



SERMAYE PİYASASI KURULU
ARAŞTIRMA RAPORU
(KAMUYA AÇIK)



BORSA ŞİRKETLERİNİN
SEKTÖREL RİSK PROFİLLERİNİN
VERİ MADENCİLİĞİYLE BELİRLENMESİ

DR. ALİ SERHAN KOYUNCUGİL
İSTATİSTİKÇİ

YÖNETİCİ ÖZETİ

Günümüzün artan rekabet koşulları, geleceğe yönelik belirsizlikleri azaltmak ve potansiyel olumsuzluklara karşı öncül tedbirler almak düşüncesini bir istemden zorunluluğa dönüştürmüştür. Kayıp fonksiyonunun beklenen değeri olarak ifade edilebilen risk kavramı ise mevcut ve potansiyel sorunları ölçmede en etkin olarak kullanılan ölçüm aracı olmaktadır. Şüphesiz ki, riskin istatistiksel perspektiften objektif yorumunun yanısıra, subjektif tanımları da söz konusudur. Ancak, son dönemde risk konusunu gündeme getiren ve sistematik bir yaklaşımı zorunlu kılan BASEL-II kriterleri, subjektif risk kavramını, objektif normlara bağlamayı zorunlu hale getirmiş ve metodolojik dayatmaları içerecek hale gelmiştir.

BASEL-II perspektifinden bakıldığında, düzenleyici kurum ve düzenleme alanında yer alan kurum/kuruluş olmak üzere konuya iki açıdan yaklaşmak mümkündür. Fakat, düzenleme alanında yer alan kurum/kuruluşların niteliği, Düzenleyici Kuruluşlar açısından da bağlayıcılık getirmektedir. Bankacılık sektörü bakımından kredi kullanımını objektif, bilimsel temelli ve risk tabanlı derecelendirmeye dayandıran BASEL-II, yapılacak düzenlemeler bakımından da en yaygın etkiye KOBİ'ler açısından sahip olacaktır.

İMKB'de işlem gören şirketler incelendiğinde pek çoğunun KOBİ statüsünde olduğu görülmektedir. Dolayısıyla, sadece halka açıklık ve Borsa'ya kote olmak açısından değerlendirmenin ötesinde KOBİ perspektifinden de değerlendirme yapmak SPK için bir zorunluluk olmaktadır.

Riskin rasgelelik içeren değişken bir öge olması dinamik bir yaklaşım gerekliliğini ortaya koymaktadır. Risk dışında, son dönemin en güncel kavramlarından birisi de veri içerisinde gizli kalmış tahminsel bilgi, örüntü ve ilişkilerin otomatik olarak açığa çıkarılıp; stratejik karar-destek amaçlı kullanımı olarak tanımlanan veri madenciliğidir. Dolayısıyla, otomatik-tahminsel kelimeleri, risk gibi değişken bir kavramın belirlenmesinde oldukça uygun görünmektedir.

Bu çalışmada, Borsa şirketlerinin sektörel ayrımında tüm alt gruplarda risk belirmesini sağlayacak, risk tayininde fikir verebilecek öncü göstergelerin tanımlanmasına imkan verecek, erken uyarı özellikli veri madenciliğine dayalı bir risk modeli tanımlanmaktadır. Çalışmada, veri madenciliği yöntemlerinden CHAID karar ağaçları algoritması kullanılmıştır.

İÇİNDEKİLER

1. GİRİŞ	1
2. LİTERATÜR TARAMASI	4
3. VERİ MADENCİLİĞİ	7
3.1. VERİ MADENCİLİĞİNİN TANIMI	7
3.2. VERİ MADENCİLİĞİ YÖNTEMLERİNİN SINIFLANDIRILMASI.....	9
3.3. BAŞLICA VERİ MADENCİLİĞİ YÖNTEMLERİ.....	10
3.4. CHAID KARAR AĞACI YÖNTEMİ.....	11
4. YÖNTEM.....	13
4.1. DEĞİŞKENLER.....	13
4.2. MODEL	14
4.2.1. Verinin hazırlanması	14
4.2.2. Veri madenciliği yöntemi CHAID algoritmasının uygulanması.....	14
4.2.3. Risk profillerinin tanımlanması.....	15
5. SONUÇ.....	17
6. KAYNAKLAR.....	19

TABLolar DİZİNİ

Tablo 1. İstatistiksel analiz ve veri madenciliğinin karşılaştırılması	8
Tablo 2. Araştırmada kullanılacak değişkenler ve tanımları.....	13

ŞEKİLLER DİZİNİ

Şekil 1. CHAID karar ağacı örneği	14
-----------------------------------------	----

BORSA ŞİRKETLERİNİN SEKTÖREL RİSK PROFİLLERİNİN VERİ MADENCİLİĞİYLE BELİRLENMESİ

1. GİRİŞ

Risk ölçmeye yönelik çalışmalar insanlık tarihinin başlangıcından itibaren kendini koruma içgüdüyle ortaya çıkmasına karşın; matematiksel anlamda en sistematik ve günümüze de ışık tutan yaklaşım ‘Olasılık Teorisi’dir. Özünde, doğanın rasgeleliğinden kaynaklı belirsizliği azaltma çabasının bir ürünü olan Olasılık Teorisi, bilimsel gelişmelerin sağladığı katkıyla farklı belirsizlik tiplerinin çözüm metodolojisinde de temel olmuştur. Bir başka deyişle, bütün belirsizlikler olasılıksal değildir ve ‘Olasılık Teorisi’ farklı belirsizlik tiplerini açıklama da yetersiz kalabilmekte, sadece ‘Olasılıksal’ olarak ifade edilebilen ve ‘kısmen belirsiz’ sayılabilecek durumları açıklayabilmektedir. Bu yetersizlik, zaman içerisinde, Kaos Teorisi, Pürüzlü Küme Teorisi ve Bulanık Küme Teorisi gibi farklı metodolojik yaklaşımların ortaya çıkmasına neden olmuştur. Şüphesiz ki, sözü geçen yaklaşımlar çok genel bir bakışla ‘Analitik’ veya özelde ‘İstatistiksel’ perpektifli ve ‘Stokastik’ yaklaşımlardır. Buna karşın, stokastik olmayan (deterministik), sadece durum tespitinden ibaret yaklaşımlar özellikle sosyal bilimlerde oldukça fazla yer bulmuştur. Kişiden kişiye değişen, yoruma açık, subjektif değerlendirmeler zaman zaman oldukça büyük, telafisi olmayan zararlara yol açabilmektedir.

Risk kavramı değerlendirildiğinde, matematiksel ifadesi bakımından kayıp fonksiyonunun beklenen değeridir ki, kayıp fonksiyonunu tanımlayabilmek genelde bir fayda fonksiyonu tanımlamayı gerektirmektedir. Ayrıca, genel yaklaşım da eşit olasılıklı bir süreç yerine önsel ve sonsal olasılık biçiminde yansıyan Bayesci yaklaşımdır. Ancak, özellikle finansal çalışmalara bakıldığında risk tanımlamasının olasılıksal veya istatistiksel yaklaşımdan uzak biçimde ‘Oran Analizi’, ‘Trend Analizi’ gibi yoruma açık yaklaşımlarla ifade edildiği görülmektedir ki, risk tanımlamasının olasılıksal doğasına aykırı bir yaklaşımdır. Öyleki, değerlendirme sadece mevcut durum için yapılabilmekte, olasılıksal değil, kesin değerlerle ifade edilmekte ve nadir gözlenme sıklığı ile sık gözlenme sıklığına sahip durumlar beraber dikkate alınmaktadır. Dolayısıyla, bir anlamda kaybetme veya başarısızlık olasılığını ortaya koyan risk kavramı öncül gösterge veya erken uyarı niteliğini kaybetmekte ve işlevsizleşmektedir.

Risk kavramı bugünlerde yoğun biçimde gündemi meşgul etmektedir. Bu gündem yalnız Türkiye ile sınırlı kalmamakta, tüm dünyada risk konusunda hararetli çalışmalar süregelmektedir. Risk kavramını son dönemde bu kadar güncel kılan ise BASEL – II kriterlerinin risk ölçümlemesini temel almasıdır. Şüphesiz ki, düzenleyici kuruluşlar gerekli düzenlemeleri yapacaktır. Fakat, düzenleme alanı içinde kalan başta bankalar olmak üzere tüm potansiyel kredi müşterileri kendisini finansal yönetim anlamında risk indirgeyici tedbirler almaya yönlendirmelidir. Kredi kullanımı başta olmak üzere, taşıdığı risk ölçüsünde faizlendirme en temel uyum çıktısıdır. Uyumlulaştırmada, şirketlerin taşıdıkları riski belirleyip yönetebilmeleri gerekmektedir ki, yoruma açık subjektif risk yaklaşımı bunu imkansız kılmaktadır. Dolayısıyla, oran ve trend analizi ile benzeri deterministik yaklaşımlar ihtiyaca cevap vermeyecektir. Bunun yerine olasılıksal bir yaklaşım izlenmesi gerekmektedir. Bu noktada en temel sorun küçük ve orta boyutlu (KOBİ) işletmelerde çıkmaktadır. Yalnız Türkiye'nin değil ABD ve AB'nin de ekonomik lokomotifleri olan KOBİ'ler, finansal yönetim konusunun ötesinde finans kavramından dahi uzaktır. Temel sorun da bu noktada çıkmaktadır. KOSGEB tanımına göre 1-250 kişi arasında çalışana sahip olma koşuluyla KOBİ tanımına giren şirketleri BASEL-II kriterlerine uyumlaştırma adına finansal yönetim ve ötesinde de risk kavramıyla tanıştırmamızın nasıl mümkün olacağı çok ciddi bir soru işaretidir. Olaya Sermaye Piyasası Kurulu perspektifinden bakıldığında, bugün İMKB şirketlerinin büyük bir kısmı KOBİ tanımındadır. Dolayısıyla, BASEL – II sorumluluklarını yerine getirmesi adına doğrudan SPK ile de ilgilidir. Bu ilgi, uyumlulaştırmanın ötesinde risk yönetim metodolojisi ile sistemini kurumsallaştırmayı ve etkin bir gözetim ile denetim mekanizması kurmayı gerektirmektedir. Bu noktada, hızlı karar vermeyi sağlayacak ve subjektiviteye yer vermeyecek, analitik derinliğe sahip bir risk yaklaşımı benimsenmesi gerekmektedir. Şüphesiz ki, SPK sorumluluğunda yer alan ve risk bazlı analitik yaklaşıma sahip olması gereken tek çalışma alanı İMKB şirketleri değildir. Ancak, ekonominin barometresi denilebilecek bir alan tartışmasız bir önceliğe sahiptir. Bu nedenle, çalışma alanı olarak İMKB seçilmiştir.

Risk ölçümlenmesi konusunda 'Risk Teorisi' açısından olaya bakıldığında, belirtildiği gibi, bir fayda fonksiyonu ve ona bağlı bir kayıp fonksiyonu tanımlamak gerekmektedir. Şüphesiz ki, doğası gereği, doğrudan veya dolaylı olarak pek çok etkiye açık sistemler olan finansal sistemlerde değişimi ölçebilecek bir fonksiyon veya fonksiyonlar grubu tanımlamak oldukça zordur. Buna karşın, dolaylı olarak riski ölçmeye yarayacak bazı öncü veya erken uyarı sinyali vazifesi görecektir. Göstergeler tespit edebilmek daha rasyoneldir.

Ancak, Hisse Senetleri Piyasası gibi rasgeleliğin de ötesinde kaotik yapıya sahip bir alanı bir tek fonksiyon ile modellemek bilinen veya bilinmeyen pek çok değişkeni göz ardı etmeye yol açacak veya çok hantal ve işletilemeyecek bir model tanımlamaya yol açacaktır ki, ihtiyaca çok fazla cevap veremeyecektir. Her faktörü içeren ama açıklama düzeyi düşük bir model yerine, değişimi algılayabilen ve kendini yenileyebilen parametrik olmayan bir model ihtiyaca daha fazla cevap verecektir. Bu noktada, soruna en uygun çözüm veri madenciliğidir. Makine Öğrenimi ve Yapay Zeka yöntemleriyle desteklenmiş İstatistiksel yöntemler sunan Veri Madenciliği, günümüzün en popüler analitik yaklaşımıdır. Bilgi teknolojilerinin yaygınlaşmasıyla beraber 2000’li yılların başından itibaren özellikle bankacılık kesimine, ticari hedeflere yönelik olarak hizmet vermeye başlayan veri madenciliği, bugün yaygın olarak pek çok alanda kullanılmaktadır. Kullandığı yöntemler bakımından varsayımları sağlamak adına tanımında ‘Büyük veri hacmi’ ön koşuluna sahip olan veri madenciliği, ilk başta sınırlı sayıda hiyerarşik veri depolama imkanına sahip kurum ve kuruluşlara katkı vermekteydi. Ancak, gerek analitik yöntemlerin gerekliliğinin gün geçtikçe daha belirgin biçimde ortaya çıkması, gerekse depolama ünitelerinin yaygınlık kazanıp; veri tabanı ve veri ambarı gibi hiyerarşik veri depolama sistemlerinin yaygınlık kazanmasına paralel olarak veri madenciliği de üstel olarak artacak biçimde kurum ve kuruluşlara yerleşmeye başlamıştır. Veri madenciliğine tanım olarak bakıldığında, ‘Büyük hacimde veri içerisinde gizli kalmış, tahminsel ilişki, örüntü ve bilgilerin otomatik olarak keşfedilmesi ve stratejik karar-destek amaçlı olarak kullanılması’ olarak ifade edilmektedir. Tanıma bakıldığında ‘Tahminsel’ ve ‘Otomatik’ olmak üzere iki anahtar sözcük ön plana çıkmaktadır.

Bu araştırma raporunda, İMKB şirketleri özelinde risk ölçümlemesindeki değişimleri otomatik olarak değerlendirmeyi ve öncü risk göstergelerini sektörel bazda belirlemeyi sağlayacak veri madenciliğine dayalı bir sektörel risk modeli oluşturulmaktadır. Risk kavramı finansal performansa dayandırılmakta ve finansal performans açısından sektörel benzeşim ve ayrışmaları gösteren tüm alt grupları ortaya koyacak risk profilleri tanımlanmakta; bu sayede, her alt grup içerdiği risk düzeyi bakımından ayrı ayrı değerlendirilebilmektedir. Ayrıca model, her bir sektör ve alt grup bazında riske işaret eden göstergeleri ve limit değerlerini de tanımlamakta olup; bu niteliğiyle de bir Erken Uyarı Sistemi özelliğine sahiptir. Modelin sunduğu Erken Uyarı Sistemi özelliği sayesinde potansiyel riskleri öngörebilmek ve risk gerçekleşmeden tedbir alabilmek kısacası ‘proaktif’ olabilmek mümkün kılınmaktadır. Böylece, karar mekanizmalarına, strateji belirleme de analitik derinlikli, objektif normlara sahip, delile dayalı ve ispatlanabilir bilgi sağlamak imkanına erişilebilmektedir.

2. LİTERATÜR TARAMASI

Finansal olarak başarılı ve başarısız işletmeleri ayırmaya yönelik çabalar Beaver'ın (1966) rasyoları kullanarak gerçekleştirdiği tek değişkenli ve Altman'ın (1968) Çoklu Diskriminant analizine dayalı Z-skorları ile başlamıştır. Çok değişkenli istatistik modelleri kullanan diğer önemli çalışmalar; Çoklu Diskriminant modelini kullanan, Altman vd. (1977); Lojit ve Probit modellerini kullanan Zmijewski (1984), Zavgren (1985), Jones (1987), Pantalone ve Platt (1987), Rudolph ve Hamdan (1988), Ward (1993), Ward (1994), Jones ve Hensher (2004) örnek olarak sayılabilir (Koyuncugil ve Ozgulbas, 2006c).

Yapay sinir ağları, 1980'li yıllarda finansal başarısızlık ve iflas gibi sorunların belirlenmesinde kullanılmaya başlanmış ve Hamer (1983), Coats ve Fant (1992), Coats ve Fant (1993), Chin-Sheng vd. (1994), Klersey ve Dugan (1995), Boritz vd. (1995), Tan ve Dihardjo (2001) ve Anandarajan vd. (2001) gibi araştırmacılar araştırmalarında yapay sinir ağlarına yer vermiştir (Koyuncugil ve Ozgulbas, 2006c).

Gaytan ve Johnson (2002) bankacılık krizleri için erken uyarı sistemlerine yönelik bir literatür taraması yapmış ve taramada ağırlıklı olarak Probit ve Logit modellere dayalı sistemler önerildiği gözlenmiştir (Koyuncugil 2006). Collard (2002) erken uyarı sistemlerinin önemine vurgu yapmış ve işte başarısızlık ve riske yönelik 10 erken uyarı sinyali sunmuştur. Gunter ve Moore (2003) bankanın finansal durumunu göstermek için bir erken uyarı modeli geliştirmeyi hedeflemişlerdir. CAMELS derecelendirme sisteminin kapsamında yer alan finansal faktörlerin ana kategorileri altında yer alan 12 finansal göstergelyi içeren bir istatistiksel model kurmuşlardır.

Jacops ve Kuper (2004) Asya'daki 6 ülke için bir erken uyarı sistemi sunmuşlardır. Finansal kriz olasılığını hesaplamak için çok değişkenli binomial bir model kurmuşlardır.

Liu et al. (2006) dünyada meydana gelen finansal krizler üzerine odaklanmışlardır. Bulanık C-ortalamalar yönteminin ekonomik ve finansal krizlerin tanımlanmasında erken uyarı sistemi olarak yardımcı olacak biçimde nasıl kullanılabileceğini göstermişlerdir.

Ülkemizde erken uyarı sistemlerine ilişkin yapılan çalışmalardan biri olan Çilli ve Temel'in (1988) çalışmasında, bankacılık krizlerini tespit etmeye yönelik olarak Diskriminant Analizi ve Faktör Analizi'ne dayalı bir erken uyarı sistemi önerilmektedirler.

TCMB, Bankacılık Genel Müdürlüğü bünyesinde yer alan Bankalar Veri Kontrol Müdürlüğü ve Bankalar Gözetim Müdürlüğü tarafından yürütülen, bir bankacılık gözetim sistemi de söz konusudur. Gözetim sistemi, Bankalardan elde edilen verilerin içsel

tutarlılıkları kontrol edildikten sonra 75 adet rasyo ile özkaynak yeterliliği, aktiflerin niteliği, karlılık, likidite ve pasif yapısı incelenerek, uyarı raporları hazırlanmaktadır. Bunun yanı sıra, performans raporları, bilanço ve kar-zarar tablolarında yer alan kalemlerin oransal-yapısal gelişmeleri, yabancı para pozisyonu, likidite, krediler portföyü, nakit akımları analizi gibi raporlar da TCMB Bankacılık Genel Müdürlüğü tarafından hazırlanmaktadır (Gönül ve Eroğlu 1999).

Gerni vd. (2005) ise Türkiye'deki ekonomik krizleri tahmin etmek üzere oluşturdukları erken uyarı modelinde, bir Logit model kurmuşlardır.

Canbas vd. (2006) İMKB'de finansal sorun yaşayan şirketlerin tespitini sorgulamışlardır. Önemli Bileşenler Analizi ve Diskriminant Analizi'nden oluşan entegre bir erken uyarı modeliyle finansal sorunları tahmin etmeye çalışmışlardır.

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, 2000). 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.

Pek çok şirket ve bunların yöneticileri gelmekte olan bir finansal kriz ve yaşadıkları risk semptomlarının farkında değillerdir. Bununla beraber, semptomlar ve sinyaller başladığı zaman ilk olarak ne yapmaları gerektiği ve riski nasıl yöneteceklerini de bilmemektedirler. Finansal sorunlara yönelik erken uyarı sinyallerinin tanımlanmasıyla, yöneticiler bu sorunları ve riskleri ortadan kaldıracabilecek veya en azından üstesinden gelebileceklerdir. Ağırlıklı olarak veri madenciliğine dayalı olmak üzere analitik teknikleri baz alan erken uyarı sistemi çalışmaları aşağıda yer almaktadır.

Apoteker ve Barthelemy (2005) gelişmekte olan piyasalardaki finansal krizlere odaklanmışlardır. Ülkelerin risk sinyallendirilmesinde yeni geliştirilen parametrik olmayan bir metodoloji kullanmışlardır. Gelişmekte olan piyasalardaki finansal krizleri tahmin etmek için 9 tane erken uyarı sinyali geliştirmişlerdir.

Ko ve Lin (2005) veri madenciliğine dayalı modülerize bir finansal bozulma öngörü mekanizması geliştirmiştir.

Chan ve Wong (2007) finansal bozulmaları ve bütün olası senaryolar için gelecekteki finansal krizleri bulmaya çalışmışlardır. Bu hedefe ulaşmak için esnekliği ölçmek üzere veri madenciliğine dayalı bir erken uyarı sistemi kullanmışlardır.

Kamin et al. (2007) gelişmekte olan piyasalardaki krizlerde iç ve dış faktörlerin rolünü tanımlamak için erken uyarı sistemi kullanmışlardır. Gelişmekte olan 26 ülkenin kur krizleri için pek çok probit modelle tahminde bulunmuşlardır. Bu modeller, iç ve dış faktörlerin kriz olasılığına katkısını ayrı ayrı tanımlamak için kullanılmıştır.

Tan ve Quekuan (2007) Genetik Tamamlayıcı Öğrenimi (GTÖ) hisse senetleri piyasası tahmin edicisi olarak kullanmış ve banka başarısızlığı için erken uyarı sistemini sorgulamışlardır. Deneysel sonuçlar göstermiştir ki, GTÖ hisse senetleri piyasası tahmin edicisi ve banka başarısızlığı erken uyarı sistemi olarak oldukça yetenekli bir sayısal finans aracıdır.

Borsalardaki şüpheli işlemler için kullanılan Stock Watch, ASAM (Automated Search And Match) ve ATOMS (Automated Tools for Market Surveillance) gibi erken uyarı veya gözetim sistemleri de bulunmaktadır. Stock Watch, New York Borsası'nın, fiyat ve miktar hareketlerindeki yasadışı işlemleri gösteren, en gelişmiş bilgisayarlı gözetim birimidir. Ayrıca, halka açık bilgiler arasındaki çapraz tutarlılığı araştırıp inceleyen bir sistemdir (www.nyse.com, 2007). Tayland Borsası (SET) da, piyasa gözetimi için bilgisayar destekli bir sisteme sahiptir. ATOMS adı verilen sistem, tüm işlemleri takip etmekte, sıradışı işlemleri analiz ederek şüpheli durumları sorgulayarak raporlamaktadır (www.set.or.th/en/index.html, 2007).

Koyuncugil (2006) ise İMKB Hisse Senetleri Piyasası'nda işlem manipülasyonu tespitine yönelik bulanık veri madenciliğine dayalı bir erken uyarı sistemi tasarlamış ve gerçek verilerle test ederek Sistem'in 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 (2006b) İMKB'deki KOBİ'lerle ilgili bir başka çalışmada, KOBİ'lerin finansal risk ve performansını etkileyen faktörleri göz önüne alarak, bir finansal performans kriteri tanımlamışlardır.

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 finansal başarısızlığa etki eden faktörler belirlenmiştir. Ayrıca, Özgülbaş ve Koyuncugil (2006) ve Özgülbaş vd. (2006) tarafından yapılan finansal erken uyarıya dayalı farklı çalışmalarda KOBİ'lerin finansal anlamda güçlü ve zayıf yönlerinin yanında, finansal performans ile KOBİ'lerin büyüklüğü (boyutu-ölçeği) arasındaki ilişki belirlenmiştir.

Raporun özüne yönelik en güncel çalışma ise Koyuncugil ve Özgülbaş (2007) tarafından gerçekleştirilen, İMKB şirketleri için veri madenciliğine dayalı olarak geliştirilen finansal erken uyarı sinyalleri tespit sistemidir.

3. VERİ MADENCİLİĞİ

3.1. Veri Madenciliğinin Tanımı

Veritabanları günümüzde terabaytlarla ifade edilen boyutlara ulaşmıştır. Zaman içerisinde bu büyük hacimde verinin içinde stratejik önem taşıyan gizli enformasyon yattığı belirlenmiştir. 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, Klasik İstatistik temel olmak üzere Yapay Zeka, Makine öğrenimi ve Örüntü Tanıma gibi pek çok disiplinin kesişimi olan bir alandır. Her ne kadar Klasik İstatistik temeline dayansa da, Moss and Atre (2003) istatistiksel analizler ve veri madenciliğinin karşılaştırması ve farklılaştığı noktaları Tablo 1'de vermişlerdir.

Tablo 1. İstatistiksel analiz ve veri madenciliğinin karşılaştırılması

<i>İstatistiksel Analiz</i>	<i>Veri Madenciliği</i>
İstatistikçiler genellikle bir hipotez ile başlarlar.	Veri madenciliği hipoteze gerek duymaz.
İstatistikçiler hipotezlerini eşleştirmek için kendi eşitliklerini geliştirmek zorundadırlar.	Veri madenciliği algoritmaları eşitlikleri otomatik olarak geliştirir.
İstatistiksel analizler sadece sayısal verileri kullanır.	Veri madenciliği farklı tiplerde data kullanır (örneğin metin, ses) sadece sayısal veriyi değil.
İstatistikçiler kirli veriyi analizleri sırasında bulur ve filtre ederler.	Veri madenciliği temiz veriye dayanır.

Koyuncu (2004) ise Veri Madenciliği'ni aşağıdaki gibi açıklamıştır:

‘Veri Madenciliğini İstatistiksel Tekniklerin evrim geçirmiş hali olarak düşünmenin de mümkün olmasının yanı sıra; Veri Madenciliği'nin, aşağıda detaylandırılacak olan, kendisine has betimleyici karakteristikleri vardır. Veri Madenciliği, metodoloji olarak güçlü teorik dayanaklara sahip olsa da; uygulamada yazılımlarla karşımıza çıkmaktadır. Veri Madenciliği kavramı yazılım bazında incelendiğinde; karşılaşılan temel karakteristiklerinden birisi bilgilerin otomatik olarak açığa çıkarılması, keşfedilmesidir. Veri Madenciliği'nin ikinci betimsel karakteristiği ise tahmin edici modelleme yapmasıdır. Dolayısıyla, Veri Madenciliği bir anlamda otomatik tahminsel modelleme ile bilgi keşfi olarak da tarif edilebilir. Bu ifade bile Veri Madenciliği'nin temel dayanaklarını açığa çıkarabilmektedir:

- Tahminsel modelleme sağlayacak istatistiksel yöntemler,
- Bunların otomatik olarak gerçekleştirilmesini sağlayacak akıllı algoritmalar.

Veri Madenciliği'nin betimleyici karakteristiklerinden olan ‘Tahmin Edici Modelleme Yapma’ özelliği, Veri Madenciliği'nin arkasında çalışan mantığın istatistik olduğunu ortaya koymaktadır. Bunun ötesinde, Veri Madenciliği'nin tanımından gelen ve temel dayanağı olan; genelde de artık terabaytlarla ifade edilen veri miktarı ile gerçekleşen ‘büyük hacimde veri’ ile;

- Veri madenciliği yöntemi olarak telaffuz edilen İstatistiksel Analiz

Yöntemlerinden pek çoğunun ön koşulu olan Normal Dağılım varsayımlarını sağlamak,

- Etüd edilen ve modellemede kullanılacak verinin, olası tüm davranışlarını gözlemleyebilmek,

hedeflerine işaret edilmektedir.’

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 söz konusu 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,
- Sınır 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).

3.4. CHAID Karar Ağacı Yöntemi

Karar ağacı adından da anlaşıldığı gibi ağaç olarak görünen, tahminsel bir modeldir. Ağacın her dalı bir sınıflandırma sorusu ve yaprakları da veri setinin bu sınıflandırmaya ait parçalarıdır. Ağaç yapısından ve kolay kural çıkarımına imkân tanınması nedeniyle karar ağaçları anlaşılabilir modeller kurmak için oldukça faydalı bir tekniktir. Karar ağacı teknolojisi veri setlerinin ve iş problemlerinin keşfi için kullanılabilir. Bu genellikle ağacın her bir bölümündeki tahmin edicilere ve değerlerine bakarak yapılabilir. Sıklıkla bu tahmin ediciler kullanılabilir içerik sağlayabilir veya cevaplanması gereken sorular önerebilir. Eğer ağaç tek bir kayıt kalana kadar büyümesine devam ederse, pek çok soru ve dal yaratılacağı tasavvur edilebilir. Ağacı bu kadar büyütme hem hesaplama açısından pahalı hem de gereksizdir (Berson vd., 2000).

Son 30 yılda pek çok karar ağacı öğrenim metodu geliştirilmiştir (Quinlan, 1993; Mitchell, 1997) ve kredi başvurusunda risk değerlendirmesi gibi finansal ve bankacılık uygulamalarında başarılı olarak kullanılmaktadır (Kovalerchuk ve Vityaev, 2000). En sık kullanıma sahip karar ağacı modelleri ID3 ve daha gelişmiş modeli C4.5, Sınıflandırma ve

Regresyon Ağaçları (Classification and Regression Trees-CART) (Breiman vd., 1984) ve CHAID'tir (Chi-Square Automatic Interaction Detector- Otomatik Ki-Kare İlişki Belirleyici) (Ribic ve Miller, 1998).

CHAID ile diğer yöntemler arasındaki en önemli farklılıklarından birisi, ağaç türetimidir. ID3, C4.5 ve CART ikili ağaçlar türetirken, CHAID ikili olmayan çoklu ağaçlar türetir (Berson vd., 2000). CHAID sürekli ve kategorik tüm değişken tipleriyle çalışabilmektedir. Bununla beraber, sürekli tahmin edici değişkenler otomatik olarak analiz amacına uygun olarak kategorize edilmektedir. CHAID, Ki-Kare metriği vasıtasıyla, ilişki düzeyine göre farklılık rastlanan grupları ayrı ayrı sınıflamaktadır. Dolayısıyla, ağacın yaprakları, ikili değil, verideki farklı yapı sayısı kadar dallanmaktadır. CHAID algoritması aşağıda verilmektedir:

1. Her bir tahmin edici değişken X için, X 'in, Y hedef değişkenini dikkate alan en az öneme sahip kategori çiftini bul (bu, en büyük p değerine sahip olandır). Yöntem, Y 'nin ölçüm düzeyine bağlı olarak p değerlerini hesaplayacaktır.
 - a. Eğer Y sürekli ise F testini kullan.
 - b. Eğer Y isimsel ise X 'in kategorileri satırlarda ve Y 'nin kategorileri sütunlarda olacak biçimde iki yönlü çapraz tablo düzenle. Pearson ki-kare testini veya olabilirlik oranı testini kullan.
 - c. Eğer Y sıralı ise bir Y birliktelik modeli (Clogg ve Eliaisin, 1987; Goodman, 1979; Magidson, 1992) uygundur. Olabilirlik oranı testini kullan.
2. En büyük p değerine sahip X 'in kategori çifti için, p değerini önceden belirlenmiş alfa düzeyi $\alpha_{birleş}$ ile kıyasla.
 - a. Eğer p değeri $\alpha_{birleş}$ 'den büyük ise bu çifti bir tek kategori altında birleştir. X 'in yeni kategori kümesi için süreci Adım 1'den başlat.
 - b. Eğer p değeri $\alpha_{birleş}$ 'den küçük ise Adım 3'e git.
3. X 'in ve Y 'nin kategori kümesi için uygun Bonferroni düzeltmesini kullanarak, düzeltilmiş p değerini hesapla.
4. En küçük düzeltilmiş p değerine sahip X tahmin edici değişkenini seç (en önemli olan). Bunun p değerini önceden tanımlanmış alfa düzeyi $\alpha_{böl}$ ile kıyasla.
 - a. Eğer p değeri, $\alpha_{böl}$ değerinden küçük veya eşit ise düğümü X 'in kategori kümesini temel alarak böl.

b. Eğer p değeri, $\alpha_{böl}$ değerinden büyük ise düğümü bölme. Bu düğüm uç düğümdür.

5. Ağaç büyütme sürecini durma kuralları görülene kadar sürdür (SPSS, 2001).

4. YÖNTEM

4.1. Değişkenler

Borsa şirketlerinin sektörel risk profillerinin belirlenmesinde kullanılacak değişkenler Tablo 2’de verilmiştir.

Tablo 2. Araştırmada kullanılacak değişkenler ve tanımları

Değişkenler	Tanımlar
Sınıflandırma Değişkeni	
Sektör	Ana faaliyeti içeren sektör
Finansal Değişkenler	
Cari Oran	Dönen Varlıklar/ Kısa Vadeli Yab. Kay.
Likidite Oranı (Asit-Test Oranı)	(Kasa+Bankalar+Serbest Men. Kıy.+ Alacaklar)/ Kısa Vad. Yab. Kay.
Nakit Oranı	Hazır Değerler (Kasa+ Bankalar+ Ser. Men. Kıy.)/ Kısa Vad. Yab. Kay.
Borç Oranı	Borç/Toplam Varlık
Sermaye Yapısı Oranı	Borç/Özsermaye
Özsermaye Varlık Oranı	Özsermaye/Toplam Varlık
Uzun Vadeli Yabancı Kaynak	Uzun Vadeli Yab. Kay./ Devamlı Sermaye
Devamlı Sermaye Oranı	
Maddi Duran Var. Özkaynak	Maddi Duran Varlık/Özkaynak
Oranı	
Maddi Duran Varlık Uzun Vadeli	Maddi Duran Varlık/Uzun Vadeli Yab. Kay.
Yabancı Kaynak Oranı	
Dönen Varlık Devir Hızı	Net Satışlar / Ortalama Dönen Varlıklar
Duran Varlık Devir Hızı	Net Satışlar / Ortalama Duran Varlıklar
Varlık Devir Hızı	Net Satışlar / Ortalama Varlıklar
Alacakların Ortama Tahsil Süresi	Ortalama Alacaklar/ (Net Satışlar/360)

Kısa Vadeli Alacak Varlık Oranı	Kısa Vadeli Alacak/Varlık
Özsermayenin Karlılığı	Net Kar / Özsermaye
Varlıkların Karlılığı	Net Kar/ Toplam Varlık

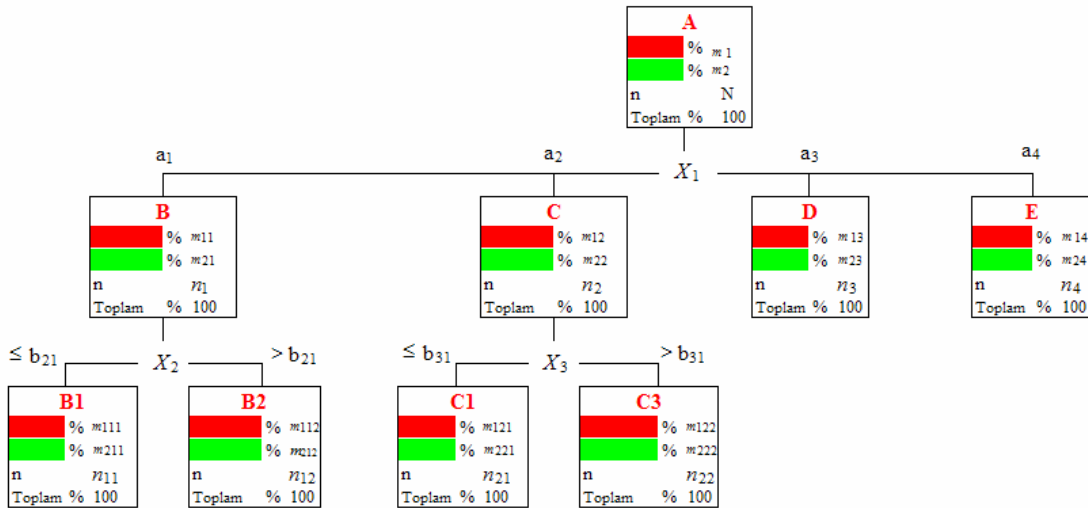
4.2. Model

4.2.1. Verinin hazırlanması

Varsayalım ki, $X_1, X_2, \dots, X_{N-1}, X_N$ Tablo 1'de verilen kesikli ve sürekli değişkenleri ve Y 'de finansal performansı gösterebilir. Her bir şirketin öncelikle sektör ayrımı gerçekleştirilir. Ardından, her bir şirketin, finansal değişkenleri dikkate alınarak, finansal durumu 'İyi' veya 'Kötü' olarak belirlenir. Sektörler ve finansal durum tanımlandıktan sonra, $k=1, 2, 3, \dots, m$ olmak üzere $X_1 \in \{a_k\}$ sektörler, $X_2 \in [a_2, b_2], \dots, X_N \in [a_N, b_N]$ finansal değişkenler ve $Y \in \{\text{Kötü}, \text{İyi}\}$ finansal durum olacak biçimde belirlenir.

4.2.2. Veri madenciliği yöntemi CHAID algoritmasının uygulanması

Y finansal performans hedef değişken ve $X_1, X_2, \dots, X_{N-1}, X_N$ açıklayıcı değişkenler olmak üzere CHAID karar ağaçları algoritması uygulanacaktır. Uygulama sonucunda, kırmızı çubuk kötü finansal performansı, yeşil çubuk ise iyi finansal performansı göstermek üzere Şekil 1'de örnek olarak verilen karar ağacı elde edilir.



Şekil 1. CHAID karar ağacı örneği

CHAID algoritması hedef değişken ve tahmin edici değişkenler arasında Ki-Kare Bağımsızlık Testi uygulayarak istatistiksel açıdan en güçlü ilişkiye sahip değişkenden başlayarak; hedef değişkenden, tahmin edici değişkenleri dallandırır. Şekil 1'deki karar ağacı örneğine bakıldığında, Y hedef değişkeni ile en güçlü ilişkiye X_1 tahmin edici değişkeni sahiptir. Şekil 1'den de görüldüğü gibi CHAID, diğer karar ağacı algoritmalarının aksine ikili değil çoklu olarak dallanmaktadır. Böylece, veri içerisindeki bütün önemli detayları ve ayrışmaları tespit etmek mümkün olmaktadır. Dolayısıyla, çalışma özünde bütün farklı risk profillerini tanımlamaktadır. Bu çalışmada kapsanan risk, şirketin finansal başarısızlığına yol açacak risk olmaktadır.

Şekil 1'den görüldüğü gibi X_1, X_2 ve X_3 olmak üzere sadece 3 değişken hedef Y değişkeniyle istatistiksel açıdan önemli ilişkiye sahiptir. Değişken tanımlamasında X_1 değişkeni sektör değişkeni olarak tanımlanmıştır ki, çalışmanın temel değişkenidir. X_1 değişkeninin Y değişkeni ile istatistiksel açıdan önemli ilişkiye sahip olmaması durumunda, X_1 değişkeni manuel olarak karar ağacına yerleştirilip dallanma her şekilde X_1 değişkeninden başlatılacaktır. Modelin anlaşılabilirliği için bu zorunlu koşulun modelde gerçekleştiği varsayılmıştır.

Şekil 1'deki karar ağacı dikkate alındığında:

- X_1 değişkeni Y hedef değişkeniyle istatistiksel açıdan en önemli ilişkiye sahiptir.
- X_2 değişkeni X_1 değişkeniyle $X_1 = a_1$ olması koşuluyla istatistiksel açıdan önemli ilişkiye sahiptir.
- X_3 değişkeni X_1 değişkeniyle $X_1 = a_2$ olması koşuluyla istatistiksel açıdan önemli ilişkiye sahiptir.

4.2.3. Risk profillerinin tanımlanması

Şekil 1 incelendiğinde 6 risk profili olduğu görülmektedir. B1, B2, C1, C2, D ve E yapraklarındaki risk profillerinin açıklamaları aşağıda verilmiştir.

1. Profil: Profil B1 değerlendirildiğinde,

- $X_1 = a_1$ ve $X_2 \leq b_{21}$ olmak üzere n_{11} şirket içerdiği,
- % m_{111} şirketin kötü finansal performansa sahip olduğu,

• % m_{211} şirketin iyi finansal performansına sahip olduğu, görülmektedir.

2. Profil: Profil B2 değerlendirildiğinde,

• $X_1 = a$ ve $X_2 > b_{21}$ olmak üzere n_{12} şirket içerdiği,
• % m_{112} şirketin kötü finansal performansına sahip olduğu,
• % m_{212} şirketin iyi finansal performansına sahip olduğu, görülmektedir.

3. Profil: Profil C1 değerlendirildiğinde,

• $X_1 = a_2$ ve $X_3 \leq b_{31}$ olmak üzere n_{21} şirket içerdiği,
• % m_{121} şirketin kötü finansal performansına sahip olduğu,
• % m_{221} şirketin iyi finansal performansına sahip olduğu, görülmektedir.

4. Profil: Profil C2 değerlendirildiğinde,

• $X_1 = a_2$ ve $X_3 > b_{31}$ olmak üzere n_{22} şirket içerdiği,
• % m_{122} şirketin kötü finansal performansına sahip olduğu,
• % m_{222} şirketin iyi finansal performansına sahip olduğu, görülmektedir.

5. Profil: Profil D değerlendirildiğinde,

• $X_1 \leq a_3$ olmak üzere n_3 şirket içerdiği
• % m_{13} şirketin kötü finansal performansına sahip olduğu,
• % m_{23} şirketin iyi finansal performansına sahip olduğu, görülmektedir.

6. Profil: Profil E değerlendirildiğinde,

• $X_1 = a_4$ olmak üzere n_4 şirket içerdiği,
• % m_{14} şirketin kötü finansal performansına sahip olduğu,
• % m_{24} şirketin iyi finansal performansına sahip olduğu,

görölmektedir.

Her bir profil ayrı ayrı değeriendirilirse:

Profil B1'e göre herhangi bir şirket için $X_1 = a_1$ ve $X_2 \leq b_{21}$ ise kötü finansal performansa sahip olma oranı veya bir başka deyişle risk oranı $R_{B1} = m_{111}$ olacaktır.

Profil B2'e göre herhangi bir şirket için $X_1 = a_1$ ve $X_2 > b_{21}$ ise kötü finansal performansa sahip olma oranı veya bir başka deyişle risk oranı $R_{B2} = m_{112}$ olacaktır.

Profil C1'e göre herhangi bir şirket için $X_1 = a_2$ ve $X_3 \leq b_{31}$ ise kötü finansal performansa sahip olma oranı veya bir başka deyişle risk oranı $R_{C1} = m_{121}$ olacaktır.

Profil C2'e göre herhangi bir şirket için $X_1 = a_2$ ve $X_3 > b_{31}$ ise kötü finansal performansa sahip olma oranı veya bir başka deyişle risk oranı $R_{C2} = m_{122}$ olacaktır.

Profil D'e göre herhangi bir şirket için $X_1 = a_3$ ise kötü finansal performansa sahip olma oranı veya bir başka deyişle risk oranı $R_D = m_{13}$ olacaktır.

Profil E'ye göre herhangi bir şirket için $X_1 = a_4$ ise kötü finansal performansa sahip olma oranı veya bir başka deyişle risk oranı $R_E = m_{14}$ olacaktır.

5. SONUÇ

Veri tabanı sistemleri içerisinde zaman içerisinde biriken verilerin değeriendirilmesi noktasından hareketle yola çıkılması, bugün en gelişmiş analitik yöntem olarak kabul edilen Veri Madenciliği'nin doğmasına yol açmıştır.

Veri madenciliği 1990'ların sonuna doğru başlayan ortaya çıkışını, 2000'lerden itibaren hızlandırmaya başlamış ve artarak hemen her alanda yaygın kullanıma sahip olmaya başlamıştır. Veri madenciliği, istatistiksel yöntemleri, yapay zeka, makine öğrenimi ve örüntü tanıma gibi disiplinlerle birleştiren metodolojisiyle; veri içerisindeki gizli kalmış, bilgi, örüntü ve ilişkilerin otomatik olarak keşfi ve elde edilen bilgilerin stratejik karar destek amaçlı kullanımı olarak tanımlanmaktadır.

Veri madenciliğinin temel karakteristiğini ortaya koyan ve diğer analitik yöntemlerden ayrışmasını sağlayan 'tahminsel' ve 'otomatik' olmak üzere iki sözcük esastır. Ayrıca, sürecin 'keşifsel' olması gerekmektedir.

Otomatik olarak tahminsel bilgi sağlayan veri madenciliği, zaman içerisinde büyük hacimde veriye sahip her alanda kullanım yaygınlığına kavuşmuştur. Finansal sistemler açısından bakıldığında ise ilk uygulamalar bankacılığa yönelik olarak ortaya çıksa da,

sermaye piyasaları ve hisse senetleri piyasaları özelinde de gün geçtikçe artan sayıda uygulama göze çarpmaktadır.

Artan rekabet koşullarının zorunlu kıldığı bir diğer kavram da 'risk' kavramıdır. Belirsizliği bilindir kılma çabası gün geçtikçe risk kavramının daha yaygın kullanımına yol açmaktadır. Önleyici tedbirler alma noktasında, mevcut ve gelecekteki durumu bilmenin olasılıksal ifadesi sayılabilen risk kavramı, subjektivite yerine, objektif normları esas kılmaktadır.

Borsa ve Düzenleyici Kurumlar perspektifinden bakıldığında veri madenciliğinin yansımaları gözetim ve denetim alanlarında görülmektedir. İnsanın elle yapacağı hesaplamaların Bilişim Teknolojileri yardımıyla otomatik olarak yapılması; zaman içerisinde tahminsel bilgiyi de bünyesini eklemiştir.

BASEL- II normlarının temel aracı olan risk kavramı, finans ve finans ile ilgili tüm kurumları bu yönde düşünmeye, çalışmaya ve uyumlulaştırmaya yol açmıştır. Otomatik tahminsel bilgi sağlayan veri madenciliğinin, risk uygulamalarında da başarıyla uygulanabileceği düşüncesi bu çalışmanın temel motivasyonunu oluşturmuştur.

Şüphesiz ki, konu risk olunca,

- Yatırımcı davranışları,
- Aracı kuruluş sermaye yeterliliği,
- Döviz-kuru, Borsa Endeksi ilişkisi,
- Düzenlemelerin etkisinin geri bildirimi,
- Manipülasyon ve içerden öğrenenlerin ticaretinin belirlenmesi,
- Yatırımcı suistimali,

vb. pek çok gözetim, denetim, hilekarlık tespit ve erken uyarı sistemiyle risk belirlemesine gitmek mümkündür. Buna karşın, bu çalışmada veri madenciliğinin risk belirlemesinde kullanımına bir örnek verilmesi istenmiş ve anlaşılabilirliği ön planda tutmak hedeflenmiştir.

Bu noktadan hareketle, Borsa Şirketleri esas alınmış ve pek çok dış etkiye açık olan Borsa Şirketleri'nin, düzenleme değerlendirmelerinde bütünsel olarak değil, sektörel ve hatta alt-sektör bazında farklılık gösteren risk düzeylerinin nasıl belirlenebileceğine yönelik bir yöntem sunmak hedeflenmiştir.

Çalışma neticesinde Borsa'da sektörel risk tespitine yönelik erken uyarı özelliği sağlayan ve öncü göstergeleri belirleyebilen bir model elde edilmiş ve modelin kullanım ve yorumuna yönelik açıklamalara yer verilmiştir. Çalışma, bundan sonraki sermaye piyasalarına yönelik analitik yaklaşımlarda yol gösterici olması ve yeni yöntemlerin sürekli olarak

gelişmiş ülkelerde uygulanıp, başarıya ulaştıktan sonra ülkemizde uygulanması yerine, zaman içerisinde Ülkemiz uygulamalarının gelişmiş ülke uygulamalarına örnek teşkil etmesi iddiası ve ümidiyle gerçekleştirilmiştir.

Bu çalışmanın devamı olarak sayısal uygulamalarla alternatif yöntemlerin mukayesesi ve veri madenciliğinin tartışmasız üstünlüğünün ortaya konulması hedeflenmektedir. Ayrıca, Sermaye Piyasası Kurulu'nun sorumluluk alanına giren diğer konularda da veri madenciliğine yönelik çalışmalar akademik-bilimsel platform perspektifinden bahse konu olacaktır.

6. KAYNAKLAR

- Altman, E. 1968. Financial Ratios, Discriminant Analysis And The Prediction Of Corporate Bankruptcy. *Journal of Finance* September. p:589-609.
- Altman, E., Haldeman, G. ve Narayanan, P. 1977. Zeta Analysis: A new model to identify bancruptcy risk of corporations. *Journal of Banking and Finance* June, p.29-54.
- Anandarajan, M., Picheng, L. ve Anandarajan, M. 2001. Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of Artificial Neural Networks. *International Journal of Intellegent Systems in Accounting, Finance and Management*, Vol: 10, p. 69-81.
- Apoteker, T. ve Barthelemythierry, S. 2005. Predicting financial crises in emerging markets using a composite non-parametric model. *Emerging Markets Review*, 6 (4), 363-375.
- Beaver, W. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research*, p. 71-111.
- Berson, A., Smith, S. ve Thearling, K. 2000. *Building Data Mining Applications for CRM*, McGraw-Hill, USA. 510 p.
- Boritz, E. J. ve Kennery, D. 1995. Effectiveness of Neural Network types for prediction of business failure. *Expert Systems with Applications* . Vol.9, p. 503-512.
- 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.
- Canbas, S., Onal, B. Y., Duzakin, H. G. ve Kilic, S.B. 2006. Prediction of financial distress by multivariate statistical analysis: The case of firms taken into the surveillance market in the Istanbul Stock Exchange. *International Journal of Theoretical & Applied Finance*, 9 (1), 133.

- Chan, N. H. ve Wong, H.Y. 2007. Data Mining of resilience indicators. IIE Transactions, 39, 617-627.
- 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.
- Chin-Sheng, H., Dorsey, R. E. ve Boose, M. A. 1994. Life insurer financial distress prediction: A Neural Network Model. Journal of Insurance Regulation, Vol: 13(2) (Winter), p. 131-168.
- Coats, P.K. ve Franklin, F.L. 1992. A Neural Network approach to forecasting financial distress. The Journal of Business Forecasting Methods & Systems, Vol: 10, p. 9-12.
- Coats, P.K. ve Franklin, F.L. 1993. Recognizing financial distress patterns using a Neural Network tool. Financial Management, Vol: 22 (3), p. 142-155.
- Collard, J. M. 2002. Is Your Company At Risk? Strategic Finance, 84 (1), 37-39.
- Çilli, H. ve Temel, T. 1988. Türk bankacılık sistemi için bir erken uyarı modeli. T.C. Merkez Bankası, Araştırma, Planlama ve Eğitim Genel Müdürlüğü, APE tartışma tebliği no: 8814.
- 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.
- Gaytan, A. ve Johnson, A.J. 2002. A review of the literature on early warning systems for banking crises. Central Bank of Chile, working papers no: 183.
- Gerni, C., Emsen, S. ve Değer, M.K. 2005. Erken uyarı sistemleri yoluyla Türkiye'deki ekonomik krizlerin analizi. Ekonometri ve İstatistik, 2; 39-61.
- Gönül, A. ve Eroğlu Z. A. 1999. Etkin bankacılık denetiminde temel prensipler: Türkiye ve diğer ülke uygulamaları. Devlet Planlama Teşkilatı. 93. Ankara
- Gunther, J.W. ve Moore, R.R. 2003. Early Warning Models in real time. Journal of Banking, 27 (10), 1979-2001.
- Hamer, M. 1983. Failure prediction: Sensitivity of classification accuracy to alternative statistical method and variable sets. Journal of Accounting and Public Policy. Vol: 2. p. 289-307.
- 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 bechmarking. Hospitals & Health Networks, Vol: 77, p. 37-44.
- Inegbenebor, A. U. 2006. Financing small and medium industries in Nigeria-case study of the small and medium industries equity investment scheme: Emprical research finding. Journal of Financial Management & Analysis; 19,1, 71-80.
- Jacobs, L.J. ve Kuper, G.H. 2004. Indicators of financial crises do work! An early-warning system for six Asian countries. International Finance, 0409001, 39.
- Jones, F. 1987. Current techniques in bankruptcy prediction. Journal of Accounting Literature 6.p. 131-164.
- Jones, S. ve Hensher, D.A. 2004. Predicting firm financial distress: A mixed Logit Model. The Accounting Review, Vol. 79 (4), p. 1011-1038.
- Kamin, S.B., Schindler, J., ve Samuel, S. 2007. The contribution of domestic and external factors to emerging market currency crises: An Early Warning System. International Journal of Finance and Economics, 12 (3), 317-322.
- Klersey, G.F. ve M.T. Dugan. 1995. Substantial Doubts: Using Artificial Neural Networks to Evaluate Going Concern, In Advanced In Accounting Information Systems. JAI Press, Greenwich.
- 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.
- Ko, P.C. ve Lin, P.C. (Eds. Ghosh, A. ve Jain, L.C.). 2005. Evolutionary Computation in Data Mining, Germany: Springer-Verlag Berlin Heidelberg, 249-263
- Kovalerchuk, B.ve Vityaev E. 2000. Data Mining in Finance. USA, Kluwer Academic Publisher, Hingham MA.
- Koyuncugil, A. S. 2004. Veri Madenciliği veya Bir Başka Deyişle 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.
- Koyuncugil, A. S. ve Özgülbaş, N. 2007. Detecting Financial Early Warning Signs In Istanbul Stock Exchange By Data Mining. IABE (International Academy of Business and Economics) Annual Conference. 14-17 Ekim 2007, Las Vegas, Nevada, ABD.
- 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 Support Systems, Vol:18 , p.63-73.
- Liu, S. ve Lindholm, C. K. 2006. Assessing early warning signals of currency crises: A Fuzzy Clustering Approach. Intelligent Systems in Accounting, Finance and Management, 14 (4), 179-184.
- 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.
- Mena, J. 2003. Investigative Data Mining for Security and Criminal Detection. USA: Elsevier Science.
- 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.
- 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.
- Pantalone, C. ve Platt, M. 1987. Predicting failures of savings and loan associations. *AREUEA Journal*, Vol: 15, p. 46-64.
- Ribic C. A. ve Miller T. W. 1998. Evaluation of Alternative Model Selection Criteria in The Analysis Of Unimodal Response Curves Using CART. *Journal of Applied Statistics*. Vol: 25. p: 685-98
- Rudolph, P. ve Hamdan, B. 1988. An analysis of post-deregulation savings and loan failures. *AREUEA Journal*. 16. 17-33.
- SPSS. 2001. *AnswerTree 3.0 User's Guide*, SPSS Inc. USA 226 p.
- 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.
- Tan, C. N. ve Dihadjo, H. 2001. A Study on using Artificial Neural Networks to develop an early warning predictor for credit union financial distress with comparison to the probit model. *Managerial Finance*. Vol: 27 (4), p. 56-78.
- Tan, Z. ve Quektuan, C. 2007. Biological Brain-Inspired Genetic Complementary Learning for stock market and bank failure prediction, *Computational Intellegence*, 23 (2), 236-242.
- Thearling, K. 2005. Web sayfası. [Erişim Tarihi: 18.06.2004] [http:// www.thearling.com/](http://www.thearling.com/)
- Türker M., Yarbaş E. ve Okay E. 2005. The Experience of SMEs in Turkey: A Compulsive Process
- Ward, T.J. 1993. Is the scaling measure used for cash flows important in predicting financially distressed firms? *Journal of Applied Business Research*. 9 (4) (Fall) 134-141.
- Ward, T.J. 1994. Cash flow information and the prediction of financially distressed mining, oil and gas firms: A comparative study. *Journal of Applied Business Research*. 10 (3) (Summer) 78-86.
- www.nyse.com. 2007. Web site: The New York Stock Exchange.
- www.set.or.th/en/index.html. 2007. Web site: The Stock Exchange of Thailand.
- Zavgren, C. 1985. Assessing the vulnerability to failure of American Industrial Firms: A Logistics Analysis. *Journal of Accounting Research*, Vol:22, p. 59-82.

Zmijewski, M. E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research (Supplement)* p.59-82.