sanayi ve hizmet sektörlerinde faaliyet gösteren 173 işletmenin 2004–2006 yıllarına ait yıllık finansal göstergelerinden yararlanarak veri madenciliği tekniklerinden birisi olan karar ağaçları tekniği uygulanmıştır. Seçilen finansal göstergelere göre sanayi ve hizmet sektörlerinde faaliyet gösteren firmaları ayıran en önemli değişkenler saptanmıştır. Ayrıca Ali Sait Albayrak tarafından gerçekleştirilen başka bir çalışmada, yerli ve yabancı olarak önceden grup üyeliği belirlenmiş bankaların sınıflandırmasında yaygın olarak kullanılan veri madenciliği tekniklerinden, diskriminant, lojistik regresyon ve karar ağacı modelleri karşılaştırılmıştır. Üç sınıflandırma tekniği, bankalarla ilgili seçilmiş likidite, gelir-gider, karlılık ve faaliyet oranları kullanılarak karşılaştırılmaktadır. Araştırmanın sonuçları, bankaların sınıflandırmasında karar ağacı modelinin geleneksel diskriminant ve lojistik regresyon modellerine üstünlük sağlayarak alternatif etkili bir sınıflandırma tekniği olarak kullanılabileceğini göstermiştir (Albayrak ve Yılmaz; Albayrak, 2009).

H. Ali Ata ve İbrahim H. Seyrek tarafından 2009 yılında gerçekleştirilen bu çalışmada, denetçiler tarafından yaygın olarak bilinmeyen bazı veri madenciliği teknikleri, finansal tablolardaki hileleri tespit etmeye yardımcı olmak üzere kullanılmıştır. Çalışma İMKB’de işlem gören ve imalat sektöründe faaliyet gösteren 100 firmanın bilgilerine dayalı olarak gerçekleştirilmiştir. Araştırma sonucunda kaldıraç oranı ve aktif karlılık oranının finansal tablo hilesini tespit etmede önemli finansal oranlar olduğu belirlenmiştir (Ata ve Seyrek, 2009).

İpek Savaşçı ve Rezan Tatlıdil tarafından 2006 yılında müşteri ilişkileri yönetimi üzerine bir çalışma gerçekleştirilmiştir. Bu çalışmada bireysel bankacılık alanında uygulanan müşteri ilişkileri yönetim süreci incelenmiş ve müşteri sadakatinin yaratılmasını sağlayan kredi kartlarında uygulanan CRM stratejileri değerlendirilmiştir (Savaşçı ve Tatlıdil, 2006).

#### **3.4. Eğitim Alanında Gerçekleştirilen Veri Madenciliği Uygulamaları**

Konya Selçuk Üniversitesi'nde Onur İnan(2003) tarafından, hazırlık sınıfı, birinci sınıf ve mezun durumunda olan öğrenciler üzerinde, üniversite veri tabanındaki veriler kullanılarak; öğrencilerin başarılarını etkileyen etmenler, başarı düzeyleri, üniversiteyi kazanan öğrenci portföyleri ve mezun olamayan öğrencilerin okulu bitirmelerini etkileyen etmenler üzerinde çalışmalar gerçekleştirilmiş ve sonuçları yorumlanmıştır (İnan, 2003).

Serdar Çiftçi(2006) tarafından gerçekleştirilen çalışmada, uzaktan eğitime katılan öğrencilerin ders çalışma etkinliklerinin değerlendirilmesi için yapılan anketler ve log dosyaları karşılaştırılarak, sonuçların farklı olup olmadıkları incelenmiştir (Çiftçi,2006). Bu çalışmaya benzer bir çalışma olarak Serdar Savaş ve Nursal Arıcı tarafından 2009 yılında gerçekleştirilen bir çalışmada, web tabanlı uzaktan eğitim için video destekli ve animasyon destekli öğretim modeline uygun iki farklı öğretim materyali, bu materyallerin öğrenci başarısı üzerindeki etkilerinin incelenmesi için hazırlanmıştır. Analiz sonucunda video destekli öğretim materyallerinin animasyon destekli öğretim materyallerine göre öğrenci başarısını daha olumlu etkilediği belirlenmiştir (Savaş ve Arıcı, 2009).

Y. Ziya Ayık, Abdülkadir Özdemir ve Uğur Yavuz tarafından yapılan çalışmada, Atatürk Üniversitesi öğrencilerinin mezun oldukları lise türleri ve lise mezuniyet dereceleri ile kazandıkları fakülteler arasındaki ilişki, veri madenciliği teknikleri kullanılarak incelenmiştir (Ayık vd., 2007).

Ahmet Selman Bozkır, Ebru Sezer ve Bilge Gök (2009) tarafından gerçekleştirilen bir çalışmada, ÖSYM tarafından 2008 ÖSS adayları için resmi internet sitesi üzerinden yapılan anket verileri üzerinde veri madenciliği yöntemleri kullanılarak, öğrencilerin başarılarını etkileyen faktörler araştırılmıştır. Bu çalışmada, veri madenciliği yöntemlerinden karar ağaçları ve kümeleme kullanılmıştır (Bozkır vd., 2009). Buna benzer bir çalışma olarak Şenol Zafer Erdoğan ve Mehpere Timor tarafından 2005 yılında gerçekleştirilen bir çalışmada, öğrencilerin üniversite giriş sınavı sonuçları ve öğrencilerin başarıları arasındaki ilişki, kümeleme analizi ve k means algoritması teknikleri uygulanarak incelenmiştir (Erdoğan ve Timor, 2005). Bu çalışmanın KPSS'ye uygulanmış bir modeline benzeyen çalışmayı Hüseyin Özçınar 2006 yılında gerçekleştirmiştir. Frekans analizi ve regresyon analizi yöntemleri kullanılarak derslere ve yıllara göre verinin özellikleri incelenmiştir. Oluşturulan regresyon modeli ile KPSS sonuçlarının değişimi üzerinde anlamlı katkısı olan değişkenler incelenmiş ve oluşturulan modellerin tahmin doğrulukları,

ortalama mutlak hata ve ortalama hata kareler kökü değerleri kullanılarak karşılaştırılmıştır (Özçınar, 2006).

Ahmet Selman Bozkır ve Ebru Sezer tarafından 2009 yılında gerçekleştirilen başka bir çalışmada ise Hacettepe Üniversitesi Beytepe Kampüsü'ndeki öğrenci ve çalışanların, gıda tüketim desenleri incelenmiştir. Çalışmada, karar ağaçları ve birliktelik kuralları uygulanmıştır ve çalışma sonunda %80 başarıyla, gıda tüketim deseninin ortaya çıkarıldığı görülmüştür (Bozkır ve Sezer, 2009).

Hidayet Takçı ve İbrahim Soğukpınar tarafından 2002'de gerçekleştirilen bir çalışmada kütüphane sitesi web günlüklerine dayalı olarak kütüphane kullanıcılarının erişim örüntüleri bulunmaya çalışılmıştır. Bu çalışma yapılırken istatistiksel yöntemler kullanılmıştır (Takçı ve Soğukpınar, 2002).

Murat Kayri tarafından 2008 yılında gerçekleştirilen bir çalışmada, öğrencilerin performans göstergelerinin sürekli izlenebilmesi ve ürünler arasındaki örüntünün bilgisayar sistemleri tarafından oldukça kolay yapılabildiği e-portfolyo değerlendirmeleri için veri madenciliğinde kullanılan yöntemlerin alternatif bir ölçme yaklaşımı olarak kullanımı önerilmektedir (Kayri, 2008).

### **3.5. Ticari Alanda Gerçekleştirilen Veri Madenciliği Uygulamaları**

Anarberk Kalikov(2006) tarafından, bir yayınevi firmasının internet sitesindeki veriler dikkate alınarak, veri madenciliği birliktelik kuralları tekniği ile sepet ve sipariş tabloları incelenmiştir. Hangi ürünlerin kategorisinin değiştirilmesi gerektiği, kullanıcıların meslek ve ilgi alanı dağılımları, müşteri ilgi alanlarına göre satış grafikleri ve kullanıcıların ödeme seçenekleri ile ilgili bir veri madenciliği uygulaması gerçekleştirilmiştir (Kalikov, 2006).

Sinem Akbulut(2006) tarafından yapılan çalışma, bir kozmetik markasının müşteri gruplarını ve ayrılma eğilimi gösteren müşteri kesitini belirleyerek; bu müşterilere özel pazarlama stratejileri geliştirilmesini hedeflemektedir. Bölümlenme için kümeleme teknikleri, ayrılacak müşteri kesitini belirlemek için sınıflama teknikleri kullanılmıştır (Akbulut, 2006).

Feridun Cemal Özçakır ve A. Yılmaz Çamurcu (2007) tarafından gerçekleştirilen bir çalışmada, bir firmanın pastane satış verileri üzerinde veri madenciliği uygulamak için birliktelik kuralları ile bir yazılım tasarlanmıştır. Genelde aynı ürün grubuna ait ürünlerin, en sık birlikte satın alınan ürünler olduğu görülmüştür (Özçakır ve Çamurcu, 2007).

Feyza Gürbüz, Lale Özbakır ve Hüseyin Yapıcı(2008) tarafından gerçekleştirilen başka bir çalışmada, Türkiye'de bir hava yolu işletmesinin parça söküm raporları üzerinde veri madenciliği çalışması gerçekleştirilmiştir. Çalışmanın amacı, uçaklarda kullanılan parçaların, herhangi bir arıza oluşmadan önce düzeltici ve önleyici işlemlerin yapılması için ikaz seviyelerinin tespit edilmesine yönelik kural



geliştirmektir. Sonuç olarak parçaların ikaz seviyelerini temsil edecek anlamlı bir kural elde edilmiş ve bulunan kurallar doğrulukları ve güvenilirlikleri bakımından test edilmiştir (Gürbüz vd., 2009).

Mehmet Aydın Ulaş tarafından 2001 yılında yapılan bir yüksek lisans çalışmasında, sepet analizi gerçekleştirilmiştir. Büyük süpermarket zinciri olan Gima Türk A.Ş.'nin verileri üzerine Apriori algoritması uygulanmış ve ortaya çıkan sonuçlar incelenmiştir. Ayrıca mal satışları arasındaki ilişkileri bulmak amacıyla da, bileşen analizi ve k-ortalama öbeklemesi metotları kullanılmıştır (Ulaş, 2001).

Çağatan Taşkın ve Gül Gökay Emel tarafından 2010 yılında veri madenciliğinde kümeleme yaklaşımları ve Kohonen ağları ile perakendecilik sektöründe bir uygulama gerçekleştirilmiştir. Bu uygulamada, bir perakende işletmenin müşterilerinin Kohonen ağları ile kümelenebileceği ele alınmıştır. Kümeleme analizinin amacı; ele alınan işletmeye, pazar bölümlendirmesi ve hedef pazar seçimi gibi stratejik pazarlama kararlarında yardımcı olması için önceden bilinmeyen kritik müşteri özellikleri ve önem derecelerini de ortaya çıkararak gerekli öngörüleri sağlamaktır (Taşkın ve Emel, 2010).

Fatma Güntürkün 2007 yılında işletmelerin kalite iyileştirmelerini araştıran bir yüksek lisans çalışması hazırlamıştır. Ayrıca bu çalışmada, sürücü koltuğu kalitesi için müşteri memnuniyeti verisi analiz edilmiştir. Müşterinin sürücü koltuğundan memnuniyetini etkileyen en önemli değişkenlerin belirlenmesi için karar ağaçları yaklaşımı uygulanmıştır. Bu uygulamadan elde edilen sonuçlar diğer bir çalışmada aynı veri kümesine uygulanmış ve lojistik regresyon analizinden elde edilen sonuçlarla karşılaştırılmıştır (Güntürkün, 2007).

### **3.6. Telekomünikasyon Alanında Gerçekleştirilen Veri Madenciliği Uygulamaları**

Umman Tuğba Şimşek Gürsoy tarafından 2010 yılında Türkiye’de telekomünikasyon sektöründe faaliyet gösteren büyük bir firmanın, ayrılma eğilimi gösteren müşterileri belirlenerek; bu müşterilere özel pazarlama stratejileri geliştirilmesi hedeflenmiştir. Ayrılacak müşteri profilini belirlemek için Lojistik Regresyon Analizi ve sınıflandırma tekniklerinden Karar Ağaçları kullanılmış ve uygulamanın sonuçları sunulmuştur (Gürsoy, 2010).

Selman Bozkır, S. Güzin Mazman ve Ebru Akçapınar Sezer tarafından 2010 yılında sosyal ağ kullanımına yönelik bir çalışma gerçekleştirilmiştir. Bu çalışmada güncel sosyal paylaşım sitesi facebook üzerinde kullanıcı şablonları incelenmiştir. Facebook kullanım süresi ve erişim sıklığı 570 facebook kullanıcısı üzerinde incelenerek sonuçları ortaya konmuştur (Bozkır vd., 2010).

## **4. SONUÇLAR**

Tüm dünyada olduğu gibi ülkemizde de veri madenciliğine verilen önem ve gösterilen ilgi her geçen yıl artmaktadır. Veri madenciliğinin kullanım alanları genişleyerek yayılmaktadır. Bu çalışmada Türkiye’de yapılan veri madenciliği uygulamaları incelenmiş ve geçmişten günümüze kadar gerçekleştirilen veri madenciliği çalışmaları anlatılmıştır.

Türkiye’de gerçekleştirilen veri madenciliği çalışmaları, eğitim, ticari, mühendislik, bankacılık, borsa, tıp ve telekom olmak üzere, kullanım alanlarına ayrılmış ve her alanda gerçekleştirilen uygulamalar kendi içinde değerlendirilmiştir. Buna göre eğitim alanında gerçekleştirilen analizlerin çoğu öğrenci başarısı üzerine analizler gerçekleştirmek için yapılmıştır. Bu alanda gerçekleştirilen analiz uygulamalarının, sonraki nesiller için öngörü oluşturmak adına kullanılması, eğitim faaliyetlerine çok yararlı olduğu ve olacağı düşünülmektedir. Ticari alanda gerçekleştirilen uygulamaların tamamına yakını müşteri analizi ve pazar analizi ile ilgili olmuştur. Gerçekleştirilen bu çalışmalar sayesinde firmaların yeni pazarlar elde etmesi, mevcut pazarları koruması ve geliştirmesi, müşteri memnuniyeti, yeni müşteri kazanma ve var olan müşteriyi koruma gibi bilgileri sağlama amaçlanmaktadır. Mühendislik alanında gerçekleştirilen veri madenciliği uygulamalarının çoğu, yeni algoritmalar ortaya çıkarma veya var olan algoritma ve teknikleri geliştirme yönünde olmuştur. Bunun sonucu olarak da kullanılan alanla ilgili daha uygun algoritmalar ve var olan algoritmaların türevleri ortaya çıkmıştır. Ancak geliştirilen algoritmaların sadece belirli algoritmalar üzerinde olması bu alanda bir eksi olarak görülmüştür. Bankacılık ve borsa alanında gerçekleştirilen çalışmalar daha çok tahmin gerçekleştirmek amaçlı yapılmıştır. Mevcut müşteri portföyünden, gelecekte karşılaşılabilecek kredi vb. mali konularda risk analizleri gerçekleştirilmiştir. Ayrıca şirket profillerinin incelenmesi ve hisse senetleri üzerine de araştırmalar gerçekleştirilmiştir. Veri madenciliğinin tahmin yönünün en etkili kullanıldığı alanlardan bir tanesi olarak bankacılık ve borsa alanını göstermek mümkündür. Tıp alanında gerçekleştirilen uygulamaların hastalık belirtileri ve var olan belirtilerden şablon ortaya çıkarma amacıyla yapıldığı görülmüştür. Ülkemizde veri madenciliği çalışmalarının tıp alanında çok daha çeşitli ve etkili kullanılması gerektiği, bu konuda eksikliğin olduğu görülmüştür. Son olarak telekom alanında yapılan çalışmaların müşteri ve kullanıcı profili analizleri için gerçekleştirildiği ortaya çıkmıştır.

Dünyadaki teknolojik gelişmelere paralel olarak ülkemizde de veri madenciliği uygulamaları gittikçe artmaktadır. Ancak incelenen çalışmalar da göstermektedir ki kurum ve kuruluşların çoğu müşteri/kullanıcı analizlerine yönelmiştir. Bunun yanı sıra kurum ve kuruluşların kendi bünyelerinde veri madenciliğini kullanarak, gelişmelerini buna dayalı gerçekleştirmeleri faydalarına olacaktır. Ayrıca veri madenciliğinin kullanıldığı alanların çeşitlendirilmesi de, gerek ülkemiz kurum ve kuruluşlarına, gerekse bu kurum ve kuruluşlardan hizmet ve ürün alan ülkemiz insanlarına büyük faydalar sağlayacaktır.

**KAYNAKLAR (REFERENCES)**

**Adriaans, P. ve Zantinge, D.**, (1997), *Data Mining*, Boston, MA, USA Addison Wesley Longman Publishing.

**Akbulut, S.**, (2006) *Veri Madenciliği Teknikleri ile Bir Kozmetik Markanın Ayrılan Müşteri Analizi Ve Müşteri Segmentasyonu*, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.

**Akgöbek, Ö. ve Çakır, F.**, (2009), “Veri Madenciliğinde Bir Uzman Sistem Tasarımı”, *Akademik Bilişim 09*, 11-13 Şubat Harran Üniversitesi, Şanlıurfa, 801-806.

**Akpınar, H.**, (2000), “Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği”. *İ.Ü. İşletme Fakültesi Dergisi*, Cilt 29, S 1, 1-22.

**Aksoy, B.**, (2009), *Cluster Analysis Of Decompression Illness*, Galatasaray University, Institute of Science and Engineering.

**Albayrak, A.S. ve Yılmaz, Ş.K.**, (2009), “Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama”, *S.D.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 14, No 1, 31-52.

**Albayrak, A.S.**, (2009), “Classification of Domestic and Foreign Commercial Banks in Turkey Based On Financial Efficiency: A Comparison of Decision Tree, Logistic Regression and Discriminant Analysis Models”, *S.D.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 14, No 2, 113-139.

**Albayrak, M.**, (2008), *EEG Sinyallerindeki Epileptiform Aktivitenin Veri Madenciliği Süreci ile Tespiti*, Doktora Tezi, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü.

**Ata, A.H. ve Seyrek, İ.H.**, (2009), “The Use of Data Mining Techniques in Detecting Fraudulent Financial Statements: An Application on Manufacturing Firms”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, Cilt 14, No 2, 157-170.

**Ata, N., Özkök, E. ve Karabey, U.**, ( 2008), “Survival Data Mining: An Application To Credit Card Holders”, *Sigma Mühendislik ve Fen Bilimleri Dergisi*, Cilt 26, No 1, 33-42.

**Ayık, Y.Z., Özdemir, A. ve Yavuz, U.**, (2007), “Lise Türü ve Lise Mezuniyet Başarısının Kazanılan Fakülte ile İlişkinin Veri Madenciliği Tekniği ile Analizi”, *Sosyal Bilimler Enstitüsü Dergisi*, Cilt 10, No 2.

**Baloğlu, U.B.**, (2006), DNA Sıralarındaki Tekrarlı Örüntülerin ve Potansiyel Motiflerin Veri Madenciliği Yöntemiyle Çıkarılması, Fırat Üniversitesi, Fen Bilimleri Enstitüsü.

**Bilgin, T.T. ve Çamurcu, A.Y.**, (2008), “Çok Boyutlu Veri Görselleştirme Teknikleri”, Akademik Bilişim 2008, 30 Ocak - 01 Şubat. Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, 107-112.

**Bilgin, T.T.**, (2009), “Veri Akışı Diyagramları Tabanlı Veri Madenciliği Araçları ve Yazılım Geliştirme Ortamları”, Akademik Bilişim 09, 11-13 Şubat, Harran Üniversitesi, Şanlıurfa, 807-814.

**Bozkır, A.S. ve Sezer, E.**, (2009), “Usage of Data Mining Techniques in Discovering The Food Consumption Patterns of Students and Employees of University”, Balkan-Kafkas ve Türk Devletleri Uluslararası Mühendislik Sempozyumu, 22-24 October, Isparta, 104-109.

**Bozkır, A.S., Mazman, S.G., ve Sezer, E.A.**, (2010), “Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case”, 2nd International Symposium on Information Management in a Changing World", 22-24 September, Hacettepe University, Ankara, 145-152.

**Bozkır, A.S., Sezer, E. ve Gök, B.**, (2009), “Öğrenci Seçme Sınavında (ÖSS) Öğrenci Başarımını Etkileyen Faktörlerin Veri Madenciliği Yöntemleriyle Tespiti”, 5. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09), 13-15 Mayıs, Karabük Üniversitesi, Karabük, 37-43.

**Çalışkan, S.K. ve Soğukpınar, İ.**, (2008), "KxKNN: K-Means ve K En Yakın Komşu Yöntemleri ile Ağlarda Nüfuz Tespiti", 2. Ağ ve Bilgi Güvenliği Sempozyumu, 16-18 Mayıs, Girne, 120-124.

**Çiftci, S.**, (2006), Uzaktan Eğitimde Öğrencilerin Ders Çalışma Etkinliklerinin Log Verilerinin Analiz Edilerek İncelenmesi, Yüksek Lisans Tezi, Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü.

**Danacı, M., Çelik, M. ve Akkaya, A.E.**, (2010), "Veri Madenciliği Yöntemleri Kullanılarak Meme Kanseri Hücrelerinin Tahmin ve Teşhisi", Akıllı Sistemlerde Yenilikler ve Uygulama Sempozyumu, 21-24 Haz. 2010, Kayseri, 9-12.

**Doğan, Ş. ve Türkoğlu, İ.**, (2008), “Iron-Deficiency Anemia Detection From Hematology Parameters By Using Decision Trees”, International Journal of Science & Technology, Cilt 3, No 1, 85-92.

**Doğan, Ş., ve Türkoğlu, İ.**, (2007), " Hypothyroidi and Hyperthyroidi Detection from Thyroid Hormone Parameters by Using Decision Trees", Doğu Anadolu Bölgesi Araştırmaları Dergisi, Cilt 5, No 2, 163-169.

**Doğan, Y.**, (2004), A Data Mining Based Classification Algorithm for Tactical Underwater Sensor Networks, Yüksek Lisans Tezi, Turkish Naval Academy, Computer Engineering.

**Duru, N. ve Canbay, M.**, (2007), “Veri Madenciliği ile Deprem Verilerinin Analizi”, International Earthquake Symposium, Kocaeli, 556-560.

**Erdem, C.**, (2006), Density Based Clustering Using Mathematical Morphology, Yüksek Lisans Tezi, Middle East Technical University, Information Systems.

**Erdoğan, Ş.Z. ve Timor, M.**, (2005), “A Data Mining Application In A Student Database”, Journal Of Aeronautics and Space Technologies, Cilt 2, No 2, 53-57.

**Güntürkün, F.**, (2007), A Comprehensive Review Of Data Mining Applications In Quality Improvement And A Case Study, Yüksek Lisans Tezi, Middle East Technical University, Statistics.

**Gürbüz, F., Özbakır, L. ve Yapıcı, H.**, (2009), “Türkiye’de Bir Havayolu İşletmesine Ait Parça Söküm Raporlarına İlişkin Veri Madenciliği Uygulaması”, Gazi Üniversitesi Mimarlık Mühendislik Fakültesi Dergisi, Cilt 24, No 1, 73-78.

**Gürsoy, U.T.Ş.**, (2010), “Customer Churn Analysis in Telecommunication Sector”, İstanbul University Journal of the School of Business Administration, Cilt 39, No 1, 35-49.

**Hand, D.J.**, (1998), “Data Mining: Statistics and More?”, The American Statistician, Cilt 52, 112-118.

**İnan, A.**, Privacy Preserving Distributed Spatio-Temporal Data Mining, Yüksek Lisans Tezi, Sabancı University, Computer Science and Engineering, 2006.

**İnan, O.**, (2003), Veri Madenciliği, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.

**Jacobs, P.**, (1999), “Data Mining: What General Managers Need to Know”, Harvard Management Update, Cilt 4, No 10, 8.

**Kalkov, A.**, (2006), Veri Madenciliği ve Bir E-Ticaret Uygulaması, Yüksek Lisans Tezi, Gazi Üniversitesi, Fen Bilimleri Enstitüsü.

**Kastro, Y.**, (2006), A Defect Prediction Method For Software Versioning, Yüksek Lisans Tezi, Boğaziçi University, Computer Engineering.

**Kayaalp, K.**, (2007), Asenkron Motorlarda Veri Madenciliği ile Hata Tespiti, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü.

**Kayri, M.**, (2008), “Elektronik Portfolyo Değerlendirmeleri İçin Veri Madenciliği Yaklaşımı”, Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, Cilt 5, No 1, 98-110.

**Kılınc, Y.**, (2009), Mining Association Rules For Quality Related Data In An Electronics Company, Middle East Technical University, Industrial Engineering.

**Kitler R. ve Wang W.**, (1998), “The Emerging Role of Data Mining”, Solid State Technology, Cilt 42, No 11, 45.

**Özçakır, F.C. ve Çamurcu, A.Y.**, (2007), “Birliktelik Kuralı Yöntemi İçin Bir Veri Madenciliği Yazılımı Tasarımı ve Uygulaması”. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, Yıl 6, No 12, 21-37.

**Özçınar, H.**, (2006), KPSS Sonuçlarının Veri Madenciliği Yöntemleriyle Tahmin Edilmesi, Yüksek Lisans Tezi, Pamukkale Üniversitesi, Fen Bilimleri Enstitüsü.

**Özekes, S.**, (2003), "Data Mining Models and Application Areas", İstanbul Commerce University Journal of Science, No.3, 65-82.

**Piramuthu, S.**, (1998), “Evaluating Feature Selection Methods For Learning in Data Mining Applications”, Thirty-First Annual Hawaii International Conference on System Sciences, IEEE Computer Society, 6-9 January, Kohala Coast Hawaii USA, 294.

**Savaş, S. ve Arıcı, N.**, (2009), Web Tabanlı Uzaktan Eğitimde İki Farklı Öğretim Modelinin Öğrenci Başarısı Üzerindeki Etkilerinin İncelenmesi, 5. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09), 13-15 Mayıs, Karabük Üniversitesi, Karabük, 1229.

**Savaşçı, İ. ve Tatlıdil, R.**, (2006), “Bankaların Kredi Kartı Pazarında Uyguladıkları CRM (Müşteri İlişkiler Yönetimi) Stratejisinin Müşteri Sadakatine Etkisi”, Ege Akademik Bakış Dergisi, Cilt 6, No 1, 62-73.

**Shearer, C.**, (2000), “The Crisp-DM Model: The New Blueprint for Data Mining” Journal of Data Warehousing, Cilt 5 No 4, 13-23.

**Sıramkaya, E.**, (2005), Veri Madenciliğinde Bulanık Mantık Uygulaması, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.

**Takçı, H. ve Soğukpınar, İ.**, (2002), İ., "Kütüphane Kullanıcılarının Erişim Örüntülerinin Keşfi", Bilgi Dünyası, Cilt 3, Sayı 1, 12-26.

**Taşkın, Ç. ve Emel, G.G.**, (2010), “Veri Madenciliğinde Kümeleme Yaklaşımları Ve Kohonen Ağları İle Perakendecilik Sektöründe Bir Uygulama”, Süleyman

Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt 15, No 3, 395-409.

**Thuarisingham, B.M.**, (2003), *Web Data Mining and Applications in Business Intelligence and Counter Terrorism*, CRC Press LLC, Boca Raton, FL,USA.

**Toprak, S.**, (2004), *Data Mining For Rule Discovery in Relational Databases*, Middle East Technical University, Computer Engineering.

**Toprak, S.D.**, (2005), *A New Hybrid Multi-Relational Data Mining Technique*, Yüksek Lisans Tezi, Middle East Technical University, Computer Engineering.

**Ulaş, M.A.**, (2001), *Market Basket Analysis For Data Mining*, Yüksek Lisans Tezi, Boğaziçi University, Computer Engineering.

**Yavaş, G.**, (2003), *Using A Data Mining Approach For The Prediction of User Movements in Mobile Environments*, Yüksek Lisans Tezi, Bilkent University, Institute of Engineering and Science.

**Yıldırım, P.**, (2008), Uludağ, M. ve Görür, A., “Hastane Bilgi Sistemlerinde Veri Madenciliği”, *Akademik Bilişim 2008*, 30 Ocak - 01 Şubat, Çanakkale Onsekiz Mart Üniversitesi, Çanakkale, 429-434.

**Yıldız, B.**, (2010), *Impacts Of Frequent Itemset Hiding Algorithms On Privacy Preserving Data Mining*, İzmir Institute of Technology, Computer Engineering.