

Veri Madenciliğinin Finansal Kararlarda Kullanımı*

Mehmet ÖZKAN

*Marmara Üniversitesi, İşletme Fakültesi, Muhasebe Finansman Anabilim Dalı,
mozkan@marmara.edu.tr*

Levent BORAN

Sorumlu Yazar, Yapı Kredi Bankası A.Ş., Bilgi Teknolojileri, levent.boran@yapikredi.com.tr

Öz

Çağımız bilgi çağı olup, her türlü bilgiye sahip olmayı gerektirmektedir. Veri içinden analizler yapılarak anlamlı örüntüler çıkartıldığında pek çok kimse tarafından bilinmeyen, rakiplerin önüne geçilmesini olanak veren anlamlı bilgiler elde edilebilir. İlgili yöntemler, şirketlerin finansal göstergeleri olan finansal tablolara uygulandığında, şirketlerin finansal davranışlarına ait faydalı bilgiler elde etmek mümkün olacaktır. Veri madenciliği, finansal analiz yöntemleri olarak çok yaygın kullanılan yatay analiz, dikey analiz, eğilim yüzdeleri analizi ve oran analizinin yanında beşinci finansal analiz yöntemidir. Veri madenciliğinin söz konusu bu analiz tekniklerine göre daha üstün olmasının nedeni, çok yoğun verileri kullanabilmesi ve pek çok kimse tarafından bilinmeyen bilgileri kullanıma sunmayı amaçlamasıdır. Veri madenciliğinin uygulanmasında en önemli iki kısıt; veri madenciliği ve veri madenciliğinin uygulanacağı alan bilgisine sahip uzman bulmaktaki zorluk ile kullanılacak bilgisayar yazılım ve donanımının oldukça maliyetli olmasıdır.

Anahtar Kelimeler: Finansal Tablo Analizi, Veri Madenciliği, Kümeleme Analizi.
JEL Sınıflandırma Kodları: G10, G17, G28.

Usage of Data Mining at Financial Decision Making

Abstract

The knowledge age requires controlling every kind of information. Recognition of patterns in data may provide previously unknown and useful information that can provide competitive advantages. If related techniques are applied on financial statements, it is possible to acquire valuable information about companies' financial situations. It is considered that data mining could be an alternative of common financial analysis techniques such as vertical analysis, horizontal analysis, trend analysis and ratio analysis. Against existing financial analysis methods, data mining provides some advantages, which are ability of manipulation of huge data and competence of obtaining previously unknown information. There exist two major constraints of data mining implementation that are lack of experts on both data mining and related domains and cost of computer software and hardware used.

Keywords: Financial Statement Analysis, Data Mining, Clustering Analysis.
JEL Classification Codes: G10, G17, G28.

* Bu çalışma, L. Boran (2012). "Veri Madenciliğinin Türk İşletmelerinin Finansal Tablolarına Uygulanması ve Uygulama Örneği" isimli doktora tez çalışmasından üretilmiştir.

1. Giriş

Günümüzde teknoloji çok büyük miktarlarda verinin elde edilerek, kaydedilmesini olanaklı kılmakta olup; bilgi çağında sayısal modellerle veri setleri içinde bulunan örüntü, eğilim ve aykırılıkların özetlenmesi, diğer bir deyişle; verinin bilgiye dönüştürülmesi daha büyük önem ve anlam kazanmıştır. Bununla beraber bu veriler içinde anlamlı ve yararlı bilgilerin elde edilme-üretile (işleme) sürecinde büyük zorluklar yaşanmaktadır. Geleneksel istatistiksel yöntemlerle büyük boyuttaki verilerin çözümlenmesi kolay olmamaktadır. Bu nedenle, verileri işlemek ve çözümlmek için, özel yöntemlere gereksinim duyulmuş olup; bunun sonucunda *Veri Madenciliği Yöntemleri* bu gereksinimi karşılamak üzere ortaya çıkmıştır.

Veri tabanı sistemlerinin gelişmesine paralel olarak firmalar artık çok daha fazla veri toplamakta olup, günümüzde bu faydalı bilgilerin toplanması endüstriyel faaliyetlerin odağı haline gelmiştir. Veri madenciliği bu bilgi keşfinin anahtar bileşeni olarak kullanılmaktadır (Kepili, 2010).

Veri madenciliğinin amacı, yoğun veriler içinden geleneksel yöntemlerle bulunamayacak, fark yaratacak bilgileri bulmanın yanı sıra, özellikle günümüzün rekabetçi koşullarına uygun olarak rakiplere ve emsallere karşı fark yaratmaktır. Bu işlem bir madencinin taş ve kaya parçaları arasından değerli bir madeni çıkartması ve işlemesi faaliyeti gibidir.

Öte yandan finansal tablo hazırlanması düşüncesi son yüzyılda ortaya çıkmış olup, sosyal bir gelişim göstermiştir. Finansal tablolar, bir şirketin geçmiş ve şimdiki performansları hakkında karşılaştırmalı bilgiler vermekte ve finansal tablolar aracılığı ile sunulan bilgiler, şirket yöneticileri, hissedarlar, tedarikçiler, kredi sağlayan kurumlar, sendikalar, kamu kurumları vb. gerçek ve tüzel kişiler/kurumlar için oldukça değerlidir. Bu bilgiler sayesinde, söz konusu kurumlar finansal tablo analizi yöntemlerini kullanarak bilgi alma gereksinimlerini karşılamakta, şirketin mevcut durumu hakkında bilgi sahibi olmakta ve gelecekteki finansal durumunu öngörmektedirler.

Son birkaç yılda veri madenciliği yöntemlerinin finans dünyasında oldukça sıklıkla kullanıldığı görülmektedir (Terzi, 2012, 169). Bu çalışmada, veri madenciliği teknikleri ile finansal tablo analiz yöntemleri birlikte ele alınarak incelenmiş ve veri madenciliği tekniklerinin finansal tablo analiz yöntemi olarak kullanılabilirliği değerlendirilmiş olup, çalışma yeni bir finansal analiz tekniği tanımlamayı amaçlamıştır. Bu çalışmayı önemli kılan üç neden şunlardır:

- **Birinci neden;** ekonomik ve finansal sistemlerin temel taşları olan işletmelerin; finansal tablolarının analizinde yeni bir yöntemin uygulamaya konulmasının amaçlanmasıdır. Bilindiği üzere günümüzün rekabetçi koşulları finansal tablo kullanıcılarının her kesimini daha net bilgilere ulaşmaya

zorlamakta olup, kullanılacak yöntemler yeni bilgilere ulaşılmasını kolaylaştıracak ve firmanın rakipleri karşısında avantaj ve ayrıcalıklara sahip olmasını sağlayacaktır.

- **İkinci neden;** finans ve bilgisayar bilimleri gibi, iki farklı disiplin arasında işbirliği ve paylaşımı sağlamak, daha güvenilir bilgileri üreterek, firma etkinlik ve üretkenliğini arttırmaktır. Sayısal ve matematiksel finans konularını bir araya getiren finans mühendisliği^β alanının gelişimine de katkıda bulunulması söz konusudur. Günümüzde her konuda uzmanlaşma büyük gelişim göstermekte olup, bunu disiplinler arası uzmanlaşma izlemektedir. Bilgisayar bilimi geleceğe ışık tutacak yazılım ve donanımları geliştirme konumunda önder iken, finans bilimi ise; ekonomileri ayakta tutan şirketlerin mali durumlarının yani varlık ve kaynak yapıları ile aralarındaki ilişkilerin asli göstergeleridir. Bir diğer ifadeyle, ekonomik (aktif yan) finansal (pasif yan) yapılarını ve bu yapılarda oluşan gelişim ve değişimleri göstermekte ve gösterimde finansal tablolardan yararlanılmaktadır. Buna bağlı olarak da, bu disiplinlerin birer temsilcisi olan finansal tablo analizi ve veri madenciliği birlikte incelenerek aralarında işbirliği ve paylaşımın sağlanma koşulları detaylı bir biçimde ve karşılaştırmalı olarak araştırılmıştır.
- **Üçüncü neden ise;** Türkiye’de akademik anlamda bu tip bilimsel bir çalışmanın bugüne kadar yapılamamış olmasıdır. Bugün veri madenciliği konusunda büyük bir eksiklik söz konusu olup, bu eksiklik henüz giderilememiştir. Kamu ve özel sektör şirketleri, özellikle pazarlama şirketleri ile bankalar veri madenciliğiyle ilgili yurt dışı destekli çalışmalardan yararlanmaktadır. Her ne kadar kamunun elinde oldukça geniş bir veri tabanı mevcut olsa dahi, özel sektörün bu konudaki gereksinimleri tam anlamıyla karşılanamamıştır. Bu nedenle de yabancı danışmanlık firmalarının verdiği hizmeti almak zorunda kalmaktadırlar.

Yukarıda belirtilen nedenlerle,

- Kullanılmayan büyük miktarlarda veri bulunması,
- Rekabetin şiddetinin artması ve etkisinin yaygınlaşması,
- Hazır teknoloji ve teknolojide sürekli yenilenme ve gelişim,

faktörleri “*Neden Veri Madenciliği?*” sorusunun yanıtıdır.

2. Temel Kavramlar

“*Muhasebe, finansal (mali) özellik taşıyan ticari işlemler ile kısmen de olsa benzeri ekonomik olayları, para cinsinden sistemli bir biçimde, saptama, toplama,*

^β Finans mühendisliği; ekonomi, bilgisayar bilimleri ve uygulamalı matematik gibi alanların araç ve bilgi birikimlerini kullanarak mevcut finansal problemlerin çözümünü arayan ve yenilikçi finansal ürünler ortaya koyan; ticari bankalar, yatırım bankaları, sigorta acenteleri ve yatırım fonları tarafından kullanılan finans alanıdır.

kaydetme, sınıflandırma ve anlamlı olarak özetleme işlemleriyle, ulaşılan sonuçları yorumlayarak ticari kişiliğin sahip ya da ortaklarına, yöneticilere ve bu kişilikle ilişkisi bulunan diğer kimselere, gereksinme duydukları uygun ve yararlı bilgileri sağlayan sanat, bilim ve uygulama sistemidir” (Altuğ, 2005, 5).

Finansal tabloların temelinde muhasebe kayıtları yer almaktadır. Finansal muhasebe kapsamında toplanan ve işlenen iş olguları, Genel Kabul Görmüş Muhasebe İlkeleri ve Muhasebe Standartlarına göre defterlere kaydedildikten sonra, belirli dönemlerde ilgililere sunulmak üzere finansal tablolara dönüşürler (Arat, 2005, 10). Bu nedenle finansal tablolar; *“Muhasebe sistemi içinde kaydedilen ve toplanan bilgilerin, belirli zaman aralıklarıyla bu bilgileri kullanacak olanlara iletilmesini sağlayan araçlardır”* biçiminde tanımlanmaktadır (Akdoğan ve Tenker, 2010, 4).

Finansal tablolar, büyük ölçüde benzerlik göstermekle birlikte; bazı sosyal, ekonomik ve hukuksal nedenlerle ulusal düzenlemelerin oluşturulması sırasında farklı ülkeler tarafından farklı finansal tablo kullanıcılarının gereksinimlerini dikkate alması gibi nedenlerle bazı farklılıklar gösterebilir. Söz konusu farklılıklar finansal tabloların unsurları için farklı tanımlamaların kullanılmasına neden olmuştur. Ayrıca söz konusu farklılıklar finansal tablolarda yer alan kalemlerin muhasebeleştirilmesinde farklı kriterlerin kullanılmasına ve bu kalemler için farklı ölçüm esaslarının benimsenmesine yol açmıştır (UFRS, 2008, 3). Ancak dünya genelinde bu farklılıklara rağmen birçok firma tarafından finansal tabloların (kısmen işletme içi olmak üzere) işletme dışı kullanıcılarının bilgi gereksinimlerini karşılama amacı değişmez. Bunun yanı sıra veri madenciliğinde kullanılacak örnek verilerin taşınması gereken bazı özellikler söz konusu olup, bu özellikler şunlardır (Sinanovic ve Gvozden, 2002, 27):

- Anlaşılabilirlik,
- Yenilik,
- Kullanılabilirlik,
- Geçerlilik,
- Önemlilik.

Ülkemizde Tekdüzen Hesap Planı çerçevesinde düzenlenecek finansal tablolar (Akgüç, 2011, 2);

- Bilânço,
- Gelir Tablosu,
- Satışların Maliyeti Tablosu,
- Fon Akım Tablosu,
- Nakit Akım Tablosu,
- Kar Dağıtım Tablosu,
- Öz Kaynaklar Değişim Tablosu,
- Net Çalışma Sermayesi Tablosudur.

Finansal tablolar analiz edilerek; hile arařtırmalarını, kredi kullanım kararlarını, yatırım kararlarını destekleyecek vb. yararlı bilgiler elde edilir. Finansal tablo analizlerinde amaç karar alma mekanizmasının somut bulgularla desteklenmesi ve karar almanın kolaylaştırılmasıdır. En çok kullanılan finansal tablo analiz yöntemleri; *eğilim yüzdeleri analizi*, *yatay analiz*, *dikey analiz* ve *oran analizi* yöntemleridir.

- Geleneksel finansal tablo analiz yöntemlerinden biri eğilim yüzdeleri analizi yöntemidir (Paramasivan ve Subramanian, 2009, 17). Bu yöntemde, temel yıl ilk kolona yerleştirilip, diğer yıllar sonraki kolonlara yerleştirilir (Gropal, 2008, 38). Sonrasında diğer yılların temel yıllara göre yüzdesel değişimi hesaplanır.
- Diğer bir finansal tablo analiz yöntemi, yatay analiz yöntemidir. Yatay analiz, şirketlerin finansal değişkenlerinin yıllara göre, değişimini izlememize yardımcı olur. Kalemlerin karşılaştırılması gelişim eğilimini gösterir. Bu sayede şirketin güçlü ve zayıf yönleri ile sorunlarının ortaya konulması amaçlanır. Bu tip analiz bir yıllık veri yerine verinin yıllara göre değişimine odaklanır (Gropal, 2008, 38).
- Dikey analiz yönteminde her kalemin toplam içindeki yüzdesi hesaplanır. Dikey analiz şirketlerin birbirleriyle karşılaştırılması durumunda oldukça faydalıdır. Büyüklük içindeki yüzdeye bakılarak, şirketlerin büyüklük farkı sorunu ortadan kaldırılmış olur (Rana, 2007, 207). Bilançonun aktif tarafında varlık toplamı 100 kabul edilip, her bir kalem bu değerle oranlanır. Benzer şekilde toplam varlığa eşit olan toplam yükümlülük tutarı da her yükümlülük ve özkaynak kalemiyle oranlanır (Kramer ve Johnson, 2009, 214).
- Oranlar yoluyla analiz (oran analizi) yönteminde, yaygın ve geniş kullanım alanı bulmuş oranlar finansal analiz araçları olarak kullanılmaktadır. Bir oran, iki niceliğin matematiksel olarak ilişkisini ifade eden bir göstergedir. Oranların hesaplanması kolay olmasına rağmen, elde edilen rakamların anlamlandırılması oldukça güçtür. Oranın anlamlı olması için, satış fiyatı ile maliyet ilişkisi gibi ekonomik olarak anlamlı ilişkileri açıklıyor olması gerekmektedir.

3. Veri Madenciliğinin Karar Almada Kullanımı

Ekonomik değeri olan herhangi bir maddeyi yeryüzüne çıkarıp, o maddeyi paraya dönüştürme işlemine "*Madencilik*" denir. Elmas, bor, kömür gibi farklı maddelerin madenlerden çıkarılması da farklı şekillerde gerçekleşmektedir. Örneğin elmasın, gizlendiği derinliklerden bulunup çıkartılması çok uzun ve zorlu uğraşlar gerektirmektedir. Yalnızca 1 karatlık bir pırlanta üretebilecek ham elmas kütlenin çıkarılması için, yaklaşık 250 ton toprağın büyük bir sabırla kazılması gerekmektedir. Bugünkü ileri teknoloji dahi, bu işlemi kolaylaştırmaya yetmemektedir (Gürsoy, 2009, 1).

Öte yandan veri madenciliği veri noktaları-kaynakları ve düzenlenmesi-işlenmesi arasındaki ilişkilerin belirlenmesinde, mevcut veriler bir dizi işlemin iletim

gücünün büyük oranlarda kullanılmasıdır. Bunun içinde algoritmalar, matematiksel teknik ve istatistikler, makinesel öğrenme uygunlukları vb. düzenlenerek ve örnek bulgular kullanılarak veri bankaları otomatik olarak aranmaktadır. Bu arama-veri madenciliği, veri bankalarından bilginin keşfedilmesidir (“Was ist Daten-Bergbau?”, 2013).

Veri madenciliği yapısal bir teknik olarak endüstrinin genelinde özellikle *Kredi Kartı Kullanımında, AVM’lerde, Finansal Hizmet İşletmelerinde, Bankalarda, Tele Pazarlamacılık ve Doğrudan Satışlarda, Hava Yolu Taşımacılığında, Üretim İşletmelerinde-İmalat Firmalarında, Telefon-İletişim Hizmetlerinde, Sigortacılıkta* (Sinanovic ve Gvozden, 2002, 19-21) yaygın bir biçimde kullanılmakta ve “*KDD = Knowledge Discovery in Data Bank = Wissenentdeckung in Datenbanken*” olarak tanımlanmaktadır. 1989 yılında Gregory Pratesky-Shapiro tarafından kullanıcılarına sunulan KDD, bilgiler için gerekli verileri ve ham verileri analiz ederek sonuçları açıklayan bir teknik olarak geliştirilmiştir. Bu yöntem kullanıcılarına algoritmalarından örnekler bulmalarına olanak vermekle beraber verilerin genelinde daima aynılık ve doğruluk söz konusu değildir. Çünkü hataları ortaya koyan tüm yöntemler birleştirilmektedir (“Was sind der unterschiedliche Daten Bergbau-Verfahren?”, 2013).

Veriler, veri tabanlarından çıkarılmayı bekleyen değerli madenler gibidir. Veri madenciliği teknikleri veriden bilginin çıkarılması noktasında devreye girerek etkilerini göstermektedir. Veriden anlamlı ilişkiler ve örüntüler (patterns) çıkarma sürecine yazında “*Veri Madenciliği*”, “*Bilgi Çıkarımı*”, “*Bilgi Keşfi*”, “*Veri Arkeolojisi*” ve “*Veri Şablon İşleme*” gibi isimler verilmektedir. Veri madenciliği tanımını daha çok istatistikçiler, veri analizcileri ve yönetim bilişim sistemleri kullanıcıları tercih etmektedirler (Gürsoy, 2009, 271).

Veri madenciliği eldeki mevcut verilerden üstü kapalı çok net olmayan önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılması olup; bu da kümeleme, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir (Işıklı, 2009). Veri madenciliğinin kullanım amaçları (Kurtbauknecht, 2002, 14);

- Teşebbüsün bölüm ve toplam verimliliklerinin belirlenmesi,
- Müşteri istek ve gereksinimlerinin anlaşılması-tespiti,
- Rantabl-karlı müşterilerin ayrılması ve yenilerinin oluşturulması,
- Müşteri sadakatinin-bağlılığının yükseltilmesi,
- Yatırım üzerinden getiri (ROI) yükseltilmesi ve reklam giderlerinin azaltılması,
- Müşteri risklerinin belirlenmesi,
- Web sayfası rantabilitesi-karlılığının artırılması,
- Zararlı kullanımlar ve zararların yanıtıcı etkilerinin ortadan kaldırılması vb.

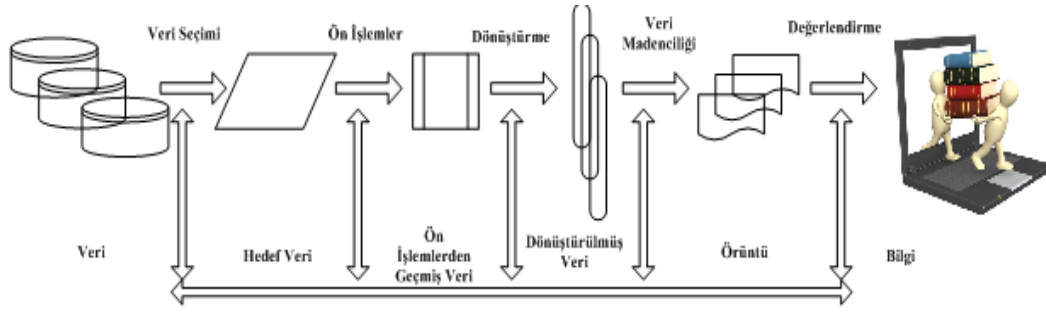
olup, veri madenciliği tekniklerinin sınıflandırılmasına temel oluşturmuştur. Bu nedenle aşağıda yer alan amaçlar veri madenciliği tekniklerinin isimleriyle aynıdır (Refaat, 2007, 7-8).

- **Sınıflandırma:** Bu tip problemlerde, veri tabanındaki her bir kayıt önceden tanımlanmış sonlu belirli bir sınıf veya kategori etiketi altında toplanır. Örneğin bir banka, her bir müşterisini yeni bir kredi kartıyla ilgilenip ilgilenmeyeceğine göre sınıflandırmak isteyebilir.
- **Tahminleme:** Bu tip problemlerde, sürekli değişkenlerin bilinmeyen değeri tahmin edilir. Örneğin, vergi otoriteleri, ev sahiplerinin gelirlerini tahmin etmekle ilgilenirler. Tahminleme probleminin mümkün olan çıktı sayısı, tanımı gereği sonsuzdur.
- **Öngörme:** Gelecekteki bir değer tahmin edilmesidir. Bu konudaki tipik örnekler hisse senedi fiyatlarının tahmini, ticari mal fiyatlarının tahmini ve hava kirliliği endeksinin gelecekteki değerinin tahmini şeklinde sıralanabilir.
- **Kümeleme:** Daha az benzer elemanlardan oluşan bir kümenin daha fazla benzer elemanlardan oluşan alt grup veya kümelere bölünmesi işidir. En tipik örnekleri müşteri ve pazarların birbirine benzer veya farklı özellikler gösterenlerinin belirlenmesidir.
- **Benzerlik Analizi:** Benzerlik analizinin diğer isimleri pazar sepeti analizi ve ilişki analizidir. Buradaki amaç, birlikte artan ya da azalan değerleri bulmaktır. Bu değerler; ürünler, işlemler, operasyon sıraları veya veri tabanına yazılan objeler olabilirler. En tipik örneği, bir süpermarkette birlikte alınan ürünlerin tespiti için yapılan pazar sepeti analizidir. Örneğin tavuk ve barbekü sosunun birlikte satın alınıp alınmadığı ile ilgilenen bir analist benzerlik analizini kullanacaktır.

Veri madenciliği uygulama aşamaları en yalın anlamda;

- Veri madenciliği öncesi işlemler,
- Veri madenciliği ve
- Veri madenciliği sonrası işlemler

olarak sınıflandırılmaktadır. Veri madenciliği öncesi işlemler, veri seçimi, verinin ön işleme ve dönüştürme olarak; veri madenciliği sonrası işlemler ise bilginin çıkartılması/kontrolü ve değerlendirilmesi şeklinde tanımlanmaktadır. Veri madenciliği uygulaması verinin seçimi ile başlayıp, verinin değerlendirilmesiyle son bulan tekrarlı süreçtir. Bunun anlamı, veri madenciliğinin en doğru sonuca ulaşılan kadar uygulama adımlarının tekrarlanması gerekliliğidir. Aşağıdaki akış diyagramında, veri madenciliğinin uygulama aşamaları gösterilmiştir.



Şekil 1: Veri Madenciliği Uygulama Aşamaları

Kaynak: Yada (2002)

4. Uygulama Örneği

Madenciliğin sonuçlarının istatistiksel olarak doğruluk düzeyi, veri seçimi ve kullanılan madencilik tekniğine bağlıdır. Veri madenciliği tekniklerinin uygulanması sonucu bulunacak sonuçlar ve ulaşılabilecek yargılar kullanılan veri kadar sağlıklı olacaktır (Kudyba ve Richard Hoptroff, 2001, 25). Bu çalışmada kullandığımız mali tabloların kaynağı İMKB (yeni ismiyle Borsa İstanbul-BIST) olup, bu tabloların kullanılmasının seçiminde; *mali tabloların hazırlama esaslarına uygunluğu, standartlara uygunluğu ve denetlenmiş olmasının* İMKB tarafından garanti edilmiş olması en önemli etkidir.

Analize Türk İmalat Sanayi Şirketlerinden 2009 yılı için 152 adedi, 2010 yılı için 151 adedi ve 2011 yılı için 148 adedi dahil edilmiştir. 2008 yılından itibaren değerlendirildiğinde 2009 yılında şirketlerin tamamı faaliyetlerine devam etmiş olup, 2010 yılında şirketlerin %99'u, 2011 yılında şirketlerin %97'si faaliyetlerine devam etmiştir. Bu sayısal verilerden 2008'de yaşanan kriz sonrası şirket sayısı anlamında istikrarın devam ettiği anlaşılmaktadır.

Bilim dallarında sağlanan ilerleme sonucu pek çok analiz tekniği kullanıcılarına sunulmuştur. Analizler sırasında tekniklerin bir ya da birkaçından yararlanılmakta olup, analistler doğru tekniği seçmekle yükümlüdürler. Bu uygulamada kullanılan analiz teknikleri değerlendirildiğinde aşağıdaki gibi benzer bir tablo oluşmaktadır.

- Çalışmada üç yıllık verilerin kullanılması ve yıllara göre değişiminde incelenmiş olmasından dolayı “**Dinamik Finansal Analiz**” yapılmıştır.
- Veri kullanımı dikkate alındığında, verileri kullanılan şirketlerin personeli olmaması ve verilerin İMKB aracılığıyla elde edilmiş olması nedeniyle “**Dış Finansal Analiz**” uygulanmıştır.
- Ayrıca finansal tablo kalemlerinin yıllara göre değişimi ya da aktif toplamı/pasif toplamı/satışlar gibi temel bir finansal tablo kalemlerine göre oranı yerine birbiriyle ilişkili olabilecek finansal kalemlerin oranlarının kullanılması itibarıyla “**Oran Analizi**” araç olarak kullanılmıştır.

- Tahmin edici modeller arasında, sınıflandırmada gruplar önceden belirli iken, kümeleme analizinde önceden belirli değildir. Cari oran, asit test oranı gibi oranların normlarının sektörden sektöre, ülkenin içinde bulunduğu duruma göre, zaman içinde değiştiği bilindiği için çalışmamızda veri sınıflarının önceden belli olduğu modeller kullanılmamıştır. Veri sınıflarının analiz sırasında belirlendiği **Kümeleme Analizi** veri madenciliği analiz yöntemi olarak tercih edilmiştir.
- Çalışma sırasında OPTICS, DBScan, K-Means ve EM algoritmaları denenmiştir. Ağırlıklı olarak **EM Algoritması** kullanılıp, **K-Means Algoritmasından** da iki algoritma sonuçlarının karşılaştırılmasında yararlanılmıştır.

Ön işleme aşamasında hesaplamayı zorlaştıracak sayıdaki özellikler makul bir sayıya çekilip, aykırılıklar ayıklanmıştır. Özellik sayısının azaltılması, gereksiz olduğu düşünülen ve sorguları yavaşlatan bazı bilgilerin silinmesi suretiyle veri temizleme aşamasıdır. (Hsu, 2008, 2585). Özellikleri bir matrisin kolonları, bilgileri de satırları olarak kabul edildiğinde, özellik sayısının azaltılması bu matrisin kolon sayısını azaltmaya, aykırılıkların ayıklanması ise yine bu matristeki satır sayısını azaltmaya karşılık gelir. Yaygın veri sayısını azaltma yöntemleri aşağıda verilmiştir (Şentürk, 2006, 17-18).

Veri Birleştirme: Bu yöntemde verilerin anlamlarını kaybetmeyecek şekilde düzenlenmesi ve verilerin birleştirilmesi amaçlanmaktadır.

Boyut İndirgeme: Çok değişkenli istatistiksel analizlerde değişkenlerin sayısını azaltmada kullanılan oldukça popüler bir yöntemdir. Özellikle bilişim teknolojilerinin ve buna bağlı olarak da veri toplama teknolojilerinin gelişmesi sonucu değişken sayısı hızla artmaktadır. Bu artış ihtiyaç olan değişkenlerin yanında yapılacak analizlerde gerekli olmayan, fazla veya işe yaramayan değişkenler de bulunmasına neden olmaktadır. Veri madenciliği çok büyük veri ambarlarından yüzlerce binlerce değişken arasından anlamlı ilişkiler çıkarmaya çalışırken performansı olumsuz etkileyecek işe yaramayan değişkenlerin elenmesi bu sürecin başarıya ulaşması bakımından önemlidir.

Veri Sıkıştırma: Orijinal verileri temsil edecek indirgenmiş veya sıkıştırılmış veriler elde etmeye dayalı bir yöntemdir. Veri şifreleme veya veri dönüştürme, veri sıkıştırmada uygulanan yöntemlerdir. Eğer sıkıştırma işlemi sonucunda elde edilen sıkıştırılmış veriler, orijinal veri kümesini hiçbir bilgi kaybı olmaksızın temsil edebilirse, kullanılan sıkıştırma tekniği kayıpsız olarak adlandırılmaktadır. Orijinal veri kümesine göre, kısmen bir bilgi kaybı varsa; bu amaçla kullanılan sıkıştırma yöntemi de kayıplı olarak adlandırılmaktadır.

Kesikleştirme: Bazı veri madenciliği algoritmaları sadece kategorik değişkenlere göre işlem yapabilmektedirler. Bu tür algoritmaların kullanılması gerektiğinde sürekli verilerin kesikli verilere dönüştürülmesi gerekmektedir. Bu şekilde sürekli

verilerin kesikli verilere dönüştürülmesiyle elde edilen kategorik değerler, orijinal veri değerlerinin yerine kullanılmaktadırlar.

Çalışma sırasında ağırlıklı olarak bilanço ve gelir tablosu kalemleri kullanılmış olup, bilgi yönünden bu iki tablonun dayanağı ticari defter kayıtları ve belgeler, şekil şartları yönünden dayanağı tek düzen hesap planıdır. Oranların bir kısmı sadece bilanço, bir kısmı sadece gelir tablosu kalemlerinin kendi aralarındaki ilişkileri gösterirken, bir kısmı da bilanço ile gelir tablosu arasındaki etkileşimi göstermektedir.

Veri madenciliğinde verinin sütun sayısının fazla olmasından çok satır sayısının fazla olması daha doğru bir analiz yapılmasını kolaylaştıracaktır. Bu nedenle Cari Oran, Asit Test Oranı, Toplam Borç/ Özsermaye Oranı (Likidite Oranları), Kısa Vadeli Finansal Borçlar / Satışlar Oranı, Ticari Borç Devir Hızı, Alacak Devir Hızı (Mali Yapı Oranları), Alacakların Tahsil Süresi, Stok Devir Hızı, İşletme Sermayesi Devir Hızı (Faaliyet Oranları), Kar / Satışlar Oranı, Kar / Özsermaye Oranı (Karlılık Oranları), Faaliyet Kaldıraç, Finansal Kaldıraç, Toplam Kaldıraç (Kaldıraç Oranları) oranları diğerlerine göre daha önemli bulunarak, bu oranlar üzerinde çalışılmıştır.

Ayrıca veri içerisinde aykırı verinin belirlenip ayıklanması analizin sağlıklı olması açısından oldukça önem taşıyan bir husustur. Veri madenciliği analizlerinin kalitesi tüm analizlerde olduğu gibi kullanılan verinin kalitesiyle çok yakından ilişkili olup, verinin kalitesi analizin kalitesini belirlemektedir. Aykırılıklar uç değerler olup; veri aralığı limitlerine yakın olabileceği gibi uzak da olabilir. Aykırılıkların tanımlanması veri girişinde hataları temsil ettikleri için oldukça önemlidir. Aykırılıklar hata değil, doğru veriler olsalar bile, çoğu yöntem aykırı verilere karşı hassas olup, düzgün olmayan sonuçlar üretebilir (Larose, 2005, 34).

Aykırı şirketlerin yıllık değişimi, ülkenin içinde bulunduğu ekonomik durumun yıllara göre değişimi hakkında, yine aykırı şirketlerin sektörlere göre dağılımı, sektörlerin birbirlerine kıyasla buldukları durum hakkında analistlere bilgi sağlayacakları açıktır. Yapılan analize imalat sanayine mensup 2009 yılı için 152 şirket, 2010 yılı için 151 şirket ve 2011 yılı için 148 şirket dahil edilmiş olup, 2009 yılı için 43, 2010 yılı için 40 ve 2011 yılı için 43 adet şirket aykırı bulunmuştur. Aykırı şirketler hesaplanan her bir oran için “**Ortalama ± 2 X Standart Sapma**” aralığının dışında kalan şirketler olarak belirlenmiş olup, şirketin tek bir aykırı oranı olsa bile analiz dışında bırakılmıştır.

Tablo 1: Veri Seti İçinde Aykırı Şirketlerin Yıllara Göre Dağılımı ve Yüzdeleri

	2009		2010		2011	
	Sayı	Yüzde	Sayı	Yüzde	Sayı	Yüzde
Aykırı Şirket Sayısı	43	28%	40	26%	43	29%
Değerlendirmeye Alınan Şirket Sayısı	109	72%	111	74%	105	71%
Toplam Şirket Sayısı	152	100%	151	100%	148	100%

Yıllar itibariyle analiz yapılırken değerlendirmeye alınan şirketlerin toplam şirketler içindeki payları yüzdesel olarak %72, %74, %71 olup aralarında önemli bir fark bulunmamaktadır. Bu sonuç ise, 3 yıl süresince kullanılan verilerin istikrarlı bir yapıya sahip olduğunu göstermektedir.

Değişkenler arasındaki aralıkların düzenlenmesi amacıyla Z-Skor Standardizasyon, Minimum-Maksimum Normalizasyon, Ondalık Ölçeklendirme Teknikleri aracılığıyla dönüştürme işlemi yapılabilir. Dönüştürme, diğer adıyla normalizasyonun amacı, gelir gibi çok büyük rakamlar ile cari oran gibi sayısal değeri küçük rakamlar arasındaki aralıkların küçültülmesidir.

Veri madencileri, sonuç üzerinde her bir değişkenin etkisini standartlaştırmak için sayısal değerleri normalize etmelidirler. Bu konuda normalizasyon işlemi yapan teknikler mevcut olup, aşağıda üçü incelenmektedir (Larose, 2005, 36).

Min.-Maks. Normalizasyon: Min.-Maks. normalizasyon işlemi alan değerinin minimum değerden ne kadar büyük olduğuna bakar ve bu farkları sıralar (Larose, 2005, s.36).

$$X^* = \frac{(X - \min(X))}{\text{aralık}(x)} = \frac{(X - \min(X))}{(\text{maks}(X) - \min(X))}$$

Ondalık Ölçeklendirme: Ondalık ölçeklendirme ondalık noktasını değiştirirken orijinal dijital değerini korur. Tipik ölçeklendirmede değer -1 ile +1 aralığında kalır. Aşağıdaki eşitlik (i) durumu için (v) özelliğinin değeri v(i) iken ondalık ölçeklendirmeyi tanımlar (Kantardzic ve Zurada, 2005, 24-25).

$$v^1(i) = V(i)/10^k$$

$v^1(i)$ en küçük k değeri yani $\max(|v^1(i)|) < 1$ için ölçeklenmiş değerdir.

Öncelikle maksimum $|v^1(i)|$ değeri veri seti içinden bulunur. Daha sonra ondalık nokta yeni, ölçeklendirilmiş ve mutlak değeri 1'den küçük oluncaya kadar hareket ettirilir. Sonrasında bölen diğer v(i)'lere uygulanır. Örneğin, en büyük değer 435

ve en küçük değer -834 iken özelliğin maksimum mutlak değeri 0.834 ve tüm v(i) için bölen 1 000 alınır (k=3) (Kantardzic, 2003, 24-25).

Z-Skor Standardizasyonu: Z-Skor Standardizasyonu, dünyada pek çok istatistiksel analizde yaygın olarak kullanılır. Alan değer ile ortalama değer arasındaki farkı standart sapma ile oranlar (Larose, 2005, s.37).

$$X^* = \frac{X - \text{ortalama}(X)}{\text{standart sapma}(X)}$$

Çalışmada aykırı oranlara sahip şirketler ayıklandıktan sonra, Minimum-Maksimum Normalizasyon ve Z-Skor Normalizasyon formülleriyle normalize değerleri hesaplanmıştır. Ancak, gerçek değer aralığından çok fazla sapma olduğu için normalize değerler yerine gerçek değerler alınmıştır. Örneğin cari oran 2009 yılı için 0,61-15,60 aralığında değer alırken Minimum-Maksimum Normalizasyon sonrası 0-1 aralığında, Z-Skor Normalizasyon sonrası -0,87-5,88 aralığında değişmiştir. Hesaplanan oranlara normalizasyon formüllerinin uygulanmış hali kümeleme analizi sonrası bulunan rakamların ilk değerlerle ve yazındaki standartlarla karşılaştırılma olanağını ortadan kaldırdığı için hesaplanan finansal oranlar normalize edilmemiştir.

Veri madenciliğinin uygulama aşamasında kullanılacak ham veriler yukarıda belirtilen yöntemlerle işlenmiş, veri madenciliği uygulanabilecek olgunluğa getirilmiştir. Veri madenciliği uygulamasının çıktısı tıpkı veriler gibi, sayısal değerlerden oluşuyor olacaktır. Bu sayısal değerler tek başlarına bir anlam ifade etmezler. Bu nedenle alan uzmanları tarafından bu sayıların yorumlanarak, yargılara dönüştürülmesi gerekmektedir.

Veri madenciliği uygulamalarında SPSS'in "Clementine", Silicon Graphics'in "MineSet", IBM'in Intelligent ve SAS Institute'ün "Enterprise Miner" yazılımları kullanılabilir (Cios ve Diğerleri, 2007, 22). Ayrıca "WEKA", "RapidMiner" gibi açık kaynak kodlu uygulamaları kullanmak da mümkündür.

Çalışma sırasında OPTICS, DBScan, K-Means, EM algoritmaları denenmiştir. OPTICS ve DBScan algoritmalarında kümelerin yeterince ayrışmadığının görülmesi üzerine bu algoritmaların kullanılmasından kaçınılmıştır. K-Means algoritması sadece merkezci elemanı döndürmesi nedeniyle sadece EM algoritmasının sonuçlarını değerlendirmek için kullanılmıştır. En son olarak denen EM algoritmasıyla kümeleşmenin yeterli sayıda olduğu görülmüştür. EM algoritması uygulanması yoluyla 2009 yılı için 3, 2010 yılı için 4, 2011 yılı için 2 adet küme elde edilmiştir.

Her bir yıl için oluşan kümelerin alt kümeleri bulunması istenildiğinde EM algoritmasıyla ikinci seviyede 2009 yılı için 7, 2010 yılı için 6, 2011 yılı için 5 adet kümenin oluştuğu ve 10 adet firmadan daha az şirket içeren kümelerin alt

kümelere bölünemediği görülmüştür. İkinci kırılımda kümelerin çok fazla parçalanmaması nedeniyle, ilk seviyenin yeterli olacağı kanaati ile sonuçların yorumları ilk kırılımda oluşan kümeler üzerinden yapılmıştır (bkz. Ek 1).

EM algoritmasının uygulanması yoluyla bulunan kümelerin homojenliğini kontrol etmek için sektörel dağılım, aktif toplamı ve satışlar, hisse senetlerinin fiyat artışı ve EM algoritması ile K-Means algoritmasının karşılaştırılması olmak üzere dört adet kontrol noktası belirlenmiştir. Kontrol sırasındaki beklentimiz, oluşan kümelere benzer sektörler ait şirketlerin yer alması ve aktif toplamı/satışlar/hisse senedi fiyat artışı gibi rakamlar açısından her bir küme içerisinde yer alan şirketlerin homojenlik göstermesidir.

İlk kontrol noktası olarak sektörel dağılıma bakılmış olup, tutarlı bir durum gözlenmiştir. Sektör bazında şirketler kümelere dağılmak yerine, aynı sektördeki şirketler ağırlıklı olarak belirli kümelere toplanmışlardır. Aynı sektörden olan şirketlerin aynı kümelere yerleşmesi sektörel karakteristiği şirketlerin yansıttığını göstermesinin yanı sıra, bu durum kümelerin doğru oluştuğunun da bir kanıtıdır (bkz. Ek 2, 3 ve 4).

İkinci kontrol noktası olan, aktif toplamı ve satışlar dağılımı da oluşan kümelerin homojenliğini desteklemiştir. Ayrıca 2009 ve 2010 yılları için şirketlerin aktif toplamları tüm kümeler için 0 – 200.000.000 TL aralığında kalmış olup, 2011 yılı için bir miktar şirketin aktif toplamları 200.000.000 TL- 400.000.000 TL aralığına sarkmıştır (bkz. Ek 5, 6 ve 7).

0 – 200.000.000 TL aralığında, oluşturulan kümeler arasında aktif toplamı ve satış gelirleri kalemleri dağılımı açısından çok büyük farklar olmayıp, daha spesifik sonuca ulaşabilmek için aralıkların daha küçük tutulması gerektiği düşünülmüştür. Daha derine inebilmek için oluşan kümeler ile aktif toplamı ve satışlar rakamlarına göre 0 – 200.000.000 TL aralığına giren şirketler tekrar incelenerek, finansal durumu daha kötü olan şirketlerden oluşan kümeler için bu iki değer alt sınırlarda toplandığı saptanmıştır.

Üçüncü kriter olarak, hisse senetlerindeki fiyat artışı belirlenmiştir. 2009 yılı için Küme 1 içindeki şirketlerin hisse senetlerine yatırım yapanlar daha fazla kar etmiş olup, bu durum 2010 yılında için Küme 2 ve Küme 3 için geçerlidir. 2011 yılı için durum bir hayli farklı olup, zarar ettiren şirket sayısı Küme 1’de daha fazladır (bkz. Ek 8, 9 ve 10).

Son kriter iki algoritmanın karşılaştırılmasıdır. Bu karşılaştırılmaya bağlı olarak EM ve K-Means algoritma sonuçlarının örtüşmediği, diğer üç kritere göre farklı bir sonuç verdiği tespit edilmiştir. EM algoritmasıyla elde edilen ortalama ve standart sapma değerleri kullanılarak “Ortalama \pm Standart Sapma” aralığı ile K-Means algoritmasıyla elde edilen ağırlık merkezi karşılaştırıldığında, ağırlık merkezi değerinin EM algoritmasıyla elde edilen aralığın dışında kaldığı

gözlenmiştir. Bu oran “**Ortalama \pm 2 x Standart Sapma**” aralığı için daha düşük olmuştur (bkz. Ek 11).

Yorumlama aşamasına girdi olabilecek veriler, veri madenciliğinin uygulanması aşamasının çıktıları olan sayılar ve görselleştirme sonucu elde edilen grafik ve diyagramlar olabilir. Elde edilen örüntünün yorumlanmasında kesin yargılara varılabiliyor olması, karar destek sistemleri açısından daha çok tercih edilir. Eğer yorumlar siyah ve beyaz alanlardan ziyade gri alanlara isabet ediyorsa, verinin gözden geçirilmesi, kullanılan tekniğin gözden geçirilmesi gibi yollara gidilip döngü tekrarlanır.

Bu inceleme sırasında şirket sayısı onun altında olan sektörler birinci derecede önemli olmadığı için dikkate alınmamış olup, Türk İmalat Sanayinin verileri içinde aykırılıklar incelediğinde, “Dokuma, Giyim Eşyası” sektörü ve “Deri ve Gıda, İçki ve Tütün” sektörü şirketleri her üç yıl için diğer sektörlerle göre daha fazla aykırı şirket içermektedir. Ayrıca bu sektörlerle ait şirketler kümelerin kesişen kısımlarında yer almış ve sınıflandırılmalarında sorun yaşanmıştır. Bunun anlamı, gerek yatırımcılar, gerekse kredi sağlayıcıların bu sektördeki şirketlerle çalışırken daha dikkatli olmaları gerekliliğidir. Bu sektörlerin daha fazla aykırı şirket içermesinin nedenleri arasında,

- Ürün sayısının bu sektörlerde fazla olması,
- Ürün fiyatlarının farklı olması,
- Üretimlerinin kolay olması,
- Bu ürünlerin temel ihtiyaç olmaması,
- Ürün fiyatlarının çok sık değişmesi ve fiyat dalgalanmalarından kolay etkileniyor olması vb. sayılabilir.

Ayrıca kümeler arasında kesişen alanlarda kalan şirketler sonuçların yorumlanmasını güçleştirmiştir.

İmalat sanayi şirketlerini kümelere ayırarak bazı yargılara varmaya çalışırken, EM algoritması ile elde edilen kümeler için birer ortalama değer ve standart sapma değeri elde edilmiştir. K-Means algoritması ise küme sayısı ve kümelerin ağırlık merkezi değerlerini vermiştir. Ortalama değer ve standart sapma değeri ile kümelerin sınırlarını belirlemek mümkün iken, ağırlık merkezi değeri ile kümelerin sınırlarını belirlemek mümkün değildir. Dolayısıyla EM algoritmasının verdiği sonuçları bazı karşılaştırmalar yoluyla kontrol etme şansı söz konusu iken, K-Means algoritmasının ürettiği sonuçların küme sınırlarının belirlenememesi nedeniyle tek başına kontrol etmek mümkün olmayacaktır. K-Means algoritmasının ürettiği sonuçların ancak başka bir algoritma desteği ile değerlendirilmesi olasıdır. Bu nedenle bu çalışmada K-Means algoritmasının ürettiği ağırlık merkezi değerlerini, EM algoritmasının ürettiği sonuçlarla karşılaştırarak her iki algoritmanın ne kadar paralel (ya da farklı) sonuçlar ürettiği

kontrol edilmiştir. Bu inceleme sonunda beklenilen aksine her iki algoritmanın ürettiği sonuçların örtüşme oranının “**Ortalama ± Standart Sapma**” aralığı için %70-%90 olarak saptanmış olup, tüm veri madenciliği algoritmalarının tüm uygulamalar için kullanılamayacağı vurgulanmıştır.

EM ve K-Means algoritmalarının ürettiği sonuçlar değerlendirerek, K-Means algoritmasıyla hesaplanmış “**Ortalama ± Standart Sapma**” aralığının dışında kalan ağırlık merkezi değerinin en çok satışlar kalemini içeren finansal oranlarda görüldüğü saptanmıştır. Bunun nedeni ise gelir tablosunda yer alan satışlar değerinin çok geniş bir aralık içinde değer almış olmasıdır.

Yapılan kümeleme çalışmasında 2008 yılında yaşanan krizin etkileri de görülmektedir. 2008 yılında belirlenen kümelerin dışında kalan 25, 2010 yılında 27 şirket varken, 2011 yılında sadece 7 şirket bulunmaktadır. 2011 yılında sınıflandırılmayan şirketlerin sayılarının diğer yıllara göre az olması, yaşanan kriz sonrası kriz yaşayan ülkelerin ekonomilerinin iyileşme eğilimine girdiğinin göstergesidir. Ayrıca 2011 yılında sadece iki kümenin oluşması 2008 yılında yaşanan krizin etkilerinin bir ölçüde geçtiğinin göstergesidir.

Bu çalışmada oluşturulmuş kümeler üzerinde çelişki oluşturan tek nokta kullanılan iki algoritmanın farklı sonuçlar oluşturması olup, bu sonuç analiz sırasında seçilecek algoritmanın doğru seçilmesi gerektiğinin ne kadar önemli bir husus olduğunu göstermektedir. Oluşan kümelerin mümkün olduğunca kesişmiyor olması ve yorumlanabilecek sayıda küme oluşması gerekmektedir. Bu çalışmada makul sayıda küme oluşmasına karşın kümeler arasında kesişmeler görülmüştür. Bunun anlamı kullanılan değerlerin normal bir dağılım göstermesidir.

5. Sonuç ve Öneriler

Finansal analizin amacı, işletmelerle ilgili olarak işletmenin öncelikle genel durum, varlık ve kaynak yapısının daha sonra ise ayrıntılı bir biçimde ele alındığında likidite durumu, işletmenin karlılık durumu, işletmenin sermaye yapısı, işletme aktiflerini kullanım durumu ve işletme hakkında önemli eğilimlerle ilgili bilgileri sağlamaktır (Ceylan ve Korkmaz, 2008, 41). Bu çalışmada da bu durum dikkate alınarak çalışma sürecinde şirketlerin ilgili finansal oranları hesaplanmış, finansal oranlar kullanılarak şirketler gruplandırılmış ve şirketler kümeleme çalışması sonucunda aynı grupta yer alan diğer şirketler ile karşılaştırılarak analiz sonucu değerlendirilmiştir.

Bu çalışmanın ilk adımında ilgilenecek veri seti belirlendikten sonra, veriler içindeki aykırılıklar belirlenmiştir. Bir sonraki aşamada veriler üzerine normalizasyon teknikleri uygulanmıştır. Normalizasyon işlemi sonunda, verinin gerçek aralığından uzaklaştığı görülmüştür. Bu durum yorumlamada zorluklara yol açacağı için normalizasyondan vazgeçilmiştir. Daha sonrasında ise verilere

bilgisayar programları yardımıyla OPTICS, DBSCAN, K-Means ve EM veri madenciliği algoritmaları uygulanmıştır.

En son adımda belirlenen kontrol noktalarına göre oluşan kümelerin doğruluğu kontrol edilmiştir. Bu aşamada aykırı şirketlerin ağırlıklı olarak aynı sektörlerden çıktığı, aynı sektördeki şirketlerin ağırlıklı olarak aynı kümelere yer aldıkları tespit edilmiştir. Yine aktif toplamı, satışlar ve hisse senedi fiyat artışı dikkate alınarak incelendiğinde benzer durumun oluştuğu görülmüş olup, tüm bu bulgular yapılan analizi desteklemektedir. Diğer bir kontrol noktası olan farklı algoritmaların ürettiği sonuçların karşılaştırılması ise; istenildiği gibi bir sonuç vermemiş olup, algoritma değiştiğinde sonuçların da değiştiği gözlenilmiştir.

Muhasebe denetimi çalışmalarında işletmelerin mali tablolarının, mali durumu ve faaliyet sonuçlarını doğru ve dürüst, genel kabul görmüş muhasebe ilkelerine ve yasal düzenlemelere uygun olarak yansıtıp yansıtmadığı konusunda bir görüş oluşturulmaya çalışılır (Bozkurt, 2006, 27). Uygulama aşamasında yapılan aykırılık analizi durumu olağandışı olan şirketler hakkında bilgi vermiştir. Bazı aykırılık analizi uygulamalarının konusu, sadece bu şirketlerin finansal tablolarında aykırı durumların tespiti olabileceği gibi, bu çalışmada aykırılık analizi sadece aykırı şirketlerin yapılan analizin dışında bırakılabilmesi amacıyla kullanılmıştır.

Uygulanan algoritmalar, çalışma için verileri gruplandırmıştır. Bir grubun bazı yönleri ön plana çıkarken, diğer bir grubun farklı yönleri ön plana çıkabilmektedir. Bu şekilde değerlendirildiğinde finansal tablo ilgililerinin ilgi alanlarına girebilecek şirket kümelerinin oluşturulması mümkün olup, bu sayede finansal tablo ilgilileri doğru gruplara odaklanarak, istedikleri bilgileri alarak zaman ve diğer maliyet unsurları açısından tasarruf sağlayabilirler.

Bu çalışmanın sağladığı bir diğer avantaj da oran analizinde olduğu gibi önceden belirlenmiş, dönemsel veya sektörel olarak değişebilecek kriterlerin temel alınması yerine her analizin kendi değerlendirme kriterlerini oluşturması ve bunları kullanılmasıdır. Örneğin kriz ortamlarında bu referans değerlerine bakıldığında tüm şirketlerin durumu kötü görünebilir. Ama bu referans değerlerinin dinamik olarak hesaplandığı bu analizde bu tip kriz ortamlarında bile gerçekten iyi durumda olan şirketleri belirlemek mümkündür.

Bu çalışmada yapılan analizin benzeri üç yıl ve yılbaşına yaklaşık 150 ve daha fazla şirket için geleneksel yollardan yapılması halinde günlerce sürmesi söz konusu olacaktır. Oysa burada yapılan analizde finansal tablolardan belirli değerlerin çekilmesinin ardından saatlerle ifade edilebilecek bir sürede analiz gerçekleştirilmiştir. Yapılan analiz sayesinde finansal açıdan kaç tip şirketin mevcut olduğu ve hangi şirketin hangi gruba dahil olduğu hızlı bir biçimde saptanmıştır.

Yatay analiz, dikey analiz ve oran analizlerinde olduğu gibi finansal tablolara veri madenciliği uygulandığında da yorumlama aşamasında uzman (ihtisas sahibi) görüşüne ihtiyaç vardır. Finansal tablolara veri madenciliği uygulaması sırasında hem veri madenciliği, hem de finans alanında uzman kişilerin bilgi ve yorumlarına ihtiyaç duyulmaktadır. Oysa diğer finansal analiz yöntemleri dört işleme dayandığı için sadece finans alanında uzmanlaşmış kişiler tarafından uygulanabilir. Bu durum da veri madenciliği uygulamalarının maliyetini arttırmaktadır.

Veri madenciliği uygulamaları hem veri madenciliği uzmanlığı hem de üzerinde çalışılan konuyla ilgili uzmanlık gerektirmektedir. Bu iki uzmanlığı bir arada bulduracak uzmanları bulmak oldukça güç olup, bu durumun sonucu olarak uygulamada birçok zorluklarla karşılaşmaktadır.

Veri madenciliği uygulamaların mutlaka bilgisayar kullanımı gerekmektedir. Bunun nedeni kullanılan algoritmalarla yapılan hesaplamalar oldukça karmaşık ve fazla sayıdadır. Ayrıca veri madenciliği uygulamalarında çok fazla veri yer almaktadır. Bu iki durum bilgisayar kullanımını bir zorunluluk haline getirmiştir. Veri madenciliğinin ilk ortaya çıktığı 90'lı yıllarda bilgisayarlar çok pahalı olduğu için veri madenciliği kullanımının yaygınlaşması gecikmiştir. Günümüzde bilgisayar ve bilgisayar programlarının fiyatları kabul edilebilir seviyeye inmiştir. Yine de bu konu veri madenciliği uygulamaları için bir kısıt oluşturmaktadır.

Gelecek çalışmalarda veri madenciliğinin farklı teknikleri finansal oranlara uygulanabilir, finansal oranlar likidite oranları, borçlanma oranları, faaliyet oranları, karlılık oranları, kaldıraç oranları şeklinde bölümlendirilerek ayrı ayrı uygulanabilir ve hatta finansal oran yerine bazı bilanço ve gelir tablosu kalemleri analiz içinde kullanılabilir. Elde edilecek sonuçlar finansal analistleri, veri madenciliği ile finansal tabloların yorumlanması konularında yeni ufuklara yönlendirebilir.

Sonuç olarak; ayrıca bu hususta akademik çalışmalara daha fazla önem verilmesi bu çalışmaların Kamu otoritesince özendirilerek teşvik edilmesi, araştırmaların ve araştırma çalışmalarının arttırılması, veri madenciliği eğitim-öğretiminin yaygınlaştırılması ve desteklenmesi ile daha büyük yararlar sağlanacaktır.

Ekler

Ek 1: Şirketlerin EM Algoritması Kullanılması Yoluyla Oluşan Kümelere Dağılımı

2009				2010				2011			
Kırılım 1		Kırılım 2		Kırılım 1		Kırılım 2		Kırılım 1		Kırılım 2	
Küme 1	28	Küme 1.1	7	Küme 1	7	Küme 1.1	7	Küme 1	64	Küme 1.1	20
		Küme 1.2	19			Çakışan	0			Küme 1.2	33
		Çakışan	2			Küme 2.1	29			Çakışan	11
Küme 2	24	Küme 2.1	13	Küme 2	35	Küme 2.2	2	Küme 2	34	Küme 2.1	2
		Küme 2.2	9			Çakışan	4			Küme 2.2	30
		Çakışan	2			Küme 3.1	25			Küme 2.3	2
Küme 3	32	Küme 3.1	17	Küme 3	40	Küme 3.2	15	Çakışan	7	Çakışan	0
		Küme 3.2	9			Çakışan	0				
		Küme 3.3	3	Küme 4	2	Küme 4.1	2				
		Çakışan	3			Çakışan	0				
Çakışan	25			Çakışan	27						

Ek 2: Ortalama \pm Standart Sapma Formülüne Göre 2009 Yılı Firmaların Grup Dağılımı

	Küme 1	Küme 2	Küme 3	Çakışan
DİĞER İMALAT SANAYİİ	100%			
DOKUMA, GİYİM EŞYASI VE DERİ		14%	21%	64%
GIDA, İÇKİ VE TÜTÜN	21%	7%	50%	21%
KAĞIT VE KAĞIT ÜRÜNLERİ, BASIM VE YAYIN	50%	25%	13%	13%
KİMYA, PETROL KAUÇUK VE PLASTİK ÜRÜNLER	11%	32%	47%	11%
METAL ANA SANAYİ	10%	50%	10%	30%
METAL EŞYA, MAKİNE VE GEREÇ YAPIM	18%	24%	35%	24%
ORMAN ÜRÜNLERİ VE MOBİLYA	50%			50%
TAŞ VE TOPRAĞA DAYALI	54%	17%	21%	8%

Ek 3: Ortalama \pm Standart Sapma Formülüne Göre 2010 Yılı Firmaların Grup Dağılımı

	Küme 1	Küme 2	Küme 3	Küme 4	Çakışan
DİĞER İMALAT SANAYİİ			100%		
DOKUMA, GİYİM EŞYASI VE DERİ		40%	47%		13%
GIDA, İÇKİ VE TÜTÜN		42%	42%		17%
KAĞIT VE KAĞIT ÜRÜNLERİ, BASIM VE YAYIN	11%	22%	22%		44%
KİMYA, PETROL KAUÇUK VE PLASTİK ÜRÜNLER	5%	30%	40%	5%	20%
METAL ANA SANAYİ		58%	25%		17%
METAL EŞYA, MAKİNE VE GEREÇ YAPIM		22%	54%	5%	18%
ORMAN ÜRÜNLERİ VE MOBİLYA					100%
TAŞ VE TOPRAĞA DAYALI	26%	21%	11%		42%

Ek 4: Ortalama \pm Standart Sapma Formülüne Göre 2011 Yılı Firmaların Grup Dağılımı

	Küme 1	Küme 2	Çakışan
DİĞER İMALAT SANAYİİ		100%	
DOKUMA, GİYİM EŞYASI VE DERİ	55%	18%	27%
GIDA, İÇKİ VE TÜTÜN	69%	23%	8%
KAĞIT VE KAĞIT ÜRÜNLERİ, BASIM VE YAYIN	63%	38%	
KİMYA, PETROL KAUÇUK VE PLASTİK ÜRÜNLER	68%	26%	5%
METAL ANA SANAYİ	82%	18%	
METAL EŞYA, MAKİNE VE GEREÇ YAPIM	73%	16%	11%
ORMAN ÜRÜNLERİ VE MOBİLYA	50%	50%	
TAŞ VE TOPRAĞA DAYALI	35%	65%	

Ek 5: 2009 Yılına Ait Veri ile Elde Edilen Kümelerin Aktif ve Satışlar Toplamına Göre Dağılımı (x 1.000.000 TL)

		Küme 1	Küme 2	Küme 3	Çakışan
0 - 200	Aktif Toplamı	14	12	20	13
	Satışlar	17	14	19	16
200 - 400	Aktif Toplamı	9	5	3	6
	Satışlar	7	3	4	5
400 - 600	Aktif Toplamı	2	3	3	1
	Satışlar	1	2	1	1
600 - 800	Aktif Toplamı	0	1	1	2
	Satışlar	2	3	5	1
800 -	Aktif Toplamı	3	3	5	3
	Satışlar	1	2	3	2

Ek 6: 2010 Yılına Ait Veri ile Elde Edilen Kümelerin Aktif ve Satışlar Toplamına Göre Dağılımı (x 1.000.000 TL)

		Küme 1	Küme 2	Küme 3	Küme 4	Çakışan
0 - 200	Aktif Toplamı	3	29	17	1	12
	Satışlar	6	31	21	1	10
200 - 400	Aktif Toplamı	3	3	8	0	5
	Satışlar	1	2	6	0	7
400 - 600	Aktif Toplamı	1	1	5	0	1
	Satışlar	0	0	3	0	3
600 - 800	Aktif Toplamı	0	0	3	1	2
	Satışlar	0	0	1	1	2
800 -	Aktif Toplamı	0	2	7	0	7
	Satışlar	0	2	9	0	5

Ek 7: 2011 Yılına Ait Veri ile Elde Edilen Kümelerin Aktif ve Satışlar Toplamına Göre Dağılımı (x 1.000.000 TL)

		Küme 1	Küme 2	Çakışan
0 - 200	Aktif Toplamı	28	13	4
	Satışlar	31	18	5
200 - 400	Aktif Toplamı	11	11	1
	Satışlar	7	9	0
400 - 600	Aktif Toplamı	3	5	0
	Satışlar	5	1	0
600 - 800	Aktif Toplamı	6	1	0
	Satışlar	4	2	0
800 -	Aktif Toplamı	16	4	2
	Satışlar	17	4	2

Ek 8: 2009 Yılı Hisse Senedi Karlılığının Oluşturduğumuz Kümeler Arasında Dağılım

	Sayı	- (%100)	(%100) - 0	0 - %100	%100 - %200	%200 -
Küme 1	28		7%	57%	32%	4%
Küme 2	24		21%	38%	33%	8%
Küme 3	32		6%	34%	44%	16%
Çakışan	25			60%	20%	20%

Ek 9: 2010 Yılı Hisse Senedi Karlılığının Oluşturduğumuz Kümeler Arasında Dağılım

	Sayı	- (%100)	(%100) - 0	0 - %100	%100 - %200	%200 -
Küme 1	7		29%	57%	14%	
Küme 2	35		26%	57%	11%	6%
Küme 3	40		20%	60%	15%	5%
Küme 4	2		50%	50%		
Çakışan	27		11%	78%	11%	

Ek 10: 2011 Yılı Hisse Senedi Karlılığının Oluşturduğumuz Kümeler Arasında Dağılım

	Sayı	- (%100)	(%100) - 0	0 - %100	%100 - %200	%200 -
Küme 1	64		64%	30%	3%	3%
Küme 2	34		68%	29%	3%	
Çakışan	7		29%	71%		

Ek 11: “Ortalama \pm Standart Sapma” ve “Ortalama ± 2 x Standart Sapma” Formüllerine Göre 2009-2010-2011 Yılları İçin EM ve KMeans Algoritma Sonuçlarının Karşılaştırılması

		\pm Standart Sapma	± 2 x Standart Sapma
2009	Aralık İçinde	57%	71%
	Aralık Dışında	43%	29%
2010	Aralık İçinde	59%	86%
	Aralık Dışında	41%	14%
2011	Aralık İçinde	64%	89%
	Aralık Dışında	36%	11%

Kaynakça

- Akdoğan, N. ve Tenker, N. (2010). *Finansal Tablolar ve Mali Analiz Teknikleri*. Ankara: Gazi Kitabevi.
- Akgüç, Ö. (2011). *Mali Tablolar Analizi*. 14. Baskı, İstanbul: Arayış Basım ve Yayıncılık.
- Altuğ, O. (2006). *Maliyet Muhasebesi*. 12. Baskı, İstanbul: Türkmen Kitabevi.
- Arat, M.E. (2005). *Finansal Analiz Aracı Olarak Oranlar: İlkeler & Yorumlar*. İstanbul: Marmara Üniversitesi Nihad Sayar Eğitim Vakfı.
- Bozkurt, N. (2006). *Muhasebe Denetimi*. 4. Baskı, İstanbul: Alfa Basım Yayınları.
- Ceylan, A. ve Korkmaz, T. (2008). *İşletmelerde Finansal Yönetim*. 10. Baskı, Bursa: Ekin Basım Yayın Dağıtım.
- Cios, K.J., Pedrycz, W., Swiniarski, R.W. ve Kurgan, L.A. (2007). *Data Mining: A Knowledge Discovery Approach*. USA: Springer Science Business Media, LLC.

- Gopal, R. (2008). *Financial Management*. New Delhi: New Age International Ltd.
- Gürsoy, U.T.Ş. (2009). *Veri Madenciliği ve Bilgi Keşfi*. 1. Baskı, Ankara: Pegem Akademi.
- Hsu, J. (2008). Data Mining and Business Intelligence: Tools, Technologies and Applications. (J. Wang, Ed.). *Data Warehousing and Mining: Concepts, Methodologies, Tool and Applications* içinde (2584-2613), USA: IGI Global.
- Işıklı, B., <http://www.burakişikli.wordpress.com> (Erişim Tarihi: 19.08.2013)
- Kantardzic, M.M. ve Zurada, J. (2005). *Next Generation of Data-Mining Applications*. New Jersey: Institute of Electrical and Electronics Engineers Inc.
- Kepil, S. <http://www.serdarkepili.wordpress.com> (Erişim Tarihi: 19.08.2013)
- Kramer, B.K. ve Johnson, C.W. (2009). *Financial Statements Demystified*. USA. McGraw Hill Companies, Inc.
- Kudyba, S. ve Hoptroff, R. (2001). *Data Mining and Business Intelligence: A Guide to Productivity*. USA: Idea Group Publishing.
- Larose, D.T. (2005). *Discovering Knowledge In Data: An Introduction to Data Mining*. New Jersey: John Wiley and Sons Inc.
- Paramasivan, C. ve Subramanian, T. (2009). *Financial Management*. New Delhi: New Age International Limited Publisher.
- Rana, G. (2007). *Production and Financial Management*. Delhi: Oxford Book Company.
- Refaat, M. (2007). *Data Preperation for Data Mining Using SAS*. San Francisco: Morgan Kaufmann Publishers.
- Sinanovic, E. ve Gvozden, S. (2002). *Seminar-Data Mining*. Kurt Bauknecht: Institut für Informatik und wirtschaftsinformatik BWZ.-Universität Wien, Seminar Business Intelligence SS 2002.
- Şentürk, A. (2006). *Veri Madenciliği Kavram ve Teknikler*. Bursa: Ekin Basım Yayın Dağıtım.
- Terzi, S. (2012). *Hileli Finansal Raporlama: Önleme ve Tespit – İMKB İmalat Sanayiinde Bir Araştırma*. İstanbul: Beta Yayıncılık.

UFRS (2008). *Uluslararası Finansal Raporlama Standartları(IFRS(IAS) ile Uyumlu TMS*. TMSK Yayınları: 3.

Yada, K. (2002). The Future Direction of Active Mining in the Business World. (H. Motoda Ed.), *Active Mining: New Direction of Data Mining* içinde (3-9), Amsterdam: IOS Press.

“Was ist Daten-Bergbau?” 2013 <http://Fragenantworten.com> (Erişim Tarihi: 20.08.2013)

“Was sind der unterschiedliche Daten-Bergbau-Verfahren?” 2013 <http://Fragenantworten.com> (Erişim Tarihi: 20.08.2013)