

## Borsa İstanbul'da Sektör Sınıflandırmasının Kümeleme Analizi İle Belirlenmesi

Sümeyra GAZEL\*  
Veli AKEL\*\*

### ÖZET

*Literatür incelendiğinde, finansal veriler üzerinde veri madenciliği tekniklerinin sıkça kullanıldığı dikkat çekmektedir. Bu çalışmada Borsa İstanbul'da (BIST) farklı sektör endeksleri içinde yer alan hisse senetlerinin, geçmiş fiyat verilerine göre kümeleme analizine tabi tutulduğunda aynı sektör içerisinde yer alıp almadığı sorusuna cevap aranmıştır. Bu amaçla, çeşitli sektörlerde yer alan 70 farklı hisse senedinin 2012-2015 dönemine ait haftalık verilerinden hareketle Hiyerarşik Yığınsal Kümeleme analizi kullanılmıştır. Sonuç olarak, bazı hisse senetlerinin BIST100'deki sektör sınıflandırmasına benzer bir şekilde kümelendiği ve Ward's Yönteminin diğer bağlantı teknikleri ile karşılaştırıldığında kümelemeyi en iyi gerçekleştiren yöntem olduğu tespit edilmiştir.*

**Anahtar Kelimeler:** Borsa İstanbul, BIST, Veri Madenciliği, Kümeleme Analizi.

**JEL Sınıflandırması:** D53, C38.

### *The Determination Of Sector Classification Using Cluster Analysis In Borsa Istanbul* **ABSTRACT**

*In the literature, the use of data mining techniques in the financial data has attracted more attention, especially in recent years. This study aims to determine whether the shares involved in the Istanbul Stock Exchange by sector based on the core activities shows the same sector classification or not when subjected to cluster analysis based on past price data. The 70 Stocks, which are traded in the Istanbul Stock Exchange and located in a variety of different sectors, were investigated with weekly data using Hierarchical Agglomerative Clustering Analysis for the period from 2012 to 2015. The results indicate that some stocks are clustered in a manner similar to the sector classification of BIST100 and Ward's Method is the best one when compared to other linking techniques.*

**Keywords:** Istanbul Stock Exchange, Data Mining, Clustering Analysis.

**Jel Classification:** D53, C38.

\* Yrd. Doç. Dr. Sümeyra Gazel, Bozok Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, sumeyra.gazel@bozok.edu.tr

\*\* Doç. Dr. Veli Akel, Erciyes Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, veliakel@erciyes.edu.tr

## 1. GİRİŞ

Son yıllarda veri madenciliği teknikleri birçok disiplinde olduğu gibi finans alanında da yaygın bir şekilde kullanılmakta ve sonuçlarının geçerliliği tartışılmaktadır. Hisse senedi fiyatları, döviz kurları ve banka iflaslarının tahmin edilmesi, finansal risklerin belirlenmesi ve yönetilmesi, kredi derecelendirmesi, borç yönetimi gibi farklı finansal konularda, sağlıklı bir değerlendirme yapmak ve büyük hacme sahip veri setlerinden yorumlanabilir, karşılaştırılabilir yeni veriler üretebilmek için veri madenciliği tekniklerinden faydalanılmaktadır. Veri madenciliğinde, karar ağacı, yapay sinir ağları, destek vektör makineleri ve kümeleme analizi en çok kullanılan teknikler arasındadır.

Veri madenciliği tekniklerinin finans alanında kullanılması konusunda izlenmesi gereken yol, Cai vd. (2012) tarafından şu şekilde özetlenmiştir. İlk olarak, büyük veri setleri arasındaki gizli ilişkileri açık bir şekilde ortaya koyabilmek için *ilişki kuralları* belirlenmelidir. Daha sonra, her bir veri setindeki nesnelerin daha önceden tanımlanan kategorilere atanması olarak tanımlanacak *sınıflandırma* yaklaşımı, hem tanımlayıcı hem de tahmin modelleri açısından faydalı sonuçlar üretmektedir. Bir sonraki aşama ise *kümeleme* olup bir veri setindeki benzer nesnelerin daha önceden *tanımlanmamış* küme veya sınıflara tıpkı sınıflandırmada olduğu gibi atanması söz konusudur. Kümeleme analizi, büyük hacimli bir veri setini sıkıştırmak suretiyle veri setinin özünü, esasını yakalamak için iyi bir başlangıç noktası olarak değerlendirilebilir. Son olarak, finansal veriler arasındaki ilişkileri tespit etmek için sınıflandırma, optimizasyon ve dinamik programlama gibi çeşitli veri madenciliği tekniklerinden faydalanmak mümkündür. Finansal sistem hakkında çıkarımlarda bulunabilmek adına bu tekniklerden yararlanmak piyasa katılımcılarının işlerini kolaylaştırmaktadır.

Türkiye’de ilk olarak 26 Aralık 1985 tarihinde İstanbul Menkul Kıymetler Borsası’nın (İMKB), kurulması ile hayat bulan sermaye piyasası finansal sistem içerisinde çok önemli bir role sahiptir. İMKB, 2013 yılında Borsa İstanbul A.Ş.’ye (BIST) dönüştürülmüş ve borsanın Nisan 2013’de Altın Borsası, Mayıs 2013’de ise Vadeli İşlemler ve Opsiyon Borsası ile birleşme işlemleri tamamlanmıştır. BIST, yeni dönemde tüm sermaye piyasası araçları ve kıymetli madenlerin alınıp satıldığı, spot ve vadeli işlem piyasalarının aynı çatı altında birleştiği sermaye piyasalarının en önemli kurumu haline gelmiştir.

Borsa İstanbul’da işlem gören her bir hisse senedi belirli bir sektör içinde yer almaktadır. Bir şirketin hangi sektörde yer aldığı ana faaliyet konusuna başka bir ifade ile gelirinin büyük bir kısmını hangi iş alanından elde ettiğine bağlıdır. Eğer bir grup içindeki hisse senetlerinin fiyat dalgalanmaları birbiri ile ilişkili ise ya da farklı gruptaki hisse senetlerinin fiyat dalgalanmaları birbiri ile ilişkisiz ise ya da aralarında çok güçlü bir korelasyon varsa bir zaman serisi kümelenmesinden söz edilebilir. Hisse senetlerinin kümeleme analizi sektör içi ve sektörler arası trendlerin etkisini nicelleştirmek, sınıflandırılmamış hisse senetlerini tespit etmek ve portföyleri değerlendirmede kullanılabilir. Bu çalışmada, BIST’te işlem gören bazı hisse senetlerinin geçmiş fiyatları dikkate alınarak her bir hisse senedinin ana faaliyet konusuna göre dâhil olduğu sektörel kategorilerin yalnızca geçmiş fiyat verileri kullanılarak elde edilip edilmeyeceğinin tespit edilmesi amaçlanmıştır.

Çalışmanın bundan sonraki kısmında, finans alanında kümeleme analizi çalışmalarının tartışıldığı literatür bölümü yer almakta olup izleyen bölümde ise veri seti ile araştırmanın yöntemi hakkında açıklayıcı bilgiler sunulmuştur. Analiz sonuçları ve ampirik bulgular dördüncü kısımda tartışılmış, beşinci bölümde ise elde edilen sonuçlara ve diğer çalışmalarla olan benzerlik ve farklılıklara yer verilmiştir.

## 2. LİTERATÜR

Veri madenciliği büyük ölçekli veri setleri arasından bilgiyi süzme ya da yararlı bilgileri ayıklama süreci olarak değerlendirildiğinde, finansal veriler gibi büyük bir veri yığınına sahip olan finansal sistem hakkında çıkarımlarda bulunabilmek adına veri madenciliği tekniklerinden yararlanmak önemli kabul edilmektedir. Veri madenciliği tekniklerini finansal verilerle birleştiren literatürdeki belli başlı çalışmalar aşağıda özetlenmiştir.

Da Costa vd. (2005), Kuzey ve Güney Amerika’da bulunan ve ortalama günlük işlem hacmi yüz bin doların üzerinde olan 1959 şirketin hissesini risk getiri kriterine göre incelemiştir. Şirketler iki ayrı dönem şeklinde 10 kümeye ayrılarak, fiyat/kazanç, piyasa değeri/defter değeri, temettü verimi gibi değişkenlerle incelenmiştir. Çalışma sonuçları riski düşük olan hisselerin getirisinin de düşük olduğunu göstermektedir.

Horobet vd. (2008), Çek Cumhuriyeti, Polonya, Romanya ve Macaristan’da bulunan, finansal aracılık, içecek ve gıda sanayi, enerji ve ilaç sektöründe faaliyet gösteren firmaların varlık ve öz kaynak karlılığını 2003-2006 dönemi için Hiyerarşik ve *K* ortalama kümeleme yöntemi ile incelemiştir. Çalışma sonuçları zamanla kümelerin yapısındaki değişime ve incelenen ülkelerin kurumsal finansal performansına işaret etmektedir. Benzer şekilde De vd. (2011), Hindistan’da çimento sektöründe faaliyet gösteren 38 şirketin 10 yıllık finansal performanslarını faktör, çoklu regresyon ve kümeleme analizi ile incelemiştir. Faktör analizi sonuçlarına göre 8 faktör tespit edilmiştir. Çalışmada faktör analizi sonuçlarını kontrol etmek amacı ile uygulanan kümeleme analizi neticesinde faktör analizi ile eşit sayıda kümenin ortaya çıktığı vurgulanmıştır. Kümeleme ve faktör analizi sonuçlarının birbirine benzerlik göstermesi çalışmanın önemli bulguları arasında yer almaktadır. Manniste vd. (2011), hisse senedi fiyatlarının 2007 yılında gerçekleşen kredi daralma vakasına gösterdikleri tepki ile ortaya çıkan büyük dalgalanmaların firmaların coğrafi konumu, büyüklüğü ve finansal performansı ile ilişkili olup olmadığını kümeleme analizi ile incelemiştir. Çalışmada 45 Avrupa ülkesinde faaliyet gösteren 705 şirket dikkate alınmıştır. Sonuçta kazandıran hisselerin çoğunun 15 Avrupa Birliği ülkesinde faaliyet gösteren büyük şirketlerden olma eğilimi gösterdiği vurgulanmakla birlikte bu hisselerin beklenen F/K oranının düşük ve kar marjlarının yüksek olduğu tespit edilmiştir. Ancak hisse fiyatlarında büyük düşüşler yaşamış olan firmaların beklenen F/K oranının ve varlıklar üzerinden beklenen getirinin diğerlerine göre daha yüksek olduğu belirtilmiştir.

Cai vd. (2012), kredi kartı sahtekârlığının tespit edilmesi, hisse senedi piyasalarında yatırım kararlarının verilmesi gibi farklı finansal veri setleri üzerine uygulanan değişik kümeleme algoritmalarının avantaj ve dezavantajlarının neler olduğunu tespit etmeye çalışmışlardır. Bu amaçla 1000 adet kredi kartı örneğinden oluşan bir veri seti kullanarak çalışmalarında farklı kümeleme tekniklerine yer vermişlerdir. Çalışma sonuçları farklı kümeleme tekniklerinin farklı sonuçlar doğurduğuna işaret etmektedir. Örneğin finansal veri

setlerinde yoğunluk tabanlı kümeleme tespit edilememişken normalleştirilmiş merkez tabanlı kümelemenin geçerli olduğu vurgulanmaktadır.

Rani ve Sikka (2012), daha önce mühendislik, finans, ekonomi, sağlık ve kamu gibi birçok alanda yapılan kümeleme çalışmalarının araştırılması ve özetlenmesini amaçlamıştır. Literatür taramasının sonuçlarına göre üç temel sınıflandırmanın varlığına işaret edilmektedir. Başka bir ifade ile zaman serilerinin kümeleme yöntemlerini, (1) doğrudan ham verilere dayanılarak yapılan kümeleme analizi, (2) dolaylı olarak ham verilere dayanan veya ham verileri temsil etme özelliğine sahip verilerden hareketle yapılan kümeleme analizi, (3) her bir zaman serisinin benzer model veya olasılık dağılımına sahip olduğu varsayımına dayanan model temelli kümeleme analizi olarak sıralamışlardır.

Akyüz vd. (2012), imalat sanayinde bulunan sektörleri likidite, mali yapı, faaliyet ve karlılık oranları yardımıyla küme ve ayırma analizini kullanarak 2007-2009 yılları arası süreç için incelemiştir. Sonuçta orman ürünleri sanayinin yer aldığı gruptakilerin diğer birkaç sanayi dalı ile aynı özellikte olduğu ve söz konusu sanayi dallarının hem ekonomik yapı olarak hem de ülke ekonomisine katkıları bakımından benzer nitelikte etkiye sahip oldukları tespit edilmiştir.

Kalfa ve Bekçioğlu (2013), BIST 100 Endeksinde yer alan gıda, tekstil ve çimento sektörlerinde faaliyet gösteren 42 şirketin finansal oranlarını 2006-2011 yılları arasında aldıkları değer ortalamalarını dikkate alarak hiyerarşik olmayan kümeleme yöntemlerinden biri olan K-ortalama yöntemine göre analiz etmişlerdir. Kümeleme analizine göre birinci kümeye 5, ikinci kümeye 19 ve üçüncü kümeye 18 şirketin atandığı gözlenirken diskriminant analizi sonucunda diskriminant fonksiyonunun doğru sınıflandırma oranının %100 olduğu tespit edilmiştir.

Özkan ve Boran (2014), Türk imalat sanayi şirketlerinden 2009 yılı için 152, 2010 yılı için 151 ve 2011 yılı için 148'ini bazı finansal oranlardan hareketle IBM'in Intelligent ve SAS Institute'ün Enterprise Miner (EM) olarak bilinen algoritması ile K-ortalama yöntemine göre kümeleme analizine dâhil etmişlerdir. Sonuç olarak, aynı sektördeki şirketlerin ağırlıklı olarak aynı kümelerde yer aldıkları tespit edilmiştir.

Arı vd. (2016), İstanbul Sanayi Odası'nın 2012 yılında yayınladığı "İlk 500 Sanayi Kuruluşu" listesinde yer alıp BIST'te işlem gören ve rastgele seçilen 90 adet firmanın performanslarını incelemişlerdir. Bu amaçla 2013 yılına ait finansal tablolar üzerinden hesapladıkları 12 farklı finansal orandan hareketle Ward'ın Minimum Varyans ile K-ortalama yöntemlerinin karması olan *İki Aşamalı Kümeleme Analizini* kullanmış ve yapılan üç farklı kümeleme analizi ile farklı sonuçlara ulaşmışlardır. Genel olarak değerlendirildiğinde karlılık oranlarında ilk kümeye dahil firmaların daha yüksek performanslara sahip olduğunu, ikinci kümedeki firmaların ise düşük karlarla ya da zarar ederek yılı kapattıklarını tespit etmişlerdir. Ayrıca oranları daha düşük olan firmaların genellikle yılı zararla kapatmaları nedeniyle farklı bir küme oluşturdukları görülmüştür.

Literatürdeki finansal veriler üzerine yapılan kümeleme çalışmalarına bakıldığında farklı sonuçlara ulaşıldığı görülmektedir. Ayrıca birçok çalışmada şirketler kümeleme analizi kapsamında finansal oranlar yolu ile araştırılmış ve beklenen kümelerin olduğu sonucuna ulaşılmıştır. Bu çalışmada geçmiş fiyat verileri baz alınarak bir kümeleme analizinin

yapılması amaçlanmaktadır. Başka bir ifade ile bu çalışmada BIST’te ana faaliyet konusuna göre sektöre dahil olan hisselerin, geçmiş fiyat verilerindeki yüzdelik değişime göre kümeleme analizine tabi tutulduğunda aynı sektörel dağılımın gerçekleşip gerçekleşmeyeceği tespit edilmek istenmiştir. Çalışma bu yönü ile Türkiye piyasalarında yapılan diğer çalışmalardan farklılık göstermektedir. Zira literatürde hisselerin geçmiş fiyat verilerindeki yüzdelik değişimi baz alınarak yapılan başka bir kümeleme çalışmasına rastlanamamıştır.

### 3. VERİ SETİ VE YÖNTEM

Bu çalışmada, BIST’te faaliyet gösteren ve çeşitli sektörlerde yer alan 70 farklı hisse senedine ilişkin 22 Ekim 2012 ve 30 Kasım 2015 tarihleri arasındaki *haftalık veriler* kullanılmıştır. Bu tarihler literatürde finansal verilerin kümeleme analizinin genellikle üç yıllık zaman dilimi için gerçekleştirilmesi sebebiyle güncel durumu ortaya koymak amacıyla analizin yapıldığı tarihten geriye üç yıl gidilerek belirlenmiştir. Analizde hisse senetleri üzerine daha önce yapılan kümeleme çalışmalarında olduğu gibi veri boyutunu azaltmak adına haftalık veriler kullanılmıştır. Veriler bir veri dağıtım şirketi olan Matriks Veri Terminalinden elde edilmiştir. Analize dâhil edilen 70 hisse senedi 9 farklı sektörden seçilmiştir. Bu sektörler: Teknoloji; İmalat Sanayi; Toptan ve Perakende Ticaret, Otel ve Lokantalar; Madencilik; Mali Kuruluşlar; İnşaat ve Bayındırlık; Ulaştırma, Haberleşme ve Depolama; Elektrik, Gaz ve Su; Eğitim, Sağlık, Spor ve Diğer Sosyal Hizmetler’dir.

Hisse senetlerinin BIST’teki sektörel dağılımı, kümeleme analizi sonuçlarını değerlendirebilmek üzere referans olarak kullanılmıştır. Sektörler analize dâhil edilirken BIST100 Endeksi içinde *ağırlığı yüksek olan sektörlere öncelik* verilmiştir. Böylece Mali, Teknoloji, İmalat, Ulaştırma, Toptan ve Perakende sektörleri analize BIST100’deki payları dikkate alınarak dâhil edilmiştir. Ayrıca bu sektörlerle ait hisseler de BIST100’deki piyasa değerlerinin büyüklüğüne göre seçilmiştir. Ancak bazı sektörlerde piyasa değeri açısından daha yüksek değere sahip olan hisselerle birlikte alt sektörlerdeki çeşitliliğe dikkat edilmiştir. Örneğin mali kuruluşlarda banka hisseleri ile birlikte holding hisselerine de yer verilmiştir. Diğer sektörlerin seçiminde ise sektöre dâhil olan hisse senedi sayısı dikkate alınmıştır. Zira üç yıllık analiz döneminden önce ya da sonra işlem gören hisseler çıkarıldıktan sonra en az beş hisse senedini içeren sektörlerle yer verilmiştir. Sektörlere ve hisselerine ait seçim Wittman’ın (2002) çalışmasından yola çıkarak Türkiye piyasalarına göre uyarlanarak gerçekleştirilmiştir. Bu çalışmada sektör ve hisse seçimi Wittman’ın (2002) çalışması referans alınarak belirlenmiştir.

Hisse senetleri piyasasının çoğu kez gerçekler üzerine hareket etmesi beklenmesine rağmen algılamalar üzerine hareket ettiğine ilişkin davranışsal finans çalışmaları dikkate alındığında, kümeleme çalışmalarında algılamaların dolaylı etkilerine bakıldığı görülmektedir. Bu kapsamda, bazı hisseler analize özellikle dâhil edilmiştir. Örneğin BIST’te Toptan ve Perakende Sektöründe faaliyet gösteren Türkiye’nin ilk teknoloji perakende şirketi Bimeks ve teknoloji perakendeciliğinde önemli bir paya sahip Teknosa özellikle bireysel yatırımcılar tarafından Teknoloji sektöründe faaliyet gösteren hisse senetleri olarak algılanabilmektedir. Benzer bir şekilde yine Toptan ve Perakende sektöründe faaliyet gösteren Mepet, imalat sektöründe faaliyet gösteren bir hisse senedi ya da Madencilik (Ham Petrol ve Doğal Gaz Üretimi) sektöründeki bir hisse senedi gibi algılanabilmektedir. Bu hisse senetleri yatırımcıların “*sektörel dağılım*” algısını tespit edebilmek üzere analizde yer almaktadır.

Çalışmadaki kümeleme analizi sonuçları BIST'teki sektörel gruplandırma ile karşılaştırılarak bu iki grup arasında fark olup olmadığı hakkında bir kanıya varılabilir.

Hisse senetlerinin kümelenmesinde temel alınan varsayım, Wittman (2002) tarafından şu şekilde ifade edilmektedir. Eğer uygulanan uzaklık ölçütüne göre iyi bir kümeleme elde edilmişse bu durumda hisse senetleri bir grup olarak hareket etme eğilimindedir. Bu çıkarım finansal piyasalar için önem arz etmektedir. Dahası kümeleme istatistikleri, hisse senetlerinin hangi düzeyde birlikte hareket ettiğini gösterebilmektedir. Bu çalışmada kümeleme analizi, yalnızca geçmiş fiyat hareketlerine dayandırıldığından BIST'te hangi sektörde yer aldığına bakılmaksızın, kümeleme analizi hangi gruptaki hisse senetlerinin geçmiş fiyat verilerine göre bir birine benzer hareket ettiğini gösterecektir.

Zaman serilerinde ölçüm yapılmak istenilen süre ne kadar kısaltılırsa o seriye ilişkin volatilité de artmaktadır. Bu sebeple zaman serileri kümelemesi üzerine daha önce yapılan çalışmalar da dikkate alınarak *haftalık veri seti* kullanılmıştır. Bir hisse senedinin fiyatı, geliri ya da bir şirketin büyüklüğünü yeterince yansıtmadığı için hisse senetlerinin fiyatları üzerinde doğrudan analiz yapmak yanıltıcı olabilir. Bu sebeple yüzdelik değişim hisse senetlerinin belirli bir zamandaki performansını iyi bir kıyaslama yaparak ölçebilir. Her bir hisse senedi için yüzdelik değişim aşağıdaki gibi hesaplanmıştır (Wittman, 2002, 3).

$$P_i(t) = \frac{s_i(t) - s_i(t-1)}{s_i(t-1)} \times 100 \quad (1)$$

Eşitlik 1'de  $P_i$  i hisse senedinin t zamandaki yüzdelik fiyat değişimini göstermekte iken  $s_i(t)$  i hisse senedinin t zamandaki fiyatını göstermektedir. Zaman serilerinin kümeleme analizinde değişkenler arasındaki uzaklık farkını bulmak için bir uzaklık ölçütüne ihtiyaç duyulur. Öklid Uzaklığı zaman serilerinde sıklıkla kullanılan, en temel uzaklık ölçütlerinden biridir. Öklid uzaklığı iki nokta arasındaki doğrusal uzaklığı ifade etmektedir. İki hisse senedi (i,j) arasındaki Öklid uzaklığı aşağıdaki gibi hesaplanmaktadır.

$$d(i, j) = \sqrt{\sum_t (P_i(t) - P_j(t))^2} = P_i - P_j \quad (2)$$

Hisse senetlerinde kümeleme analizlerinde karşılaşılan en büyük zorluklardan biri hisse senedi piyasasının bütün olarak hareket etme eğilimidir. Başka bir ifade ile hisse senetlerinin grup halinde hareket ettiği varsayılmaktadır. Gruplar arasındaki farkı görmek için grup olarak gerçekleşen hisse senedi hareketlerini devre dışı bırakıp daha ince trendlere odaklanmak gerekir. Analizde bu ayrımı dikkate almanın bir yolu verileri standartlaştırmaktır. Hisse senedi verilerini standartlaştırmada iki temel yöntem kullanılmaktadır. Bu yöntemler veriye (hisse senedine) dayalı ve zamana dayalı standartlaştırmadır. Hiyerarşik yığınsal kümelemeye ilişkin yapılan çalışmalarda zamana dayalı standartlaştırmının daha iyi sonuçlar verdiği tespit edilmiştir (Wittman, 2002, 4). Zaman bazlı standartlaştırma için Eşitlik 3 kullanılmıştır.

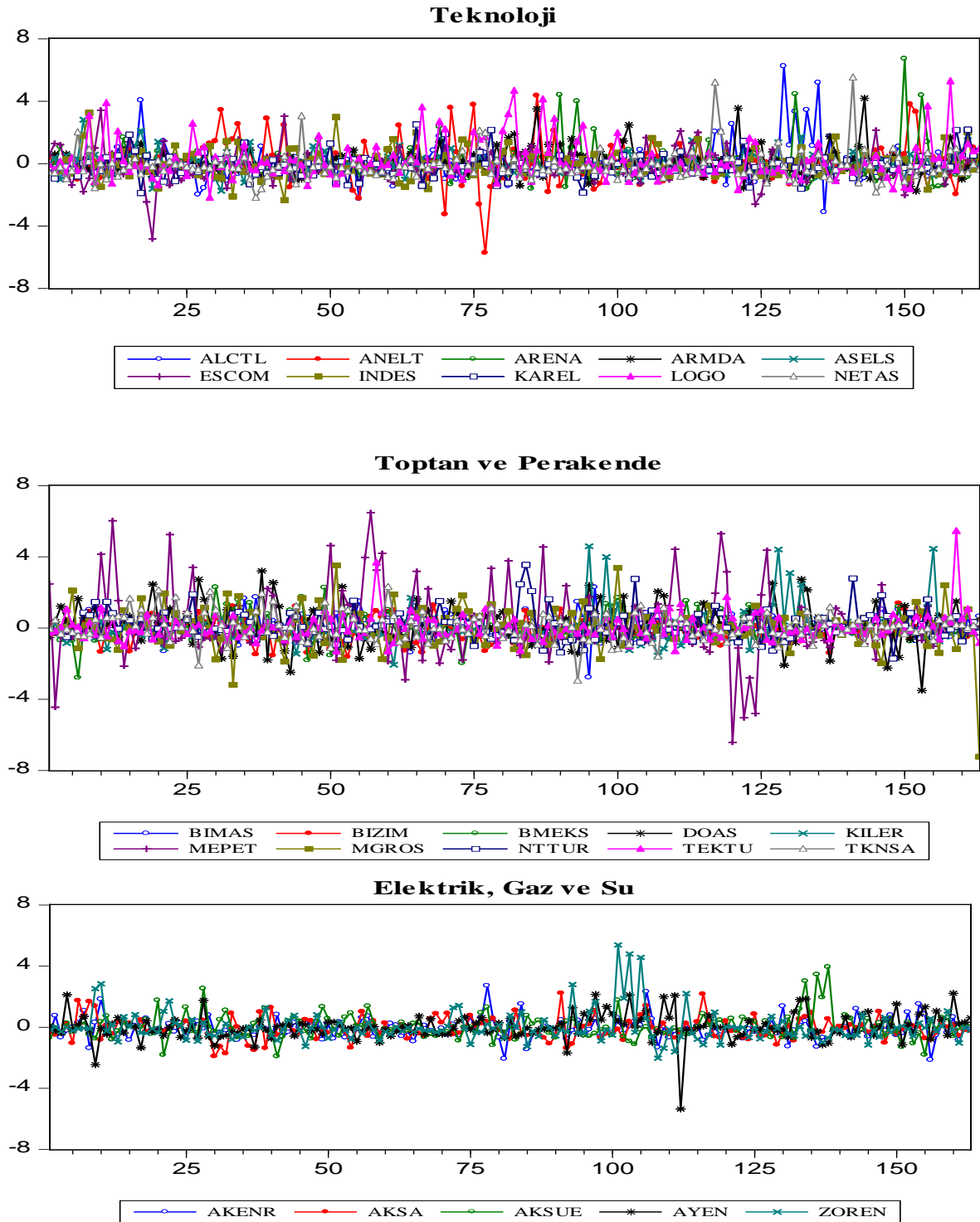
$$\tilde{P}_i(t) = \frac{P_i(t) - \mu(t)}{\sigma(t)} \quad (3)$$

Burada  $\mu(t)$  ortalama yüzde deęişimi,  $\sigma(t)$  standart sapmayı göstermektedir. Böylece zamana dayalı standartlaştırma ile hisse senedi piyasasının genel trend eğilimi ortadan kaldırılmaktadır. Kümeleme analizinde, kümeleme işlemi iki şekilde gerçekleştirilebilir. Bunlar; hiyerarşik ve hiyerarşik olmayan kümeleme. En çok kullanılan yöntemler, hiyerarşik yöntemler grubudur. Bu grup kendi içinde yığınsal ve bölücü hiyerarşik olmak üzere ikiye ayrılır. Hiyerarşik kümelemenin en etkin ve yaygın kullanılan yöntemi hiyerarşik yığınsal yöntemdir. Bu yöntem ilk olarak bütün gözlemleri tek bir kümede toplar; sonra bu kümeye en aykırı olan gözlemleri birer birer kümeden ayırarak başka kümelerin oluşmasını sağlar. Yöntem kaç tane küme oluşması gerektiğine kendisi karar verebilir (Kalaycı, 2008, 358). Finansal verilerin kümeleme analizine ilişkin literatür dikkate alındığında daha çok Hiyerarşik Yığınsal Kümeleme analizinin kullanıldığı görülmektedir (Gavrilov vd, 2000, 492; Wittman, 2002,4).

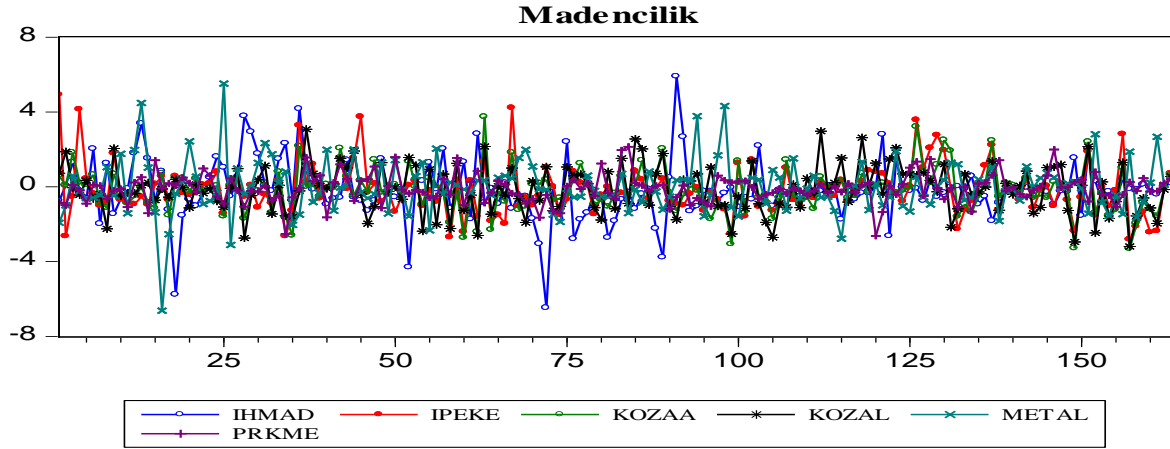
#### 4. ANALİZ SONUÇLARI

Bu çalışmada en temel kümeleme tekniklerinden biri olan Hiyerarşik Yığınsal Kümeleme yöntemi kullanılmış ve sonuçlar küme ağacı ile gösterilmiştir. Küme ağacı, soy ağacına benzer bir şekilde yukarıdan aşağıya doğru görselleştirilmiş bir dendrogramdır. Kümeleme analizlerinde küme ağaçları algoritması oluşturulurken iki aşama izlenir. Öncelikle küme ağacı oluşturulur. Sonra kümedeki aykırı veriler tespit edilerek küme ağacı budanır. Küme ağacında  $k$  tane küme elde etmek için ağacı en tepeden istenilen  $k$  seviyesine kadar ‘budamak’ gerekmektedir. Bütün verileri kümeleme analizine dâhil etmeden önce yukarıda bahsedilen “*sektörel dağılım algısı*”na daha yakından bakabilmek ve finansal verilerin kümeleme analizini daha anlaşılır kılabilmek üzere öncelikle birkaç sektörün kümeleme analizine yer vermek daha uygun olacaktır. Bu kapsamda ilk önce dört sektörün (Teknoloji; Toptan ve Perakende Ticaret, Otel ve Lokantalar; Madencilik; Elektrik, Gaz ve Su) kümeleme analizi gerçekleştirilmiştir. İlk analiz için kümelemede yaygın olarak kullanılan dört bağlantı tekniğine yer verilmiştir: (1) Tek Bağlantı (Single Link), (2) Tam Bağlantı (Complete Link), (3) Ortalama Bağlantı (Average Link) ve (4) Ward Yöntemi (Ward’s Method). Tek Bağlantı Yöntemi en kısa mesafe esasına dayanmaktadır. Başka bir ifadeyle, bu teknik birbirine en yakın iki gözlemi bularak kümeleme işlemine başlar ve birbirine yakın başka iki gözlem daha bularak kümeyi genişletir. Tam Bağlantı Yöntemi kümeleme işlemine en uzak iki gözlemden başlarken, Ortalama Bağlantı Yöntemi kümenin ortasına düşen gözlemi dikkate alır. Ward’s Yöntemi ise bir kümenin ortasına düşen gözlemin aynı kümenin içinde bulunan gözlemlerden ortalama uzaklığını dikkate alarak kümeleme işlemine başlamaktadır. Grafik 1’de ilk analiz için seçilen sektörlerde yer alan hisselerin haftalık yüzdesel deęişimine ilişkin grafikler yer almaktadır.

Grafik 1. Sektörlere Göre Hisse Senetlerinin Haftalık Fiyat Değişimleri (%)



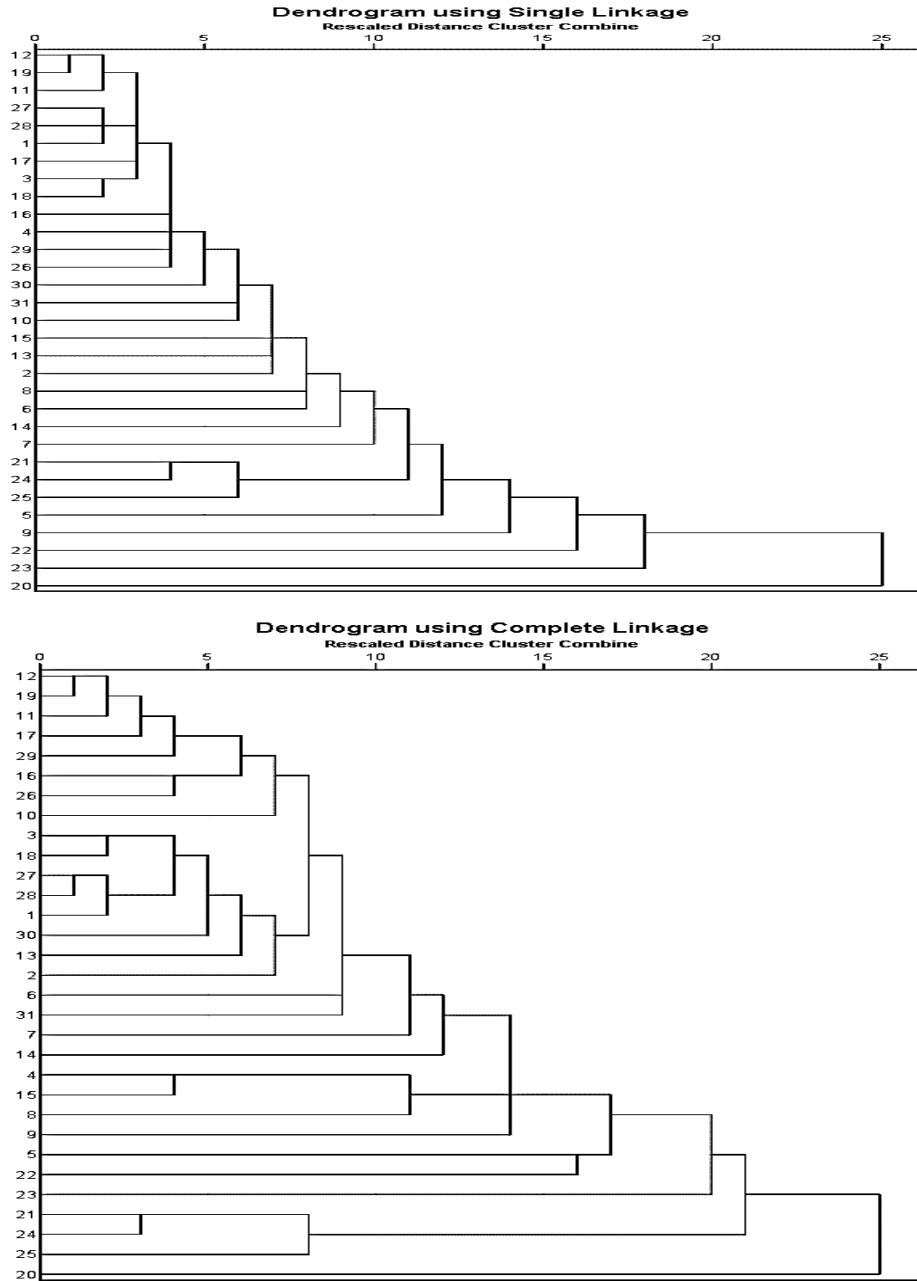


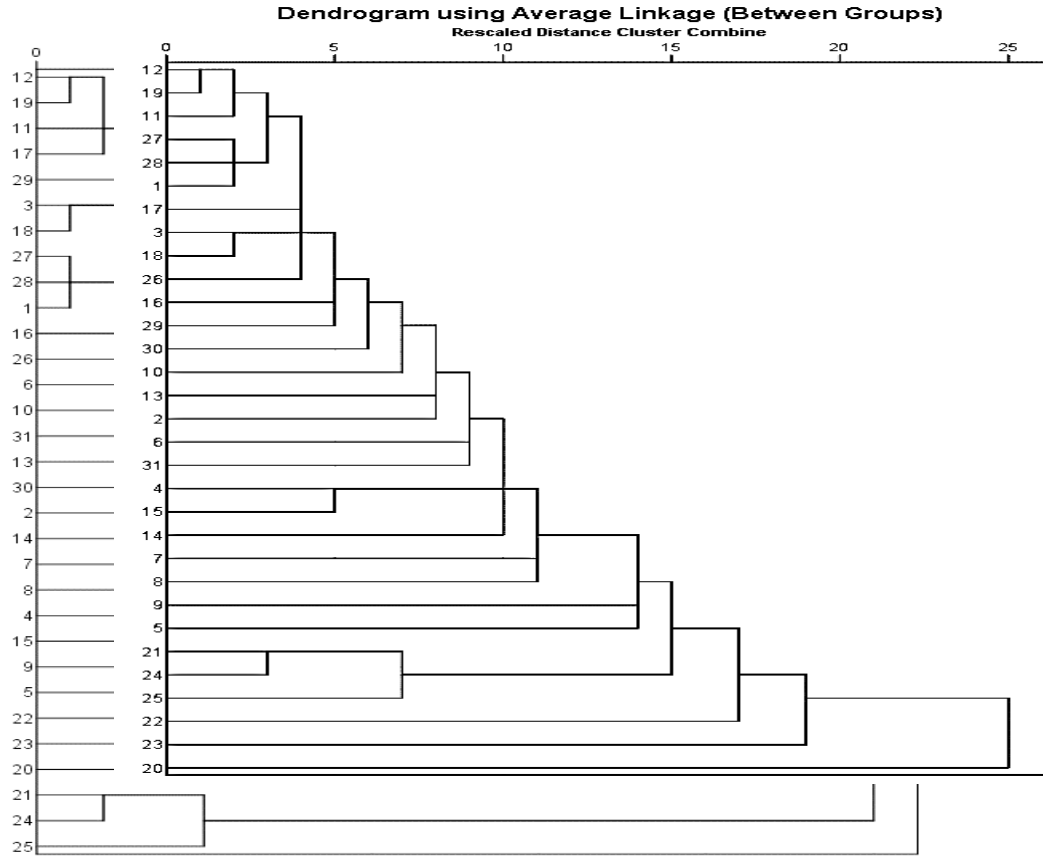


Grafik 1 incelendiğinde özellikle teknoloji sektöründe bulunan birçok hisse senedinin 0 noktasından çoğu kez yukarı yönlü olmak üzere uzaklaştığı görülmektedir. Toptan ve Perakende olarak isimlendirilen sektör hisselerinden Mepet'in diğer hisse senetlerinden çoğu kez farklı hareket ettiği görülmektedir. Bu durum Mepet'in sektördeki diğer hisse senetlerinden farklı algılanıp algılanmadığı durumunun ön göstergesi olarak değerlendirilebilir. Ayrıca, Grafik 1'de Teknosa ve Bimeks hisse senetlerinin sektördeki diğer hisse senetleriyle çoğu zaman birlikte hareket ettikleri dikkat çekmektedir. Yukarıdaki grafikler hisse senetleri hakkında bir fikir edinilmesini sağlasa da elde edilen çıkarımların kümeleme analizi ile desteklenmesi gerekmektedir.

Şekil 1'de dört farklı bağlantı metodu ile dört sektörün hiyerarşik yığınsal kümeleme sonuçları yer almaktadır. Şekil 1'deki dendrogramlar incelendiğinde Toptan ve Perakende sektöründe yer alan Teknosa ve Bimeks hisselerinin dört bağlantı yönteminde de birbirlerine yakın kümelendikleri, bu kümede toptan ve perakende sektöründen başka şirketlerin de yer alması ve küme dışı olmamaları bu şirketlerin hisselerinin birebir teknoloji şirketi olarak değerlendirilmediğini göstermektedir. Ancak Mepet'in dört diyagramda da küme dışı olduğu görülmektedir. Ayrıca Mepet'in İmalat ya da Elektrik, Gaz ve Su sektöründeki şirketlerle de kümelenmemesi Wittman'ın (2002) çalışmasına benzer sonuçları ifade etmektedir. Mepet Toptan ve Perakende sektörünün ve diğer sektörlerin dışında kümelenmiştir. Mepet'e ilişkin analiz dönemi araştırıldığında hisselerin çok fazla spekülasyon konusu olduğu dikkat çekmektedir. Yaygın olarak kullanılan Ward's Metodu incelendiğinde madencilik sektörüne ait hisselerin birlikte kümelendiği dikkat çekmektedir. Benzer sonuçların ortalama ve tam bağlantı tekniklerinde de geçerli olduğu görülmektedir. Madencilik sektörü hisselerinin BIST100'deki sektörel dağılıma benzer şekilde kümelendiğini söylemek mümkündür.

Şekil 1. Dört Sektöre Ait Farklı Bağlantı Yöntemleri Kullanılarak Elde Edilmiş Dendrogram





\*Grafikteki sayıların hangi hisse senetlerini ifade ettiği Ek 1’de gösterilmektedir.

lar

Grafik 1’e göre sektör hisselerinden bazıları birbirlerine yakın kümelenmiş olsa da BIST100’deki sektörel dağılımla örtüşmeyen birçok aykırı gözlemin olduğu söylenebilir. Tek Bağlantı diyagramı incelendiğinde kümelerin çoğunun tek hisse senedi içerdiği görülmektedir. Bu durumun sebebi *Tek Bağlantı* tekniğinin aykırı gözlemlere özellikle duyarlı olmasıdır. Kısacası, dört sektör üzerine yapılan analizde geçmiş fiyat hareketlerine göre yapılan kümeleme ile referans kabul edilen BIST100’deki sektörel dağılımın bazı kümeler dışında örtüşmediği ve bu benzeşmeyi sağlamayan birçok aykırı gözlemin olduğu görülmektedir.

Analiz dokuz sektör ve dolayısıyla 70 hisse senedi için tekrarlanıp aykırı gözlemler analizden çıkarıldığında kümeleme sonuçları daha detaylı bilgiye ulaşılmasını sağlayacaktır. Aykırı gözlemlerin analizden çıkarılmasına ilişkin birkaç yöntem söz konusudur. Bu çalışmada kullanılan finansal verilerin Wittman’ın (2002) çalışmasına benzer şekilde normal dağılıma sahip olmaması, aykırı gözlemlerin analizden çıkarılmasında aynı yöntemin izlenmesine yardımcı olmuştur.

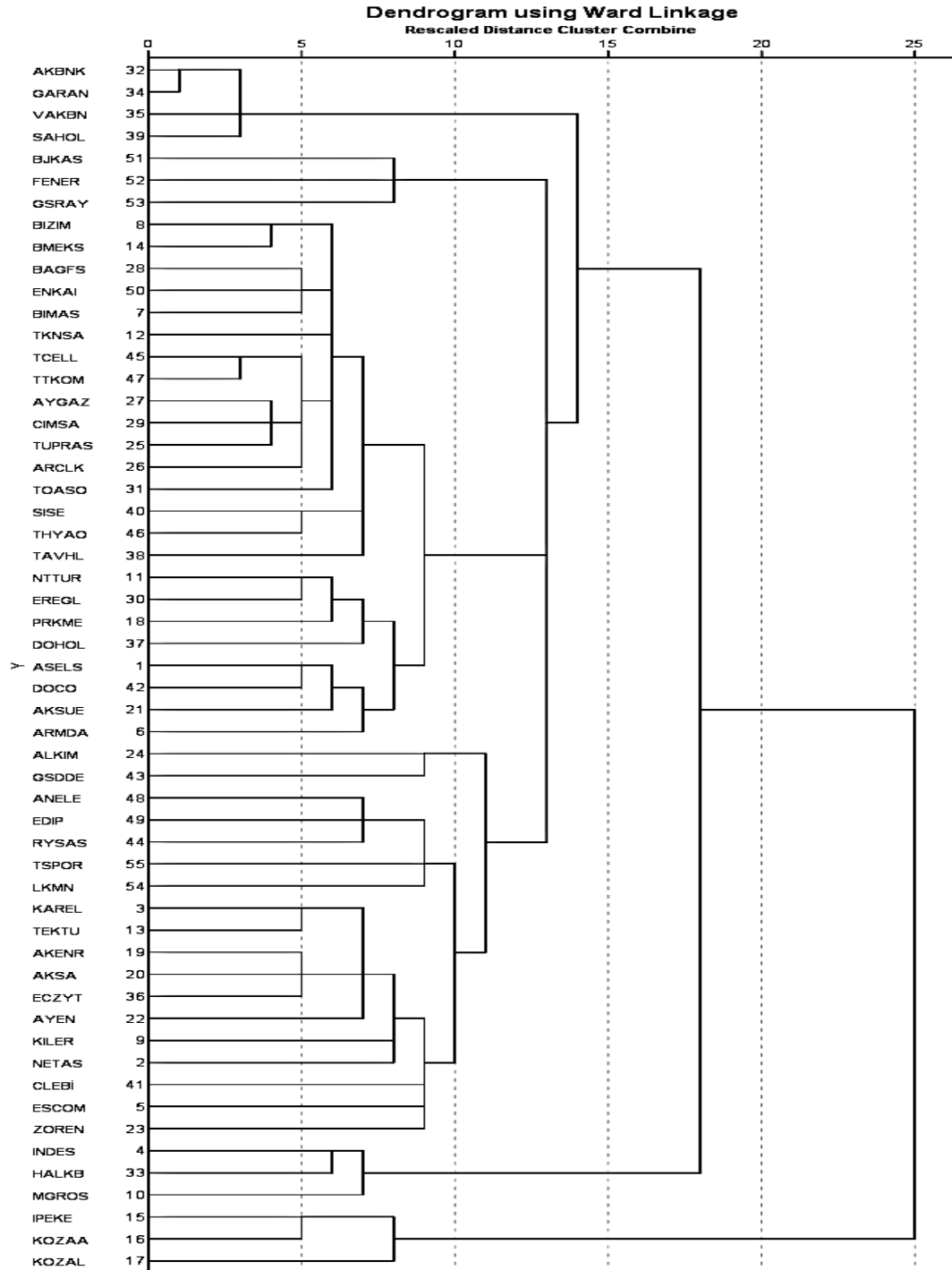
Aykırı gözlemler analizden çıkarılırken küme ağacı istenilen küme sayısına göre ‘budanmış’ ve her kümede en az iki hisse kalıncaya kadar analiz tekrarlanmıştır. Dört ayrı kümeleme metodu ile tekrarlanan analiz literatürdeki çalışmalara benzer sonuçlar

içermektedir. Başka bir deyişle, bu analiz için de Ward's Metodu kümelemeyi diğerlerine göre en iyi gerçekleştiren yöntem olmuştur. Bundan dolayı, Şekil 2'de sadece Ward's yöntemine ilişkin dendograma yer verilmiştir.

Ward's bağlantı tekniğiyle toplamda 15 aykırı gözlem tespit edilmiş ve aykırı gözlemler 7 kez tekrarlanan analizle çalışma dışı bırakılmıştır. Böylece her kümede en az iki hisse senedinin bulunduğu Şekil 2'teki dendrogram elde edilmiştir. Sonuçlar incelendiğinde İmalat, Haberleşme ve Ulaştırma, Toptan ve Perakende sektöründe yer alan hisselerin birbirine yakın kümelendiği görülmektedir. Bu sektörlerle ait hisselerin aynı kümede ve birbirine yakın gruplanmış olması fiyat hareketlerinin benzer olduğunu göstermektedir. Mali kuruluşlar; Elektrik, Gaz ve Su; Madencilik; Eğitim ve Spor sektör hisselerinin birçoğunun da BIST100'deki sektörel kategoriye göre kümelenecek olması bu analize ilişkin kümeleme sonuçlarının BIST100'deki dağılıma daha yakın sonuçlar içerdiğini göstermektedir. Ancak geri kalan kümelerdeki hisselerin sektörel bir bağ olmadan kümelendiği dikkat çekmektedir. Analiz dışında kalan aykırı gözlemler incelendiğinde teknoloji sektörü hisselerinin ağırlıkta olduğu görülmektedir. Ayrıca analiz dışında kalmayan teknoloji hisseleri için de belli bir küme söz konusu değildir. Yukarıda yer alan Teknoloji sektörüne ait grafik incelendiğinde de birçok hissenin 0 noktasından uzaklaştığı başka bir ifade ile teknoloji sektörünün diğer sektörlerle göre daha değişken olduğu görülmektedir. BIST100'deki teknoloji sektörü hisselerinin geçmiş fiyat hareketleri dikkate alındığında aynı kümede toplanmadığını söylemek mümkündür.

BIST'te Eğitim, Sağlık, Spor ve Diğer Sosyal Hizmetler sektörüne ait beş hisse senedi işlem görmektedir. Analiz kapsamında olan Beşiktaş, Fenerbahçe ve Galatasaray hisselerinin aynı kümede olması Eğitim ve Spor sektörü hisse senetlerinin geçmiş fiyat hareketlerine göre incelendiğinde de aynı grupta yer aldığını göstermektedir. Bu sektöre ait hisseler incelendiğinde analiz döneminde üç büyük takıma ilişkin hisseler için çok düşük fiyat hareketlerine rastlamak mümkündür. İncelenen dönemde sektörde yaşanan gelişmelere bakıldığında 2013 yılında UEFA'nın Fenerbahçe, Beşiktaş ve Trabzonspor ile ilgili yaptırımları, Galatasaray'ın SPK'ya bedelli sermaye artırım başvuruları ve BIST Yönetiminin dört büyük takıma ait şirketi finansal sorunları sebebi ile uyarması dikkat çekmektedir. 2013 yılında Trabzon Spor hisseleri %56,4 oranında düşüş kaydederken, Galatasaray %46,8, Fenerbahçe %29,8 ve Beşiktaş %9,1 değer kaybetmiştir. Bu sektöre ait grafik incelendiğinde Beşiktaş, Fenerbahçe ve Galatasaray hisselerinin benzer hareket ettiği ancak Trabzonspor'un daha değişken olduğu dikkat çekmektedir. Bu durumda analizde Beşiktaş, Fenerbahçe ve Galatasaray'ın birlikte kümelenecek hisseler olması anlam kazanabilir.

Şekil 2. Ward's Bağlantı Yöntemine Göre Oluşturulmuş Dendrogram



## 5. SONUÇ

Kümeleme analizi verilerin benzerliklerini dikkate alarak onları iki ya da daha fazla gruba bölmeyi amaçlamaktadır. Kümeleme analizi her ne kadar yaygın bir teknik olsa da finansal verilerin kümeleme analizi literatürde sık rastlanan bir konu değildir. Bu çalışmada

BIST100’de işlem gören hisseler sektörel bazda haftalık verilerle üç yıllık bir süreç için Hiyerarşik Yığınsal Kümeleme analizi ile incelenmiştir. Öncelikle analize özellikle dâhil edilen hisselerin de bulunduğu dört sektör, dört farklı bağlantı tekniği ile incelenmiştir. Dendrogramlar incelendiğinde bazı hisse senetlerinin BIST100’deki sektörel dağılıma benzer bir şekilde kümelendiği görülmektedir. Ancak dört bağlantı tekniğinde de çoğu hisse senedinin tek başına bir küme oluşturduğu ve aykırı gözlem olarak öne çıktığı görülmektedir. Bu durumda çoğu hisse senedinin BIST100’deki sektörel kategoriye göre kümelendiği sonucu ortaya çıkmaktadır. Başka bir ifade ile ilk analiz için seçilen sektörlerdeki hisselerin tamamı analize dahil edildiğinde BIST’te ana faaliyet konusuna göre oluşturulan sektörel dağılıma uzak bir sonuç elde edilmiştir. Bu durumda seçilen sektördeki hisse senetlerinin geçmiş fiyat hareketlerine göre kümeleme analizi ana faaliyet konusuna göre yapılan kümelemeden farklılık arz etmektedir. BIST’teki sektörel dağılıma yakın bir kümeleme sonucu elde edebilmek için sonraki aşamada aykırı gözlemler analizden çıkarılmış ve her seferinde analiz yeniden tekrarlanmıştır. İkinci analizde Ward’s bağlantı tekniği ile toplamda 15 aykırı gözlem tespit edilmiş ve aykırı gözlemler 7 kez tekrarlanan analizle çalışma dışı bırakılmıştır. Ward’s Metodu diğer bağlantı tekniklerine göre kümelemeyi en iyi gerçekleştiren yöntem olmuş ve her kümede en az iki hisse senedinin kaldığı sonuçlar elde edilmiştir.

Analiz sonuçları aykırı gözlemler çıkarıldıktan sonra İmalat, Haberleşme ve Ulaştırma, Toptan ve Perakende, Mali Kuruluşlar; Elektrik, Gaz ve Su; Madencilik; Eğitim ve Spor sektörlerine ait hisselerin BIST100’deki dağılıma benzer bir şekilde kümelendiğini göstermektedir. Bu durumda diğerleri ile farklı hareket eden hisse senetleri çıkarıldıktan sonra incelenen sektörlerin çoğunda BIST’teki sektörel dağılıma yakın bir kümelemenin elde edildiğini söylemek mümkündür. Analiz sonuçları yatırımcıların sektörel bazda yatırımlarını gerçekleştirirken bazı hisse senetlerinin sektördeki diğer hisselerden fiyat adımları açısından farklılık göstermesi nedeniyle portföyde o hisse senetleri ile çeşitlendirme yapabileceklerini göstermektedir. Ancak özellikle teknoloji ve inşaat sektörlerinde geçmiş fiyat hareketlerine göre bir kümelenemenin gerçekleşmediği dikkat çekmektedir. Bu bulgu, teknoloji ve inşaat sektör hisselerinin benzer fiyat adımlarına sahip olmadığını ifade etmektedir. Bu sektörlerin incelenen dönemde diğer sektörlerle göre daha volatil olduğu grafiklerde de görülmektedir. Finansal verilerin kümeleme analizine ilişkin daha sonra yapılacak olan çalışmalarda incelenen dönem ve hisse senetleri farklılaştırılabilir. Ayrıca kümeleme analizi birkaç menkul kıymetten oluşan portföylerin zaman serisi analizinde de kullanılabilir. Böylece portföylerin kümeleme analizinde, hangi sektörlerin portföy üzerinde etkisinin daha fazla olduğu tespit edilebilir.

## **KAYNAKLAR**

Akyüz, K. Cemil - Balaban, Yasin - Yıldırım, İbrahim (2012), “Bilanço Oranları Yardımıyla Orman Ürünleri Sanayisinin Finansal Yapısının Değerlendirilmesi”, Uluslararası İktisadi ve İdari İncelemeler Dergisi , Sayı: 9, ss. 133-144.

Arı, E. Sertaç - Özköse, Hakan - Doğan, Ahmet - Calp, M. Hanefi (2016), “İstanbul Borsası’nda İşlem Gören Firmaların Finansal Performanslarının Kümeleme Analizi İle Değerlendirilmesi”, Bilişim Teknolojileri Dergisi, Sayı: 9(1), ss. 33-39.

- Cai, Fan - Le-Khac, Nhien-An - Kechadi, M- Tahar (2012), “Clustering Approaches for Financial Data Analysis: A Survey”, In Proceedings of the International Conference on Data Mining (DMIN). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- Da Costa, Newton - Cunha, Jefferson - Da Silva, Sergio (2005), “Stock Selection Based on Cluster Analysis”, Economics Bulletin, Vol. 13 (1), pp.1-9.
- De, Anupam - Bandyopadhyay, Gautam - Chakraborty, B.N. (2011), “Application of the Factor Analysis on the Financial Ratios and Validation of the Results by the Cluster Analysis: An Empirical Study on the Indian Cement Industry”, Journal of Business Studies Quarterly, Vol. 2 (3), pp. 13-31.
- Gavrilov, Martin - Anguelov, Dragomir - Indyk, Piotr - Motwani, Rajeev (2000), “Mining the Stock Market: Which Measure is Best?” In Proceedings of the Sixth International Conference on Knowledge Discovery and Data Mining, pp. 487-496.
- Horobet, Alexandra. - Joldes, Cosmin - Dumitrescu, Dan Gavriel (2008), “A Cluster Analysis of Financial Performance in Central and Eastern Europe”, Economic Science Series, Vol. VXII (3), pp.289-294.
- Kalaycı, Şeref (2008), SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri, 3. Baskı, Asil Yayın Dağıtım
- Kalfa, V. Rıza, Selim - Bekçioğlu (2013), “İMKB’de İşlem Gören Gıda, Tekstil ve Çimento Sektörü Şirketlerinin Finansal Oranlar Yardımıyla Kümelenmesi”, Dumlupınar Üniversitesi Sosyal Bilimler Dergisi EYİ 2013 Özel Sayısı, ss. 441-463.
- Manniste, Mari - Hazak, Aaro - Listra, Enn (2011), “Typology of European Listed Companies’s Reactions to Global Credit Crunch: Cluster Analysis of Share Price Performance”, International Journal of Trade, Economics and Finance, Vol. 2(6), pp.478-483.
- Özkan, Mehmet - Boran, Levent (2014), “Veri Madenciliğinin Finansal Kararlarda Kullanımı”, Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Sayı: 4(1), ss.59-82.
- Rani, Sangeeta - Sikka, Geeta (2012), “Recent Techniques of Clustering of Time Series Data: A Survey”, International Journal of Computer Applications, Vol. 52(15), ss.1-9.
- Wittman, Todd (2002), “Time Series Clustering and Association Analysis of Financial Data”,  
<https://pdfs.semanticscholar.org/ed5c/c487cd94b9f3c76d868d326e3554f78b0fd7.pdf>

## Ek 1. Hiyerarşik Yıgınsal Kümleme Analizi ile İncelenen Hisse Senetleri

<b>Teknoloji</b>	<b>Elektrik, Gaz ve Su</b>	<b>Mali Kuruluşlar</b>
1 Aselsan	27 Ak Enerji	50 Akbank
2 Netaş Telekom	28 Aksa Enerji	51 Halkbank
3 Karel Elektronik	29 Aksu Enerji	52 Garanti Bankası
4 İndes Bilgisayar	30 Ayen Enerji	53 İş Bankası
5 Anel Telekom	31 Zorlu Enerji	54 Vakıf Bank
6 Escort Teknoloji	<b>Ulaştırma, Haberleşme</b>	55 Eczacıbaşı Yat.
7 Alcatel Lucent	32 Beyaz Filo	56 Doğan Holding
8 Arena Bilgisayar	33 Çelebi Hava Servisi	57 Tav Hava Lim.
9 Logo Yazılım	34 Do-Co	58 Sabancı Holding
10 Armada Bil.	35 Gsd Denizcilik	59 Şişe Cam
<b>Toptan Perakende Tic.</b>	36 Reysaş Lojistik	<b>İnşaat ve Bayındırlık</b>
11 Bim Mağazaları	37 Turkcell	60 Anel Elektrik
12 Bizim Mağazaları	38 Türk Hava Yolları	61 Edip Gayrimenkul
13 Kiler Alış. Hiz.	39 Türk Telekom	62 Enka İnşaat
14 Doğu Otomotiv	<b>İmalat Sanayi</b>	63 Kuyumcukent
15 Migros Ticaret	40 Afyon Çimento	64 Orge Enerji
16 Net Turizm	41 Akim Kimya	65 Yeşil Yapı
17 Teknosa	42 Tüpraş	<b>Eğitim, Sağlık, Spor</b>
18 Tek-Art İnşaat	43 Arçelik	66 Beşiktaş Futbol Yat.
19 Bimeks	44 Aygaz	67 Fenerbahçe Futbol
20 Mepet Petrol	45 Bagfaş	68 Galatasaray Sportif
<b>Madencilik</b>	46 Çimsa	69 Trabzonspor Sportif
21 İpek Doğal Enerji	47 Ereğli	70 Lokman Hekim Sağ.
22 Metal Gayrimenkul	48 Vestel	
23 İhlas Madencilik	49 Tofaş Oto. Fab.	
24 Koza Madencilik		
25 Koza Altın		
26 Park Elek. Madencilik		



Ek 2. Analizde Kullanılan Hisse Senetlerinin Sektör İçi Grafikleri

