



**VERİ MADENCİLİĞİ  
VE  
SERMAYE PİYASALARINA UYGULANMASI**

**DR. ALİ SERHAN KOYUNCUGİL  
İSTATİSTİKÇİ**

## YÖNETİCİ ÖZETİ

Bilgi teknolojilerinin gelişimi ve gündelik hayatın her aşamasında kullanılabilir hale gelmesiyle beraber, her alanda oldukça büyük miktarda veri birikmeye başlamıştır. Zaman içerisinde bu verilerin karar destek amaçlı olarak kullanımının nasıl gerçekleştirilebileceğine yönelik çalışmalar Veri Madenciliği kavramının ortaya çıkmasına yol açmıştır.

Veri madenciliği, büyük hacimde veri içerisindeki gizli kalmış önemli örüntü, bilgi ve ilişkilerin açığa çıkarılıp stratejik karar destek amaçlı kullanımı olarak tanımlanmaktadır. Büyük hacimde veri niteliğinden dolayı, veri madenciliğinin kullanıldığı alanlar başlangıçta; sağlık sektörü, telekomünikasyon, finans ve pazarlama olarak şekillenmiştir. Buna karşın, bugün güvenlik de dahil olmak üzere hemen her alanda gittikçe yaygınlaşan biçimde veri madenciliğinden faydalanılmaktadır.

Finans sektöründe, bankacılık ağırlıklı olarak uygulama yaygınlığına sahip olan veri madenciliğinin sigortacılık ve sermaye piyasası uygulamaları da gün geçtikçe artış göstermektedir.

Veri madenciliğinin sermaye piyasası uygulamalarına bakıldığında, gözetim denetim sistemleri, erken uyarı sistemleri, profillendirme çalışmalarının ağırlık kazandığı gözlenmektedir.

Bu çalışmada, veri madenciliği, kavram, tanım ve kullanılan yöntemler açısından değerlendirilmiş ve Sermaye Piyasası Kurulu çalışmaları açısından;

- Gözetim ve Erken Uyarı Sistemleri,
- Yasal Düzenlemelerin Etki ve Etkinlik Değerlendirmesi,
- Halka Açık Şirketlerin Değerlendirilmesi ve Halka Açıklığın Önündeki Engellerin Belirlenmesi,
- Yatırımcı Profilinin Tespiti,
- Aracı Kuruluş Risk Profillerinin Belirlenmesi,

konularında uygulama örnekleri verilerek değerlendirilmiştir. Çalışma, mümkün olduğunca teknik detaylardan arındırılarak, olabildiğince yalın biçimde sunulmuştur.

## İÇİNDEKİLER

<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. LİTERATÜR TARAMASI .....</b>	<b>3</b>
<b>3. VERİ MADENCİLİĞİ .....</b>	<b>6</b>
3.1. VERİ MADENCİLİĞİNİN TANIMI .....	6
3.2. VERİ MADENCİLİĞİ YÖNTEMLERİNİN SINIFLANDIRILMASI.....	6
3.3. BAŞLICA VERİ MADENCİLİĞİ YÖNTEMLERİ.....	8
<b>4. VERİ MADENCİLİĞİNİN SERMAVE PİYASALARI ÜZERİNE UYGULAMA     ÖNERİLERİ .....</b>	<b>10</b>
4.1. GÖZETİM VE ERKEN UYARI SİSTEMLERİ .....	10
4.2. YASAL DÜZENLEMELERİN ETKİ VE ETKİNLİK DEĞERLENDİRMESİ .....	12
4.3. HALKA AÇIK ŞİRKETLERİN DEĞERLENDİRİLMESİ VE HALKA AÇIKLIĞIN ÖNÜNDEKİ ENGELLERİN BELİRLENMESİ .....	12
4.4. YATIRIMCI PROFİLİNİN TESPİTİ .....	13
4.5. ARACI KURULUŞ RİSK PROFİLLERİNİN BELİRLENMESİ.....	15
<b>5. SONUÇ.....</b>	<b>16</b>
<b>6. KAYNAKLAR.....</b>	<b>17</b>

# VERİ MADENCİLİĞİ

## VE

### SERMAYE PİYASALARINA UYGULANMASI

#### 1. GİRİŞ

Sadece sayım gibi büyük ölçekli çalışmalarda kullanılan 1950'lerin devasa boyutlu ilk bilgisayarları; günümüze gelindiğinde artık avuç içine sığacak derecede küçülmüş ve günlük hayatımızın bir parçası haline gelmiştir. Neredeyse günlük yaşamın akışındaki tüm iş ve işlemler bilgisayarla veya bilgisayar destekli olarak gerçekleştirilmeye doğru ilerlemektedir. Bu açıdan bakıldığında, küçük bir işletmenin dahi günlük işlem verileri zaman içerisinde katlanarak büyümekte ve fiyatları anlamlı derecede düşen sabit diskler sayesinde saklanabilmektedir.

1990'lara kadar, yaygın olarak 'Dijital Hafıza' çağrışımı yapan bilgisayarlar, geçmiş işlemlere ait verileri gerektiğinde sunmakta ve bir takım hesaplamalara izin vermektedir. 1990'lardan itibaren ise bilgisayarlarda, çoğunlukla da veri tabanlarında, depolanan veri yığınlarının nasıl kullanılabilceği sorgulanmaya başlamış; nihayetinde yığınla verinin içerisinde gözden kaçan ve keşfedilmeyi bekleyen bilgiler olduğu düşüncesi hakim olmuştur. Bu noktadan itibaren veri yığınları arasında sıkışıp kalmış bilgilerin nasıl açığa çıkarılabileceği üzerine yoğunlaşmış ve Veri Tabanlarında Bilgi Keşfi (VTBK, Knowledge Discovery in Databases - KDD) kavramı ortaya çıkmıştır.

VTBK kavramının oluşumunda, parça parça, farklı format ve ortamlarda verinin ayrı ayrı değerlendirilmesi yerine; biraraya getirilerek, aralarında doğrudan ilişki bulunması düşünülmemeyen değişkenlerin (alanların), dolaylı birtakım ilişkileri ima edebileceği mantığından hareket edilmiştir. Veri madenciliği ise VTBK'nin özünü oluşturan keşif kısmının gerçekleştiği bir adım olarak alınabileceği gibi bağımsız bir süreç olarak da değerlendirilebilmektedir.

Veri madenciliğinin yeni gelişen bir alan olması nedeniyle, farklı kaynaklarda farklı tanımlar yapıldığı görülmektedir. Buna karşın, veri madenciliği genel anlamda; büyük miktarda veri içerisinden, gizli kalmış, değerli, kullanılabilir bilgilerin açığa çıkarılması biçiminde ifade edilebilir. Veri içerisinde keşfi bekleyen bilgilerin varlığı düşüncesi sayesinde, İstatistik farklı bir bakış açısıyla yeniden keşfedilmiş; zaman içerisinde Makine

Öğrenimi ve Yapay Zeka desteğiyle neredeyse düşünebilen Akıllı Algoritmalar ile İstatistiğin birleşimi, Veri Madenciliği'ni meydana getirmiştir. Şüphesiz ki veri madenciliğini bilişimden ayrı tutarak salt istatistik olarak düşünmek çok doğru bir yaklaşım olmayacaktır. Temel anlamda, verinin yer aldığı veri tabanları ve özel bir şekli olan veri ambarları, yapay zeka ve makine öğrenimine dayalı yöntemler de söz konusudur. Veri madenciliği ile Klasik İstatistik arasında belli noktalarda ayrışmalar söz konusu olsa da, içerisinde istatistiği telaffuz etmeden veri madenciliğinden bahsetmek mümkün olmamaktadır. Öyleki, bu konu en iyi biçimde 'İstatistikçiler, Veri Madenciliğini elle yaparlar.' ifadesiyle özetlenebilir.

Veri madenciliğinin temel malzemesinin büyük hacimde veri olduğu dikkate alındığında; veri madenciliğinin uygulama alanlarının başında finans kesimi gelmektedir. Özellikle bankacılık kesimi 2000'lerin başından itibaren Müşteri İlişkileri Yönetimi, Müşteri Profillendirmesi, Kampanya Yönetimi, Kredi derecelendirme, Risk Yönetimi konularında yaygın olarak veri madenciliğinden yararlanmaktadır. Sigorta kesimi ise daha çok Hilekarlık Tespiti yönünden veri madenciliği ile ilgilenmektedir. Sermaye Piyasaları açısından bakıldığında ise düzenleyici otoritelerin daha çok sermaye piyasası suçları ve düzenleme ihlalleri tespiti açısından veri madenciliğinden faydalandığı gözlenmektedir.

Bu çalışmada, veri madenciliğinin Sermaye Piyasası Kurulu açısından hangi alanlarda kullanılabileceği ortaya koyulmaktadır.

## 2. LİTERATÜR TARAMASI

Teknolojinin bir sonucu olan bilgi sistemlerinin etkinliği ve özeldede veri tabanlarının yaygın kullanımı sonucunda ulaşılan büyük hacimde veri, Veri Tabanlarında Bilgi Keşfi (VTBK) kavramını ortaya çıkarmıştır. VTBK ifadesi ilk defa Piatetsky-Shapiro (1991) tarafından 1989 yılında gerçekleştirilen ilk VTBK çalışma grubu toplantısında kullanılmış, konuyla ilgili kavram ve tanımlar ortaya konulmuştur. Ayrıca, Veri Madenciliği (VM) terimi de VTBK'nin bir bileşeni olarak tanımlanmıştır. Tanımlar bir taraftan tartışıldururken, veri madenciliği yöntemleri geliştirmeye yönelik çalışmalar da devam etmiş, Agrawal vd. (1996) nicel birliktelik kurallarının madenciliği için hızlı bir algoritma olan Apriori'yi önermişlerdir. VTBK'nin temel süreçleri üzerinde hiyerarşi arayışının sonucunda, Fayyad vd. (1996) veri tabanlarında bilgi keşfinin süreçlerine ve veri madenciliğinin bu süreçteki yerine yönelik bir akış sunmuşlardır. Ayrıca, veri madenciliğinin temel özelliklerini irdelemişlerdir. İş zekasının en yeni ve popüler bileşeni olan veri madenciliğinin yerini belirlemek amacıyla, Cabena vd. (1997) veri madenciliğinin diğer iş zekası çözümleriyle mukayesesini yapıp, aralarındaki hiyerarşiyi ortaya koymuşlardır. VM'nin uygulama yaygınlığı kazanmasıyla birlikte uygulamaya yönelik çalışmalar ağırlık kazanmaya başlamış ve Berson vd. (1999) VM'nin en yaygın kullanıma sahip alanı olan müşteri ilişkileri yönetimi kapsamında veri madenciliği yöntemleri ve uygulamalarına yer vermişlerdir. Ayrıca, yeni nesil yöntemlerden olan Karar ağacı yöntemlerinin başlıcalarından CART ve CHAID'i incelemişlerdir. Şüphesiz ki veri madenciliğinin vazgeçilmez bileşenlerinden birisi de kullanılan yazılımlardır. VM'nin ticari anlamda kazandığı yaygınlığa paralel olarak bilimsel gelişimini de sağlayabilmesi için akademik amaçlı çeşitli VM yazılımları geliştirilmiş ve Witten ve Frank (2000) veri madenciliği yazılımı WEKA üzerinden veri madenciliği yöntemlerinin uygulanmasına değinmişlerdir.

VM'yi bilişim ve istatistik olmak üzere iki bakış açısıyla incelemek mümkündür. Hastie vd. (2001), veri madenciliğine istatistiksel perspektiften yaklaşmış ve veri madenciliği ile istatistik arasındaki ilişkiyi 'İstatistiksel Öğrenme' kavramı ile kurmuşlardır. 1990'lar veri madenciliğinin bilişim açısından gelişim dönemi olurken 2000'li yıllardan itibaren VM'nin analitik tarafı ağırlık kazanmaya başlamış; istatistiksel yöntemlerin büyük veri setlerine uygulamaları üzerine çalışmalar ağırlık kazanmıştır. Rao (2001), istatistiğin geçmişi ve gelecek vizyonuna değinip, veri madenciliğine yönelimi istatistiksel açıdan değerlendirmiştir. Veri madenciliğinin gelişimini sürdüren bir alan olması nedeniyle; diğer analitik yöntemlerle arasındaki ayrımın belirginleştirilmesi önem kazanmıştır. Bu noktadan hareketle, Roiger ve

Geatz (2002), veri madenciliği ile SQL sorgusu ve OLAP arasındaki ayrımı ortaya koyup; bilgi ihtiyacının düzeyine göre uygulanması gereken yöneme yönelik bir çerçeve çizmişlerdir. Moss ve Atre (2003) veri madenciliği ile istatistiksel analiz yöntemlerini karşılaştırmış ve aralarındaki farklılıkları ortaya koymuşlardır. Koyuncugil (2004) veri madenciliğinin temel prensiplerinin arkasındaki istatistiksel mantığa vurgu yapmıştır. Ayrıca, Chen (2001) veri madenciliği ile belirsiz muhakeme teknikleri arasındaki ilişkiyi ve bağlantıları ortaya koymuş, bulanık teorinin veri madenciliği ile kesişimine yer vermiştir.

Hisse senetleri, döviz kurlarına ve işletme iflaslarının tahmini, finansal risk yönetimi ve belirlenmesi, borç yönetimi, müşteri profillerinin belirlenmesi ve kara para aklama analizleri veri madenciliğinin finansal çalışmalarda kullanıldığı temel alanlardır (Kovalerchuk ve Vityaev, 2002). Veri madenciliğinin finansal araştırmalarda kullanıldığı bir çok çalışma olmakla beraber veri madenciliğinin finansal çalışmalarda başarılı olarak uygulandığı araştırmalara; hilekarlığın, işletme iflas ve başarısızlıklarının belirlendiği Tam ve Kiang (1992), Lee vd. (1996) ve Kumar ve diğerlerinin (1997) çalışmaları, stratejik finansal karar alma uygulamalarında Nazem ve Shin'in (1999) çalışması, pazar uygulamalarında Bracham vd'nin (1996), finansal performans uygulamalarında Eklund vd. (2003), Hoppszallern (2003), Derby (2003), Chang vd. (2003), Lansiluoto vd. (2004), Kloptchenko vd. (2004) ve Magnusson ve diğerlerinin (2005) çalışmaları örnek olarak gösterilebilir. Koyuncugil (2006) ise Bulanık Veri Madenciliği ile Hisse Senetleri Piyasası'nda işlem manipülasyonu tespitine yönelik bir erken uyarı sistemi tasarlamış ve başarıyla çalıştığını tespit etmiştir.

Koyuncugil ve Özgülbaş (2006a) Türkiye'deki KOBİ'lerin finansal sorunlarını vurgulayarak, KOBİ Borsaları da dahil olmak üzere çözüm önerilerini irdelemişler ve sorunları çözmek için atılacak ilk adımın KOBİ'lerin finansal profillerinin belirlenmesi olduğunu ifade etmişlerdir. Yazarlar, İstanbul Menkul Kıymetler Borsasında (İMKB) işlem gören 135 KOBİ'nin 2004 yılı verilerini kullanarak Veri Madenciliği yöntemi ile finansal profillerini belirlemiştir. Koyuncugil ve Özgülbaş'ın (2006b) aynı veri seti ile gerçekleştirdikleri başka bir çalışmada da KOBİ'ler için bir finansal performans ölçütü belirlenmesi hedeflenmiştir. Koyuncugil ve Özgülbaş (2006c) tarafından yine İMKB'de 2000-2005 yılları arasında işlem gören KOBİ'lerde yürütülen çalışmada, KOBİ'lerin finansal başarısızlığa etki eden finansal faktörlerin belirlenmesi amaçlanmıştır. Veri Madenciliği yöntemlerinden CHAID (Chi-Square Automatic Interaction Detector) Karar Ağacı Algoritması ile gerçekleştirilen çalışma sonucunda KOBİ'lerin finansal başarısızlığında etkili olan faktörler belirlenmiştir. Özgülbaş ve Koyuncugil (2006) ve Özgülbaş vd. (2006)

tarafından aynı veri seti ile yapılan farklı çalışmalarda KOBİ'lerin finansal anlamda güçlü ve zayıf yönlerinin yanısıra, finansal performans durumu ile KOBİ'lerin büyüklüğü arasındaki ilişki de belirlenmiştir.



### 3. VERİ MADENCİLİĞİ

#### 3.1. Veri Madenciliğinin Tanımı

Veritabanları günümüzde terabaytlarla ifade edilmektedir. Bu büyük hacimde verinin içinde stratejik önem taşıyan gizli enformasyon yatmaktadır. Temel soru, bu gizli kalmış enformasyonun nasıl açığa çıkarılacağıdır. Bu sorunun en güncel ve popüler yanıtı Veri Madenciliği (VM) olarak karşımıza çıkmaktadır.

Veri madenciliği veri kümesi içerisinde keşfedilmemiş örüntüleri bulmayı hedefleyen teknikler koleksiyonunu betimlemektedir. Veri madenciliğinin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar-verme modelleri yaratmaktır. Veri madenciliği, William Frawley ve Gregory Piatetsky-Shapiro (1991) tarafından, ‘ ... verideki gizli, önceden bilinmeyen ve potansiyel olarak faydalı enformasyonun önemsiz olmayanlarının açığa çıkarılması...’ biçiminde yapılan bilgi keşfi tanımını destekler (Berson vd. 1999).

Veri madenciliği, pek çok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları geçerli tahminler yapmak için kullanan bir süreçtir (TWO CROWS 1999).

#### 3.2. Veri Madenciliği Yöntemlerinin Sınıflandırılması

Veri Madenciliği yöntemlerini,

- Denetimli,
- Denetimsiz,

olmak üzere iki ana kategoriye ayırmak mümkündür. Denetimli ve denetimsiz yöntemler için genel kabul görmüş tanımlar aşağıda verilmiştir.

**Denetimli (Supervised):** İyi tanımlanmış veya kesin bir hedef olduğunda denetimli (supervised) ifadesi kullanılır.

**Denetimsiz (Unsupervised):** Elde edilmesi istenen sonuç için özel bir tanımlama yapılmamışsa veya belirsizlik sözkonusu ise denetimsiz (unsupervised) ifadesi kullanılır (Hastie vd. 2001).

Denetimli ve denetimsiz ifadeleri birbirinin tersine karşılık gelmektedir. Denetimli ve denetimsiz yöntemleri sürecin bütünü açısından değerlendirmek gerekirse;

- Denetimsiz yöntemler daha çok veriyi anlamaya, tanımaya, keşfetmeye yönelik olarak kullanılan ve sonraki uygulanacak yöntemler için fikir vermeyi amaçlamaktadır,
- Denetimli yöntemler ise veriden bilgi ve sonuç çıkarmaya yönelik kullanılmaktadır,

denilebilir.

Bu nedenle denetimsiz bir yöntemle elde edilen bir bilgi veya sonucu, eğer mümkünse denetimli bir yöntemle teyit etmek, elde edilen bulguların doğruluğu ve geçerliliği açısından önem taşımaktadır.

Denetimli ve denetimsiz algoritmaların mantığını ve farklılığını en iyi aktaracak olan yöntem, Kümeleme Analizi'dir.

Aşamalı Kümelemede hem birimlerin hem de değişkenlerin birbiriyle değişik benzerlik ölçütlerine göre kümelenmesi; küme sayısı baştan verilmeden elde edilmektedir. Küme sayısı baştan belirli olmadığı için, Aşamalı Kümeleme Analizi denetimsiz (unsupervised) bir yöntemdir.

Aşamalı Olmayan Kümeleme yöntemlerinden K-Ortalamlar Kümelemede ise birimlerin uygun oldukları kümelerde toplanmaları ve 'n birimin k sayıda kümeye parçalanması' hedeflenmektedir. Küme sayısı baştan belli olduğu için, K-Ortalamlar Kümeleme Analizi denetimli (supervised) bir yöntemdir.

#### **Başlıca Veri Madenciliği yöntemleri denetimli veya denetimsiz olmalarına göre;**

- **Denetimli (Supervised)**
  - En yakın k komşuluk (k-Nearest-Neighbor)
  - K-ortalamlar kümeleme (K-means clustering)
  - Regresyon modelleri (Regression models)
  - Kural çıkarımı (Rule induction)
  - Karar ağaçları (Decision trees)
  - Sinir ağları (Neural networks)
- **Denetimsiz (Unsupervised)**
  - Aşamalı kümeleme ( Hierarchical clustering)
  - Kendi kendini düzenleyen haritalar (Self organized maps)

olarak sınıflandırılabilir (Hastie vd. 2001, Thearling 2005, Koyuncugil 2006).

### 3.3. Başlıca Veri Madenciliği Yöntemleri

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 yöntemlerde genellikle istatistiği temel alan ama daha çok makine öğrenimi ve yapay zeka destekli yeni nesil yöntemlerdir.

Veri madenciliğinde kullanılan klasik yöntemlerin başlıcaları;

- Regresyon
- K - En Yakın Komşuluk
- Kümeleme

olarak sayılabilir (Berson vd. 1999).

Yeni nesil yöntemlerin başlıcaları ise;

- Karar Ağaçları,
- Birliktelik Kuralları,
- Sinir Ağları,

olarak sıralanabilir (Berson vd. 1999).

Ayrıca diğer veri madenciliği yöntemlerinin başlıcaları da;

- Temel Bileşenler Analizi,
- Diskriminant Analizi,
- Faktör Analizi,
- Kohonen Ağları,
- Bulanık Mantığa Dayalı Yöntemler,
- Genetik Algoritmalar,
- Bayesci Ağlar,
- Pürüzlü (Rough) Küme Teorisine Dayalı yöntemler,

olarak sıralanabilir (Chen 2001).

Yukarıda sayılan yöntemlerin dışında birden fazla tekniği içine alan hibrid yöntemler ve zaman serilerine dayalı yöntemlerden de veri madenciliği yöntemi olarak faydalanılmaktadır. Özet olarak, bilgi keşfine yarayan her yöntem veri madenciliği yöntemi olarak kullanılabilir (Kovalerchuk ve Vityaev 2002).

## 4. VERİ MADENCİLİĞİNİN SERMAYE PİYASALARI ÜZERİNE UYGULAMA ÖNERİLERİ

Veri içerisinde gizli kalmış, gözlenemeyen bilgilerin açığa çıkarılıp; stratejik karar destek amaçla kullanılması tanımına sahip veri madenciliği, özellikle karar süreçlerine katkıda bulunmak üzere aşağıda sayılan başlıklarda değerlendirilebilir. Şüphesiz ki, veri madenciliğinin Kurul çalışmalarına katkısı aşağıda sayılan başlıklarla sınırlı değildir. Kurul içerisinde veri üretimi gerçekleşen her alanda veri madenciliğinden faydalanmak mümkündür.

### 4.1. Gözetim ve Erken Uyarı Sistemleri

Menfaat sağlamak üzere kasıtlı olarak menkul kıymetlerin fiyatlarıyla oynamak olarak özetlenebilecek olan manipülasyon suçu Hisse Senetleri Piyasası'nın fiyat mekanizmasını bozmaktadır. Ayrıca, gerek manipülasyon gerekse içerden öğrenenlerin ticareti haksız kazanca yol açtığı için potansiyel yatırımcıların piyasaya katılımının önünü kesmekte ve geniş anlamda düşünüldüğünde piyasanın gelişimini ve derinleşmesini de engellemektedir.

Manipülasyon ve içerden öğrenenlerin ticareti suçlarının incelemeleri, özünde fiyat ve miktar hareketlerinin incelenmesine dayanmaktadır. Olağandışı duruma sahip işlemler, paydaş olanların bir hiyerarşi içerisinde incelenmesini gerekmektedir. Dolayısıyla, manipülasyon ve içerden öğrenenlerin tespiti aslında hesaplama dayalı bir dizi işlem gerektiren analitik bir süreçtir ve hisse senedi piyasasının doğal akışı içerisinde izlediği yoldaki paydaşların arasındaki ilişkiyi ortaya koymaktır. Gözetim sistemlerinin ve erken uyarı sistemlerinin de temel mantığı, insangücüyle yapılan hesaplamalar zincirini bilgi teknolojilerinden yararlanarak gerçekleştirmek ve şüpheli bir durumda olası ilişkileri ortaya koyacak analizleri yapmaktır.

Konuya ilişkin olarak, New York Borsası (NYSE 2006) tarafından uygulanmakta olan sistem iyi bir örnek teşkil ettiğinden kısaca değinilmesi faydalı olacaktır. New York Borsası'nda 'Stock Watch' olarak adlandırılan ve manipülasyon ve içeriden öğrenenlerin ticaretini önceden tespit edebilmeye yönelik bilgisayarlı bir erken uyarı sistemi kullanılmaktadır. Stock Watch NYSE (2006) tarafından 'NYSE'de kayıtlı menkul kıymetler için yasal olmayabilecek anormal fiyat ve miktar hareketlerini gözlemleyen, NYSE'nin son teknoloji bilgisayar destekli gözetim birimidir.' biçiminde tanımlanmaktadır.

Erken uyarı sistemlerine bir başka örnek de Tayland Borsası (SET 2006) tarafından kullanılan ATOMS (Automated Tools for Market Surveillance) sistemidir. ATOMS sistemi

içerisinde yer alan Otomatik Uyarı Sistemi, gerçek zamanlı olarak fiyat ve/veya miktar hareketlerini, geçmiş değerlerle kıyaslayarak; olağandışı durumlarda sesli uyarı vermektedir. Uyarı neticesinde, olağandışı örüntü soruşturulmaktadır.

ATOMS sisteminin,

1. Olağandışı fiyat değişimleri için uyarı veren ‘Fiyat Uyarı Modülü’
2. Olağandışı miktar değişimleri için uyarı veren ‘Miktar Uyarı Modülü’
3. Fiyat ve miktar kombinasyonunda olağandışı değişimler için uyarı veren ‘Fiyat ve Miktar Uyarı Modülü’

olmak üzere 3 modülü bulunmaktadır. Ayrıca, ATOMS bünyesinde ön soruşturma ve dokümantasyon imkanı sağlayan alt sistemler de mevcuttur (SET 2006).

Stock Watch sisteminde veri madenciliğinden yararlanıldığı bilinmesine karşın; ATOMS ile ilgili böyle bir bilgi bulunmamaktadır. Ancak, her iki sistem de özünde olağandışı fiyat ve miktar değişimlerini belirlemeye yöneldiği için bir anlamda büyük hacimde veri içerisinde gizli bilgi ve ilişkileri keşfetmeye dayanmaktadır ki, veri madenciliğinin tanımıyla oldukça örtüştüğü görülmektedir.

Koyuncugil (2006) ise çalışmasında New York ve Tayland Borsası’nda kullanılan sistemlerin bir benzerini İstanbul Menkul Kıymetler Borsası Hisse Senetleri Piyasası (HSP) için geliştirmiştir. Sistem, manipülasyon tespitine yönelik veri madenciliğine dayalı bir erken uyarı sistemi olup; gerçek verilerle işletilerek Sistem’in başarıyla çalıştığı ispat edilmiştir. Sistem,

- Manipülasyona uğramış hisse senedini,
- Manipülatif işlemleri,
- Manipülatif işlemlere aracılık eden aracı kuruluşları,
- Manipülasyonu gerçekleştiren yatırımcıları,

otomatik olarak belirlemektedir. Benzer bir sistemin SPK veya İMKB’de hayata geçirilmesinin, oldukça büyük katkı sağlayacağı düşünülmektedir.

#### **4.2. Yasal Düzenlemelerin Etki ve Etkinlik Değerlendirmesi**

Kurul düzenlemeleri, temel anlamda, piyasanın tüm bileşenlerinin ihtiyaçlarına cevap verebilmek doğrultusunda şekillenmektedir. Ancak, bir düzenleme yapıldıktan sonra, olumlu olumsuz yönlerini gözlemek belirli bir zamanı gerektirdiği gibi piyasa ihtiyaçlarına tam olarak cevap verilip verilmediği de ancak kapsam içerisinde yer alan her durumla karşılaşıldığında anlaşılabilir. Dolayısıyla, her düzenlemenin olası etkisini önceden kestirebilmek veya ihtiyaç duyulan düzenlemeleri tespit edebilmek bir geri dönüşüm mekanizması oluşturmayı gerektirmektedir.

Halihazırda düzenleme yapılmış olan alanlardaki değişiminin yanısıra, tüm sermaye piyasası çevresinin, düzenleme öncesi ve sonrasının etüd edilmesi ile düzenlemenin etkisini belirlemek mümkündür.

Ayrıca, düzenleme yapılması düşünülen alana yönelik bir model geliştirilerek, sözkonusu alana yönelik etkisi olan faktörlerin belirlenmesi, alternatif senaryolarla simülasyonlar yapılması ve düzenleme öncesi önsel bilgi edinimi sağlanması mümkündür. Böylece, yanlış veya eksik bir düzenlemenin piyasaya ve geniş perspektiften tüm ekonomik sisteme vereceği zarar öngörülerek; düzenlemede gerekli iyileştirmelerin yapılması mümkün olabilir.

#### **4.3. Halka Açık Şirketlerin Değerlendirilmesi ve Halka Açıklığın Önündeki Engellerin Belirlenmesi**

Türkiye’de sermaye piyasalarının maalesef istenen gelişmişlik düzeyinde olmadığı açıktır. Mali sistem incelendiğinde fon fazlasına sahip olan tasarrufçular ve fon ihtiyacında olanlar üzere iki temel aktör söz konusudur. Buna ek olarak, bir de fon ihtiyacı ve fon fazlası olanlara aracılık eden aracı kuruluşlardan üçüncü aktör olarak bahsetmek mümkündür. Bu noktadan hareketle, konu ile ilgili olarak iki bakış açısına sahip olmak mümkündür. Bunlardan birisi halka açıklık oranının düşük olması, ikincisi de yatırımcıların tercihlerini sermaye piyasası araçlarından yana kullanmaması olarak belirtilebilir. Özünde birbirini besleyen bu iki olgunun halka açık şirket ayağı bu bölümde, yatırımcı ayağı ise Bölüm 4.4’de değerlendirilecektir.

Koyuncuğil ve Özgülbaş (2006a) çalışmalarında veri madenciliği yöntemlerinden CHAID karar ağaçları algoritması ile İMKB’de işlem gören KOBİ’lerin finansal profillerini belirlemiştir. Benzer biçimde, halka açık şirketlerin, SPK’nın elindeki verilerin yanı sıra

karakteristik verileri de anket yoluyla toplanarak; profillendirmesi yapılabilir. Profillendirmeden kasıt, halka açıklık açıklayıcı (hedef) değişken, derlenen ve toplanan veriler açıklayıcı değişkenler olmak üzere veri madenciliği yöntemlerinden;

- Lojistik regresyon,
- Karar ağaçları algoritmaları,

kullanımıyla sınıflandırmalar yapmak ve istatistiksel açıdan önemli ilişkileri ortaya koymaktır.

Halka açık şirket profilleri elde edildikten sonra, profillerden elde edilen şablon halka açık olmayan ama halka açılma potansiyeline sahip şirketler üzerine uygulanarak;

- Halka açılmaya elverişli olan şirketler,
- Halka açılmaya elverişli olmayan şirketler,

olmak üzere şirketler 2 gruba ayrıştırılabilir. Bu ayrıştırmadan sonra ‘Halka açılmaya elverişli olan’ ama buna karşın halka açık olmayan şirketlerin verileri derlenip, toplanarak; eldeki verilerin niteliğine bağlı olarak;

- Önemli bileşenler analizi,
- Faktör analizi,
- Karar ağaçları algoritmaları,

vb. veri madenciliği yöntemlerinin kullanımıyla, halka açılmanın önünde engel teşkil eden faktörler belirlenebilir.

#### **4.4. Yatırımcı Profiline Tespiti**

Mali sistemin ikinci temel aktörü olan fon fazlasına sahip tasarruf sahipleri veya yatırımcılardır. Piyasa derinliğinin artması konusunda, yatırımcıların sermaye piyasası araçlarına yönelmemesi temel sorunların başında gelmektedir. Sermaye piyasalarına yatırım yapmanın, bankacılığa göre daha fazla bilgi gerektirmesi ve riskin yatırımcıya ait olması gibi bir dizi gerekçeyle istenilen derinliğe ulaşılmaması açıklanabilir. Şüphesiz ki, bu açıklamalar oldukça gerçekçidir. Ancak, bu gerekçelerin hiçbirisi somut bulgulara dayanmamakta, çoğunlukla subjektif faraziyeler olmaktadır.



Maalesef, yatırımcılara yönelik erişilebilir veri olarak Merkezi Kayıt Kuruluşu (MKK) bünyesinde derlenen Hisse Senedi yatırımcılarının kimlik ve hesap bilgileri bulunmaktadır. MKK dışında çok fazla kullanılabilir veriye ulaşılamamaktadır. Söz konusu veriler de, Türkiye Sermaye Piyasası Aracı Kuruluşlar Birliği (TSPAKB 2006) tarafından yıllık olarak hazırlanan ‘Türkiye Sermaye Piyasası Raporu’nda tablolar halinde sunulmaktadır. Geçmiş dönük olarak bakıldığında da SPK tarafından gerçekleştirilmiş ‘Hanehalkı Tasarruf ve Yatırım Eğilimleri Anketi’ne yönelik tablolar Eser (2000) tarafından basılı hale getirilmiş olmasına karşın araştırma verilerine ulaşılamamaktadır.

Bu noktadan bakıldığında, hisse senetleri piyasasının (HSP) dışında yatırım fonları, yatırım ortaklıkları ve diğer yatırım araçlarına yönelik yatırımcı veya tasarruf sahiplerine yönelik hemen hemen hiçbir veri bulunmamaktadır. Türkiye’yi kapsayacak biçimde yapılacak bir araştırma ile yatırımcılara yönelik veriler elde edildikten sonra;

- Hisse senetleri,
- Yatırım fonları,
- Yatırım ortaklıkları,

yatırımcılarının profili ve potansiyel yatırımcıları belirlenebilir. Bu belirlemelerde,

- Yerleşim yeri,
- Yaş,
- Cinsiyet,
- Eğitim,
- Medeni durum,
- Gelir düzeyi,
- Meslek,

gibi sosyal, ekonomik ve demografik faktörlerin tasarruf ve yatırım kararlarına etkisi belirlenerek, mevcut yatırım araçlarının ve potansiyel enstrümanların hedef kitlesi belirlenebilir. Geliştirilecek enstrümanlar, potansiyel hedef kitleleri dikkate alınarak oluşturulabilir.

#### 4.5. Aracı Kuruluş Risk Profillerinin Belirlenmesi

Aracı kuruluşların risk değerlendirmesi ‘Sermaye Yeterliliği’ olarak adlandırılan mali tablolar üzerinden ardışık bir oranlamaya dayanmaktadır. Fakat, risk kavramının duruma özgü ve dinamik bir olgu olduğu düşünüldüğünde; risk belirlemesinin statik bir oranla değil olasılıklı bir modelle tespitinin daha anlamlı olacağı açıktır.

Aracı kuruluşlara yönelik geçmiş tecrübeler dikkate alındığında, aslında her vakanın bir şablon sunacağı açıktır. Bu noktadan hareketle, aracı kuruluşları bir tek oranla ifade etmek yerine, risk oluşumuna yol açacak niteliklerinin ağırlıklarına göre profillendirmek daha gerçekçi bir sonuç verecektir. Böylece, aracı kuruluşların ‘Sermaye Yeterliliği’ne bakarak risk arz edip, etmediğine değil;

- Risk arz etmeye ne kadar uzak ya da yakın olduğu,
- Mevcut seyrine devam etmesi halinde ne kadar sürede riskli konuma gireceği,
- Riskli konumdan çıkması için alması gereken önlemlerin ne olduğu,

sorularının cevapları da olasılıksal olarak verilebilecektir. Ayrıca, oran değerlendirmelerinde yer alabilecek subjektivite yerine sayıların tartışılmaz objektivitesi etkin kılınmış olacaktır. Dolayısıyla, subjektif normlar yerine, objektif normlar etkin kılınmış olacaktır.

## 5. SONUÇ

Büyük hacimde veriye sahip pek çok alanda gittikçe daha fazla kullanım yaygınlığına kavuşan veri madenciliği, finans alanında oldukça fazla uygulamaya yer bulmuştur. Özellikle bankacılık alanında özel sektör uygulamalarının ağırlık kazandığı veri madenciliği, her geçen gün sermaye piyasası ve sigortacılık uygulamalarında da artış sergilemektedir.

Bu çalışmada, veri içerisinde gizli kalmış ilişki ve bilgilerin açığa çıkarılıp, stratejik karar destek amaçlı kullanımı olarak tanımlanabilen veri madenciliği, Kurul sorumluluk alanına giren uygulama alanları yönünden değerlendirilmiştir. Değerlendirme mümkün olduğunca teknik detaylardan arıtılarak, olabildiğince yalın biçimde sunulmaya çalışılmıştır.

Genel anlamda sermaye piyasaları, özel anlamda ise Sermaye Piyasası Kurulu'nu temel alacak pek çok araştırmada veri madenciliği yöntemlerinin kullanımı mümkündür. Veri madenciliğinin, Bölüm 4'te aktarılan konular dışında da SPK çalışmalarına analitik ve stratejik karar destek amaçlı olarak katkı sağlayacağı düşünülmektedir. Ayrıca, gerekli veri altlığının hazırlanması halinde yalnızca sermaye piyasaları açısından değil tüm mali sistem açısından politika üretimi bakımından veri madenciliğinden etkin biçimde yararlanılabileceği düşünülmektedir.

## 6. KAYNAKLAR

- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. ve Verkamo, A. I. 1996. Fast discovery of association rules. *Advances in knowledge discovery and data mining*, AAAI/MIT Press, Cambridge, MA, 307-328.
- Brachman, R. J., Khabaza, T., Kloesgen, W., Piatetsky-Shapiro, G. ve Simoudis, E. 1996. Mining business databases. *Communications of the ACM*, Vol: 39, p. 42-48.
- Berson, A., Smith, S. ve Thearling, K. 1999. *Building data mining applications for CRM*. McGraw Hill, 510, USA.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. ve Zanasi, A. 1997. *Discovering data mining: from concept to implementation*. Prentice Hall PTR, Upper Saddle River, New Jersey, 195, USA.
- Chang, S., Chang, H., Lin, C. ve Kao, S. 2003. The effect of organizational attributes on the adoption of data mining techniques in the financial service industry: An empirical study in Taiwan. *International Journal Of Management*, vol: 20, p. 497-503.
- Chen, Z. 2001. *Data mining and uncertain reasoning: an integrated approach*. John Wiley & Sons, Inc., 370, Canada.
- Derby, B. L. 2003. Data mining for improper payments. *The Journal of Government Financial Management*. Vol. 52, p. 10-13.
- Eklund, T., Back, B., Vanharanta, H. ve Visa, A. 2003. Using the self-organizing map as a visualization tool in financial benchmarking. *Information Visualization*, Vol: 2 , p.171-81.
- Eser, R. 2000. *Hanehalkı Tasarruf ve Yatırım Eğilimleri Anketi Sonuçları, 1996*. Sermaye Piyasası Kurulu Yayınları, Ankara.
- Fayyad, U., Piatetsky-Shapiro G. ve Symth, P. 1996. From data mining to knowledge discovery in databases. *AI Magazine*, 17(3); 37-54 .
- Hastie, T., Tibshirani, R. ve Friedman, J. 2001. *The elements of statistical learning; data mining, inference and prediction*. Springer Series in Statistics, 533, USA.
- Hoppszallern, S. 2003. Healthcare benchmarking. *Hospitals & Health Networks*, Vol: 77, p. 37-44.

- İMKB (İstanbul Menkul Kıymetler Borsası). 2006. Web sitesi. www.imkb.gov.tr. Erişim Tarihi: 20.03.2006
- Kloptchenko, A., Eklund, T., Karlsson, J., Back, B., Vanhatanta, H. ve Visa, A. 2004 Combinig data and text mining techniques for analysing financial reports. *Intelligent Systems in Accounting Finance and Management*. Vol: 12. p. 29-41.
- Kovalerchuk, B. ve Vityaev E. 2002. Data mining in finance: advances in relational and hybrid methods. Kluwer Academic Publishers, 308, USA.
- Koyuncugil, A. S. 2004. Veri Madenciliği veya Bir Başka Değişle Akıllı Algoritmalarla İstatistik Kullanımı. *Emniyet Genel Müdürlüğü Polis Dergisi Bilişim Özel Sayısı*, Sayı:37, s.38-40.
- Koyuncugil, A. S. 2006. Bulanık Veri Madenciliği ve Sermaye Piyasalarına Uygulanması. Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Doktora tezi (basılmamış), 186 sayfa, Ankara.
- Koyuncugil, A. S. ve Özgülbaş, N. 2006a. Financial Profiling Of SMEs: An Application By Data Mining. *The European Applied Business Research (EABR) Conference*, June 26-29. Florence, Italy.
- Koyuncugil, A. S. ve Özgülbaş, N. 2006b. Is There a Specific Measure for Financial Performance of SMEs. *The Business Review, Cambridge* Vol:5, No:2. September.
- Koyuncugil, A. S. ve Özgülbaş, N. 2006c. İMKB’de İşlem Gören Kobi’lerin Finansal Başarısızlığına Etki Eden Faktörlerin Veri Madenciliği İle Belirlenmesi. *3. KOBİ ve Verimlilik Kongresi*. 17-18 Kasım. İstanbul.
- Kumar, N., Krovi, R. ve Rajagopalan, B. 1997. Financial decision support with hybrid genetic and neural based modeling tools. *European Journal of Operational Research*, Vol:103, p. 339-49.
- Lansiluoto, A., Eklund, T., Barbro, B., Vanharanta, H. ve Visa, A. 2004 Industry-specific cycles and companies’ financial performance comparison using self-organising maps. *Benchmarking*, Vol: 11. p. 267-86.
- Lee, K. C., Han, I. ve Kwon, Y. 1996. Hybrid neural network models for bankruptcy predictions. *Decision Supporte Systems*, Vol:18 , p.63-73.

- Magnusson, C., Arppe, A., Eklund, T. ve Back, B. 2005. The language of quarterly reports as an indicator of change in the company's financial status. *Information & Management*. Vol: 42, p. 561-70.
- Moss, L.T. ve Atre, S. 2003. Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley Publishing, 576, USA.
- Nazem, S. ve Shin, B. 1999. Data mining: New arsenal for strategic decision making. *Journal of Database Management*, Vol: 10, 39-42.
- NYSE (New York Stock Exchange). 2006. Web sitesi. [www.nyse.com](http://www.nyse.com). Erişim Tarihi: 15.10.2006
- Özgülbaş, N. ve Koyuncugil, A.S. 2006. İMKB'de İşlem Gören KOBİ'lerin Veri Madenciliği Karar Ağaçları Algoritmalarından CHAID ile Profillendirilmesi ve Küreselleşme Sürecinde Güçlü Ve Zayıf Yönlerinin Belirlenmesi. *10. Ulusal Finans Sempozyumu*. 1-4 Kasım. İzmir.
- Özgülbaş, N., Koyuncugil, A.S. ve Yılmaz, F. 2006. Identifying the Effect of Firm Size on Financial Performance of SMEs. *The Business Review, Cambridge*. Vol:5, No:2. December.
- Piatetsky-Shapiro, G. 1991. Knowledge discovery in real databases: a workshop report. *AI Magazine*, 11(5); 68-70.
- Rao, C.R. 2001. Statistics: reflections on the past and visions for the future. *Communication in Statistics-Theory and Methodology*, 30(11); 2235-2257.
- Roiger, R. J. ve Geatz, M. 2002. Data mining: a tutorial based primer. Addison-Wesley Publishing, 408, USA.
- SET (Stock Exchange Thailand). 2006. Web sitesi: <http://www.set.or.th/en/index.html>. Erişim Tarihi: 27.09.2006
- Tam, K. Y. ve Kiang, M. Y. 1992. Managerial applications of neural networks: The case of bank failure predictions. *Decision Sciences*, Vol:38, p. 926-48.
- Thearling, K. 2005. Web sitesi: [www.thearling.com](http://www.thearling.com). Erişim Tarihi: 18.06.2005
- TSPAKB (Türkiye Sermaye Piyasası Aracı Kuruluşları Birliği). 2006. Web sitesi: [www.tspakb.org.tr](http://www.tspakb.org.tr). Erişim Tarihi: 21.11.2006

TWO CROWS. 1999. Introduction to data mining and knowledge discovery. Two Crows Corporation, 36, USA

Witten, I. H. ve Frank, E. 2000. Data mining. Morgan Kaufmann Publishers, 371, San Francisco, California.