



## Hisse Senetlerinin Özdüzenleyici Haritalarla Kümelendirilmesi: BIST50 Endeksinde Yer Alan Hisseler Üzerine Bir Uygulama

Mehmet Özçalıcı<sup>†</sup>

(İktisadi ve İdari Bilimler Fakültesi, Kilis 7 Aralık Üniversitesi, Kilis, Türkiye)

### ARTICLE INFO

*Yayın Bilgisi*

Received/Başvuru  
18/02/2016

Accepted/Kabul  
19/04/2016

### Anahtar Sözcükler:

Özdüzenleyici Haritalar  
Hisse Senetleri  
Portföy Yönetimi  
Benzerlik Matrisi  
Silhouette grafiği

### Keywords:

Self-organizing Maps  
Common Stocks  
Portfolio Management  
Similarity Matrix  
Silhouette Graphics

### ÖZ

Etkin bir portföy yönetimi için portföyde yer alacak hisse senetlerinin özelliklerinin bilinmesi gerekmektedir. İyi çeşitlendirilmiş bir portföyde risk azalacaktır. Çeşitlendirmenin gerçekleştirilmesi için portföyde yer alma ihtimali olan senetlerin özelliklerinin iyi bilinmesi gerekmektedir. Bu çalışmada BIST50 Endeksinde yer alan 50 adet hisse senedinin günlük standartlaştırılmış getiri ve risk değerleri kullanılmış ve senetler kümelerine ayrılmıştır. 708 seansa ilişkin kapanış fiyatı bilgisi her bir hisse senedi için Borsa İstanbul Datastore departmanından edinilmiştir. Çalışmada kullanılan son gözlem 30/06/2015 tarihine aittir. Çalışmada, kümeleme işlemini gerçekleştirmek üzere yapay sinir ağlarının özel bir çeşidi olan özdüzenleyici haritalar kullanılmıştır. Ayrıca benzerlik matrisi, dağılım grafiği, silhouette grafiği ve kümelere seçilmiş senetlerin zaman serisi grafiği çizilmiştir. Sonuçlar özdüzenleyici haritalar yönteminin başarılı bir şekilde hisse senetlerini kümelendirdiğini ve görselleştirdiğini ortaya çıkarmaktadır.

### Clustering Stocks with Self-organizing Maps: An application on Stocks Listed in BIST50 Index

### ABSTRACT

The determination of characteristics of stocks are essential for an efficient portfolio management. In a well-diversified portfolio, the risk will be minimum. It is essential to know the characteristics of stocks for a better diversification. In this study, the stocks listed in BIST50 index are clustered using their standardized mean return and standard deviation of return. Closing price along 708 trading session is retrieved from Borsa İstanbul Datastore Department. For each stock, the last observation belongs to second trading session of 30/06/2015. Self-organizing maps which is a special kind of artificial neural networks are used as clustering technique. Also similarity matrix, scatter diagram, silhouette plots and time series plots for the selected stocks are drawn. Results indicate that self-organizing maps are successful at clustering and visualizing the stocks.

<sup>†</sup> mozcalici@kilis.edu.tr (M. Özçalıcı)

## 1. Giriş

Modern portföy teorisine göre yatırımcılar portföydeki riski minimize edecek veya beklenen getiriyi maksimize edecek farklılaştırmayı sağlamalıdır (Markowitz, 1952). Söz konusu teoride bir varlığın getirisi ortalama getiri ile bir varlığın riski ise getiriye standart sapma ile ölçülmektedir. Yatırım çeşitlendikçe risk de azalmaktadır. Bu nedenle etkin bir portföy oluşturmak için çeşitlendirme yapmak gerekmektedir.

Etkin portföy belirli bir risk seviyesinde en yüksek getiriyi sağlayan veya belirli beklenen gelir düzeyinde en düşük riske sahip portföylerdir (Brigham ve Ehrhardt, 2011). Hisse senetlerinin riski getirilerindeki standart sapma ile ölçülmektedir. Risk, sistematik ve sistematik olmayan risk olmak üzere iki türdedir. Sistematik risk, piyasadaki bütün işletmeleri etkileyen risktir. Buna karşın sistematik olmayan risk, bireysel şirketlerin sahip oldukları risklerdir. İyi bir çeşitlendirme ile sistematik riskler elenmektedir ve geriye sadece sistematik olmayan risk kalmaktadır (Brealey, Myers ve Marcus, 2012). Çeşitlendirme aynı zamanda uluslararası yatırımlarda da önemli bir rol oynamaktadır. Uluslararası çeşitlendirmede de temel amaç performansı birbirleri ile ilişkili olmayan yabancı yatırımları seçmektir (Madura, 2010).

Hisse senetlerinin karakteristik özelliklerinin belirlenmesi ve buna uygun olarak kümelenebilir etkin bir yatırım stratejisi, başka bir ifade ile etkin bir çeşitlendirme için gereklidir (Liu, Cai and Luo, 2012). Hisse senetlerine ilişkin tanımlayıcı bilgilerin açığa çıkartılması ise etkin portföylerin oluşturulmasında yol gösterici bir rol oynamaktadır. Kümeleme analizi ile birlikte birbirlerine benzer özelliklere sahip olan hisse senetleri bir arada toplanabilir ve böylelikle etkin portföy oluşturmak için gerekli ilk şart sağlanabilir.

Kümeleme analizi, gözlemleri kendi içinde homojen, kendi aralarında ise heterojen özelliğe sahip olacak şekilde ayrı kümelerde toplayan tekniklerdir (Gan vd., 2007:3). Farklı kümeleme yöntemleri mevcuttur. Bu yöntemlerden bazıları; k ortalamalar yöntemi, hiyerarşik kümeleme, fuzzy c ortalamalar yöntemi ve özdüzenleyici haritalardır. Özdüzenleyici haritalar eğitim sürecinde denetimsiz eğitim uygulayan yapay sinir ağlarının özel bir çeşididir. Başka bir ifade ile bir çıktı setine (öğretmene) ihtiyaç duyulmamaktadır ve bu nedenle de kümeleme analizi için ideal bir yaklaşım olarak tanımlanmaktadır. Özdüzenleyici haritalar literatürde farklı işletmecilik ve finansman problemlerinin çözümünde başarıyla kullanılmıştır. Örnek olarak perakende satış yapan işletmelerin müşterilerinin kümelendirilmesi (Taşkın ve Emel, 2010), e-bankacılık müşterilerinin kümelendirilmesi (Niyagas vd., 2006), banka müşterilerinin kümelendirilmesi (Özçalıcı, 2011), endüstriyel piyasaların bölümlendirilmesi (Kuo vd., 2002), online

alışveriş piyasasının bölümlendirilmesi (Kim ve Ahn, 2008), piyasa bölümlerinin görselleştirilmesi (Hanafizadeh and Mirzazadeh, 2011), iflas edecek bankaların önceden tahmini (Lopez-Iturriaga ve Pastor-Sanz, 2015) gibi alanlar göz önünde bulundurulabilir.

Özdüzenleyici haritalar, Türkiye dışındaki borsalarda işlem gören hisse senetlerinin kümelendirilmesinde başarıyla uygulanmasına rağmen (Nanda vd., 2010; Hsu, 2011; Huang ve Tsai, 2009; Hsu vd., 2009), Borsa İstanbul'da işlem gören hisse senetlerinin kümelendirilmesi için daha önce kullanılmamıştır. Bu çalışma ile literatürdeki bu eksiklik giderilmeye çalışılacaktır.

Bu çalışmada BIST50 Endeksinde yer alan firmaların tarihi getiri ve risk değişkenleri hesaplanmış ve söz konusu değişkenleri dikkate almak suretiyle hisse senetleri kümelere ayrılmıştır. Ayrıca küme sonuçları farklı görsel tekniklerle görselleştirilmiştir. Çalışma 5 bölümden oluşmaktadır. Bu giriş bölümünden sonra ikinci bölümde özdüzenleyici haritalar ile ilgili yapılmış önceki çalışmalar incelenecektir. Üçüncü bölümde özdüzenleyici haritalar tanıtılacaktır. Dördüncü bölümde veri seti, analiz sonuçları ve grafikler yer alacaktır. Beşinci bölümde ise sonuç yer almaktadır.

## 2. Literatür Taraması

Hisse senetlerinin kümelere ayrılması, araştırmacıların ilgisini çeken konulardan bir tanesidir. Bu nedenle de hisse senetlerinin kümelendirilmesi ile ilgili çalışmalara literatürde sıklıkla rastlanmaktadır.

Gavrilov vd. (2000) S&P500 indekisinde yer alan hisse senetlerini kümelemiştir. Kümeleme yöntemi olarak birleştirici hiyerarşik kümeleme (hierarchical agglomerative clustering) yöntemini kullanmışlardır. Çalışmaları sonucunda normalleştirme işlemi ile birlikte daha anlamlı kümelerin ortaya çıktığını ifade etmektedirler.

Basalto vd. (2005) ise çalışmada kaotik harita senkronizasyonu (chaotic map synchronization) adı verilen yöntemi kullanmışlar ve Dow Jones Endeksinde yer alan 30 adet hisse senedini kümelere ayırmışlardır. Çalışmalarının sonucunda hisse senetlerinin sermaye malları ve hammadde ile finans, hizmetler, sağlık, teknoloji, holdingler ve konjonktürel olmayan tüketim malları olmak üzere iki gruba ayrılabilirliklerini raporlamışlardır.

Xiu vd. (2009), Şangay borsasındaki hisse senetlerini fraktal teorisinden faydalanmak suretiyle kümelendirmişlerdir. Çalışma sonucunda, benzer özellik gösteren hisse senetlerinin aynı kümede bir araya getirildiği ve bu bilginin yatırımcılar tarafından kayıplardan kaçınmak için kullanılabileceği raporlanmaktadır.

Nanda vd. (2010) Bombay hisse senedi piyasasındaki 2007-2008 tarihlerinde işlem gören hisse senetlerini kümelerle ayırmışlardır. Çalışmalarında k-ortalamlar, özdüzenleyici haritalar ve bulanık c-ortalamlar yöntemlerini kullanmışlar ve kümelerin performanslarını, farklı uzaklık ölçülerine göre ölçmüşlerdir.

Liu vd. (2012) CSI300 indeksinde yer alan hisse senetlerini kümelemişlerdir. Kümeleme yöntemi olarak izometrik özellik haritalama (ISOMAP - isometric feature mapping) yöntemini kullanmışlardır. Eylül 2009 ile Kasım 2011 arasındaki verileri kullanmışlardır ve veri setindeki karmaşıklığın başarılı bir şekilde kümelendiğini raporlamışlardır. Girdi olarak 300 adet hisse senedinin kapanış fiyatlarını kullanmışlardır.

Rea ve Rea (2014), hisse senetlerinin korelasyon matrisini yeni bir kümeleme yaklaşımı olan Neighbor-Net ile görselleştirmişlerdir. Yeni Zelanda Hisse Senedi Piyasasında listelenen 48 adet hisse senedinin korelasyon matrisini hesaplamışlardır ve bu matrisi Neighbor-net ile görselleştirmişlerdir. Çalışmalarında ayrıca hiyerarşik kümeleme yöntemini de kullanmışlardır. Sonuçta önerdikleri yöntemin hisse senetleri arasındaki korelasyonun daha iyi anlaşılmasını sağladığını raporlamışlardır. Aghabozorgi ve Teh (2014) üç aşamalı yeni bir sınıflandırma yöntemi kullanmak suretiyle, fiyatları birlikte hareket eden hisse senetlerini belirlemişlerdir. Çalışmalarında, Kuala Lumpur borsasında işlem gören hisse senetlerine ilişkin veriler kullanmışlardır. Çalışmalarında önerdikleri yöntemin performansını istatistiksel uzaklık ölçü birimleri ile değerlendirmişlerdir.

Türkiye’de ise Irmak ve Çetin (2009), çalışmalarında korelasyona dayalı ve tam bağıntı yöntemini kullanan hiyerarşik kümeleme analizi uygulamışlardır. Kümeleme analizinin, halihazırda var olan sektörel kümelemelerden daha farklı sonuçların ortaya çıkardığını raporlamışlardır. Çalışma sonuçlarının riskin azaltılmasında etkin bir şekilde kullanılabileceğini ifade etmişlerdir. Karabayır ve Doğanay (2010), BIST100 Endeksinde işlem gören hisse senetlerini sınıflandırmak amacıyla, risk ve getiri bilgilerinin girdi olarak kullanıldığı hiyerarşik kümeleme analizi uygulamışlardır. Sonuçta 10 adet küme belirlenmiş ve ilk zaman aralığında oluşturulan bir portföyün ikinci zaman aralığında kazanç sağladığını raporlamışlardır. Özdüzenleyici haritalar ise Türkiye’de Vatansever ve Büyüklü (2009) tarafından kullanılmıştır. Yazarlar çalışmalarında, görsel veri madenciliği tekniklerinin kümeleme analizinde kullanımı detaylı bir şekilde raporlamaktadırlar. Ayrıca söz konusu yöntemleri Türkiye’de yer alan 81 ildeki 918 ilçenin gelişmişlik düzeylerine göre kümelemede uygulamışlardır. Sonuç olarak ilçelerin gelişmiş ve gelişmekte olan ilçeler

olmak üzere iki kümede başarılı bir şekilde kümelendiğini raporlamışlardır.

Yukarıda sıralanan ve Türkiye’de gerçekleştirilen çalışmalar (Irmak ve Çetin (2009), Karabayır ve Doğanay (2010), Vatansever ve Büyüklü (2009)) göz önünde bulundurulduğunda, özdüzenleyici haritalar yönteminin, Türkiye’de gerçekleştirilen çalışmalarda hisse senetlerini kümelemek amacıyla kullanılmadığı ortaya çıkmaktadır. Bu açıdan bu çalışma bir ilk niteliği taşımaktadır.

### 3. Metodoloji

Özdüzenleyici haritalar, yapay sinir ağlarının özel bir çeşididir. Teuvo Kohonen tarafından 1982 yılında tanıtıldığı için Kohonen ağırları olarak da adlandırılmaktadır (Kohonen, 1982). Özdüzenleyici haritalar, veri setinde var olan fakat hâlihazırda bilinmeyen örüntülerin (patterns) ortaya çıkarılmasında kullanılmaktadır (Cabanès ve Bennani, 2010).

Özdüzenleyici haritalar birbirleriyle tamamen bağlanmış iki katmandan oluşmaktadır: Girdi Katmanı ve Kohonen Katmanı olarak da adlandırılabilen Çıktı Katmanı. Kohonen katmanı (çıkı katmanı) aynı zamanda veri setindeki kümelemenin gözlemlenmesini sağlayacak haritanın oluştuğu katmandır. Özdüzenleyici haritaların sahte kodu (pseudocode) Şekil 1’de yer almaktadır.

Girdi: D: veri seti; d: özellik haritasının boyutu;

1: Ağırlık vektörlerinin başlangıç değerleri  $w_j$   $j = 1, 2, \dots, d$  ; D veri setinden rastgele gözlemler seçilerek belirlenir;

2: Aşağıdaki döngü (3-5) haritada gözle görülür bir değişiklik olmayana kadar tekrarlanır

3: D veri setinden bir x gözlemi belirli bir olasılık dâhilinde seçilir;

4: Kazanan nöron  $i(x)$ , s zaman diliminde, en düşük Öklid uzaklık ölçütü kullanılarak belirlenir:

$$i(x) = \arg \min_{1 \leq j \leq d} \|x - w_j\|;$$

5: Bütün nöronların ağırlık vektörleri yenilenir;

6: Özellik haritasından çıkılır.

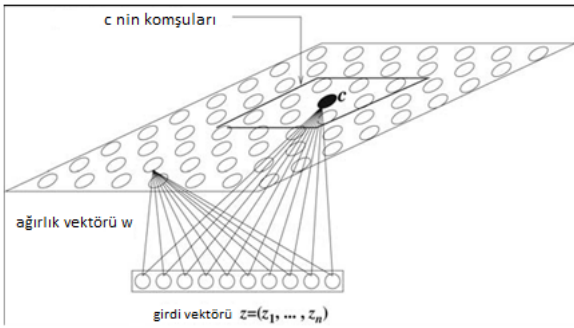
**Şekil 1. Özdüzenleyici Haritaların Sahte Kodu.**  
**Kaynak:** Gan vd. (2007).

Özdüzenleyici haritaların eğitim sürecinde nöronlar birbirleriyle rekabet halindedir. Rekabet süreci sonunda kazanan nöron ve çevresindeki nöronlar ağırlıklarını, sunulan gözleme göre yenilemektedirler (Kohonen, 2001). Eğitim süreci sonunda çıktı katmanında, benzer

örüntüler aynı nöronlarda veya komşu nöronlarda temsil edilecektir (Folguera vd 2015).

Şekil 2’de özdüzenleyici haritalardaki girdi ve çıktı katmanı şematize edilmiştir. Şekilde 9x7 nörondan oluşan çıktı (Kohonen) katmanı yer almaktadır. Girdi vektörü  $z$  ile temsil edilmiştir ve bu vektörde değişkenler yer almaktadır ( $z_n$  ifadesinde  $n$  adet değişken olduğu ifade edilmektedir). Ağırlık vektörü ise  $w$  ile temsil edilmiştir ve her bir değişkenden her bir nörona giden ağırlıkları barındırmaktadır.  $c$  kazanan nöronu temsil etmektedir ve çevresindeki nöronlar  $c$ ’nin komşularıdır.

Özdüzenleyici haritaların en önemli avantajı, çok boyutlu bir veri setini iki boyuta indirgeyebilmesidir. Bu durum karar vericilerin veri setinin tamamını kolayca değerlendirmelerine olanak sağlamaktadır.



**Şekil 2. Özdüzenleyici Haritalar Yönteminde Girdi ve Çıktı Katmanı.**

**Kaynak:** Badran vd (2004:407).

Özdüzenleyici haritaların parametrelerine kullanıcının karar vermesi gerekmektedir. Bu parametrelerin belirlenmesi için de hazır bir kural söz konusu değildir. Bu durum özdüzenleyici haritalar için dezavantaj olarak nitelendirilebilir. Kullanıcının karar vermesi gereken parametreler arasında, çıktı katmanının boyutu, çıktı katmanının şekli, eğitim algoritması da sayılabilir. Çıktı katmanının boyutuna deneme yanılma yoluyla karar verilebilmektedir. Genellikle, dikdörtgen şeklindeki çıktı katmanı, karesel şeklindeki çıktı katmanına göre daha iyi sonuç vermektedir (Kohonen, 2001). Eğitim algoritması için de sıralı (*sequential*) veya yığın (*batch*) eğitim algoritması kullanılabilir. Sıralı eğitim algoritmasında ağırlıklar, her gözlem için güncellenirken, yığın eğitimde, nöronların ağırlıkları ancak bütün veri seti işleme sunulduğunda güncellenmektedir (Haga vd., 2015). Yığın eğitim algoritması, sıralı algoritmadan daha istikrarlı (*stable*) sonuçlar üretebildiği için (Kohonen, 2001), çalışmada yığın eğitim algoritması kullanılmıştır.

## 4. Analiz

### 4.1. Çalışmanın Modeli

Çalışmanın modeli Şekil 3’te görselleştirilmiştir. Çalışmada ilk olarak BİST50 Endeksindeki hisse senetlerinin tarihi fiyat bilgilerine erişim sağlanmıştır. Her bir hisse senedi için risk ve getiri değerleri hesaplanmış ve özdüzenleyici haritalar yöntemi yardımıyla hisse senetleri kümelerine ayrılmıştır. Bu kümeleme işlemi ile birlikte, benzer özellik gösteren hisse senetleri aynı kümelerde toplanmıştır. Başka bir ifade ile hisse senetleri, homojen özellik gösterecek şekilde kümelendirilmiştir. Kümeler ise birbirleriyle heterojen özellik göstermektedir. Her bir hisse senedinin hangi kümede yer aldığı belirlendikten sonra (küme indeksleri atandıktan sonra), kümeleme işleminin ne kadar başarılı olduğunu ortaya çıkarmak amacıyla, şu grafikler oluşturulmuştur:

- Dağılım Grafikleri
- Benzerlik Matrisi
- Zaman Serisi Grafiği
- Silhouette Grafiği

### 4.2. Veri Seti

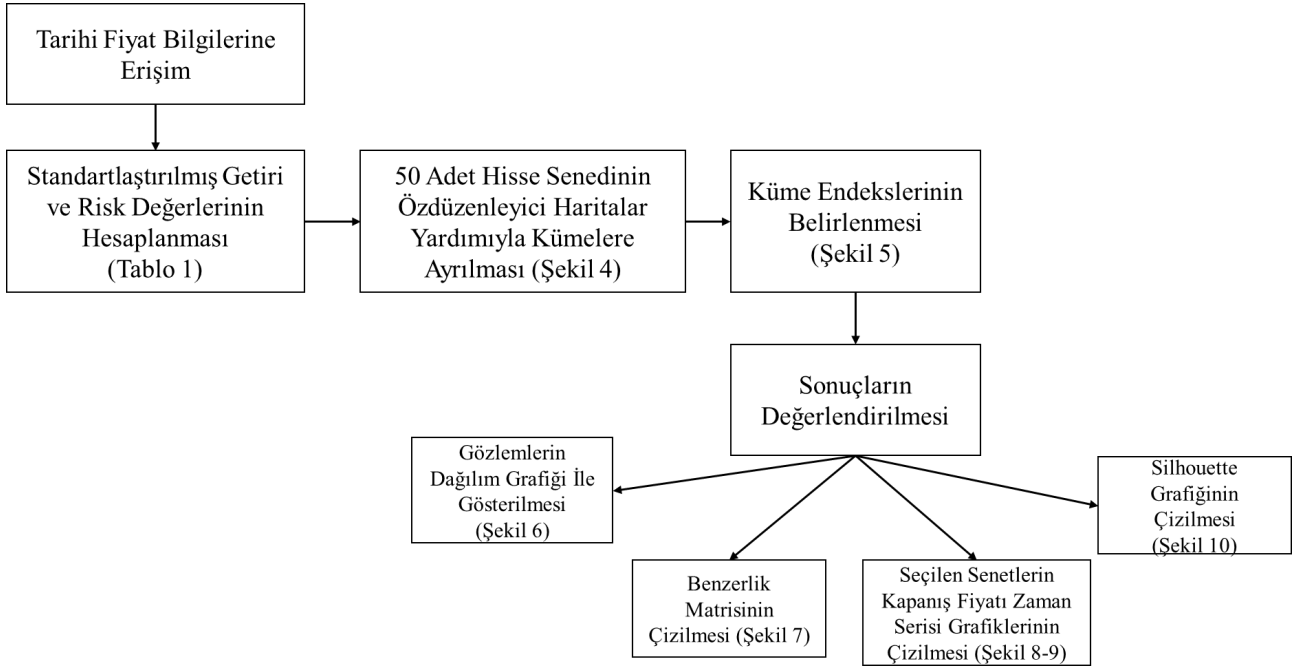
Çalışmada 2015 yılının ilk çeyreğinde BİST50 Endeksinde yer alan hisse senetleri tercih edilmiştir. BİST50 Endeksinde en fazla işlem gören hisse senetleri dâhil edilmektedir. Söz konusu bu endekste yer alan hisse senetlerinin likiditesinin fazla olduğunu söylemek mümkündür. BİST50 Endeksinde yer alan hisse senetlerine ilişkin tarihi kapanış fiyatı BIST Datastore bölümünden temin edilmiştir. Çalışmada 01/01/2014 ile 30/06/2015 tarihleri arasındaki 708 seansa ilişkin kapanış fiyatları kullanılmıştır.

Çalışmada kullanılan değişkenlerden biri ortalama getiridir ve ortalama getiri Formül 1 yardımıyla hesaplanmıştır.

$$\bar{r} = \frac{\sum_{i=1}^n r_i}{n} \quad (1)$$

Formülde  $r_i$   $i$  gününe ilişkin hisse senedi getirisini,  $n$  ise toplam seans sayısını temsil etmektedir. Hisse senedinin getirisi ( $r_i$ ) ise Formül 2 ile hesaplanmıştır.

$$r_i = \log \left( \frac{p_{t+1}}{p_t} \right) \quad (2)$$



**Şekil 3. Çalışmanın Modeli**

Formülde  $p_t$  t gününe ilişkin kapanış fiyatını temsil etmektedir.

Çalışmada kullanılan bir diğer değişken ise hisse senetlerinin risk değeridir ve risk getirideki standart sapma ile aşağıdaki formül yardımıyla ölçülmüştür.

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (r_i - \bar{r})^2}{n - 1}} \quad (3)$$

Formülde  $r_i$  ifadesi hisse senedinin  $i$  günde ki getiriyi,  $\bar{r}$  ise hisse senedinin ortalama getirisini ve  $n$  ifadesi ise hesaplamaya konu olan seans sayısını temsil etmektedir.

Daha büyük değerlere sahip değişkenler, özellikle, yapay sinir ağlarının eğitimi ve regresyon analizlerinde, daha düşük değerlere sahip özelliklerle bir arada kullanıldıklarında, analizde egemen hale gelmektedirler (Liu ve Hu, 2012). Bu olumsuz durumu ortadan kaldırmak amacıyla, birbirlerinden farklı ölçeklere sahip olan risk ve getiri değerleri, Formül 4 yardımıyla standartlaştırılmıştır.

$$x_y = \frac{x_e - x_{min}}{x_{max} - x_{min}} \quad (4)$$

Formülde  $x_y$  ifadesi yeni değeri,  $x_e$  eski değeri,  $x_{min}$  en küçük değeri,  $x_{max}$  en yüksek değeri ifade etmektedir. Standartlaştırma işlemi sonrasında değişkenlerin yeni aralığı [0,1] olmuştur.

Tablo 1’de her bir hisse senedinin kodu, çalışmada kullanılan dönem boyunca standartlaştırılmış ortalama getiri ve risk değerleri yer almaktadır.

#### 4.3. Özdüzenleyici Haritalar ile Hisse Senetlerinin Kümelendirilmesi

Çalışmada, kümeleme yöntemi olarak yapay sinir ağları tabanlı özdüzenleyici haritalar kullanılmıştır. Özdüzenleyici haritalarda, harita boyutuna deneme yanılma yoluyla karar verilmiştir. Birçok deneme sonucunda 10x5 boyutlarındaki haritanın en okunaklı harita olduğu sonucuna varılmıştır. Çalışmadaki analizi gerçekleştirmek ve şekilleri çizmek için MATLAB yazılımı kullanılmıştır.

Şekil 4’de, U matris, hit çizelgesi ve girdi panelleri yer almaktadır. U matris kümelerin olduğu Kohonen katmanıdır ve gözlemlerin birbirleri arasındaki uzaklığı temsil etmektedir (Şekil 4-a). U matriste koyu renkler daha büyük uzaklığı gösterirken, açık renkler birbirlerine yakın olan gözlemleri içermektedir. Başka bir ifade ile veri setinin kaç adet kümeye ayrıldığı U matrisi ile belirlenmektedir. U matristeki koyu renkli bağlantılar küme sınırlarını göstermektedir. Hit çizelgesinde ise nöronlarda var olan gözlem sayısı yer almaktadır (Şekil 4-b). Bu çizelge sayesinde hangi nöronda hangi hisse senedinin olduğunu belirlemek mümkündür.

**Tablo 1. Çalışmada Kullanılan Veri Seti**

Sıra No.	Hisse Senedi Kodları	Ortalama Getiri*	Risk*	Sıra No.	Hisse Senedi Kodları	Ortalama Getiri*	Risk*
01	AEFES	52.036	10.255	26	ISCTR	59.25	9.497
02	AKBNK	52.073	4.349	27	KCHOL	52.852	2.658
03	AKSA	49.206	0	28	KOZAA	54.078	8.364
04	ARCLK	35.734	21.449	29	KOZAL	50.495	15.18
05	ASELS	57.392	9.454	30	KRDMD	47.616	6.863
06	BAGFS	36.496	6.966	31	MGROS	58.642	10.003
07	BIMAS	69.908	7.924	32	OTKAR	46.255	11.357
08	BIZIM	66.388	7.37	33	PETKM	68.104	5.022
09	BJKAS	53.728	12.047	34	PGSUS	0	100
10	BRISA	53.702	10.739	35	SAFGY	46.946	17.774
11	CCOLA	63.583	16.175	36	SAHOL	14.431	0.645
12	DOAS	61.268	0.7	37	SISE	72.323	8.254
13	DOHOL	60.561	37.932	38	TAVHL	36.909	5.184
14	EKGYO	59.557	4.839	39	TCELL	85.121	19.986
15	ENKAI	25.215	14.538	40	THYAO	54.467	15.703
16	EREGL	63.277	3.52	41	TKFEN	67.442	14.14
17	FROTO	52.319	5.094	42	TOASO	9.769	70.414
18	GARAN	57.934	13.23	43	TRKCM	71.535	26.427
19	GOLTS	46.001	12.944	44	TSKB	90.046	15.74
20	GOODY	63.488	8.192	45	TTKOM	18.467	62.936
21	GOZDE	39.69	6.041	46	TUPRS	54.133	41.545
22	GSDHO	60.413	13.32	47	ULKER	66.037	20.127
23	GUBRF	51.817	7.343	48	VAKBN	43.372	20.711
24	HALKB	49.196	6.744	49	VESTL	41.338	10.041
25	IHLAS	66.872	4.232	50	YKBNK	100	39.973

\* Standartlaştırılmış değerler

Girdi panellerinde ise, her bir değişkenin nöronal ağırlıkları gösterilmektedir (Şekil 4-c). Girdi panellerindeki desenler, değişkenlerin birbirleri ile ilişkilerini ortaya koymaktadır. Simetrik (ters simetrik) desenlerin olması durumunda, değişkenlerin birbirleri ile aynı yönde (zıt yönde) hareket ettiğini söylemek mümkündür. Girdi panellerindeki açık renkler yüksek değerleri içermektedir.

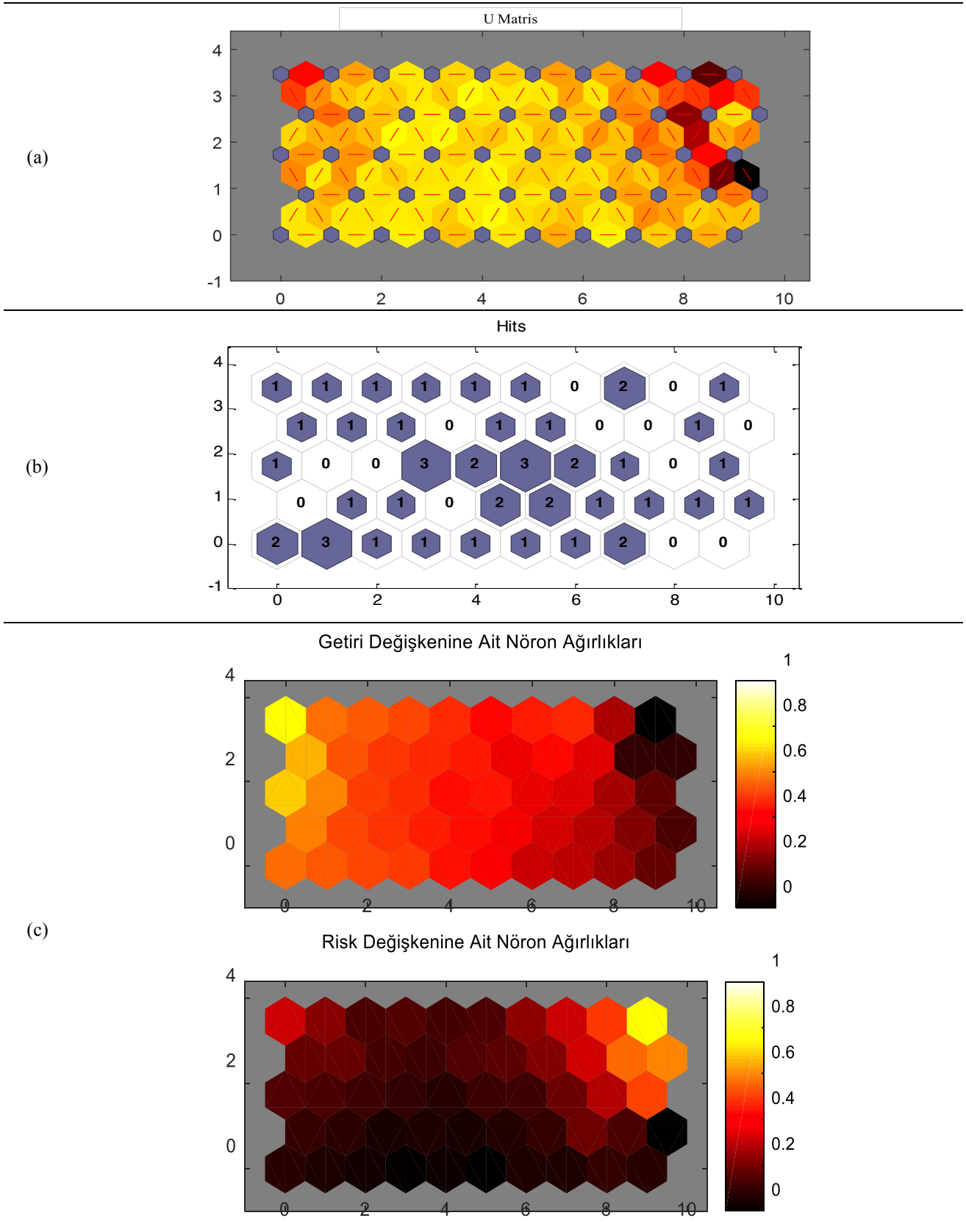
Hit çizelgesi ve U matris birlikte değerlendirildiğinde BIST50 Endeksinde işlem gören hisse senetlerini 2 grupta kümelendirilmiştir (Şekil 5).

U matriste koyu renkler daha uzak hisse senetlerini ifade ettiğinden dolayı, U matrisin sağ üst köşesinde yer alan ve etrafı koyu renkle çevrili üç adet nöronun ilk kümeyi oluşturduğunu söylemek mümkündür. Sağ üst köşedeki nöronları diğer nöronlara bağlayan uzaklıklar koyu renklerle gösterilmiştir. İki nöron arasındaki bağlantı koyu renkli ise, nöronlarda yer alan hisse senetleri birbirlerinden uzaktırlar (başka bir ifade ile birbirlerine benzememektedir). Bu nedenle koyu renkler küme sınırlarını ifade etmektedir. Söz konusu kümelere ilişkin girdi panelleri incelendiğinde sağ üst köşedeki grubun düşük getiri ve yüksek risk içerdiği görülmektedir.

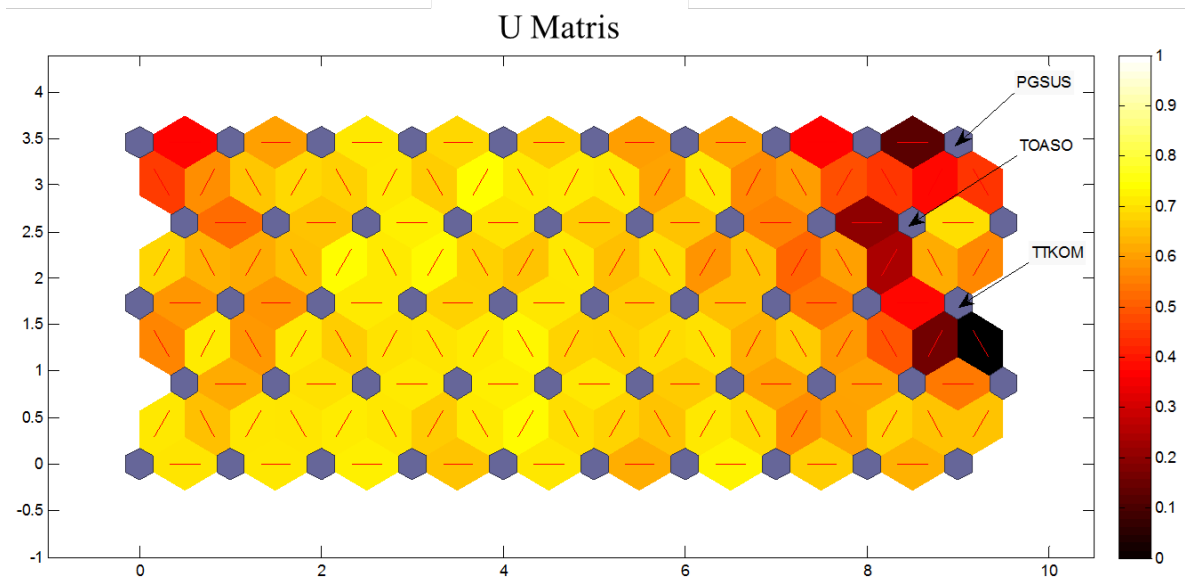
Sağ üst köşede yer alan hisse senetleri Küme 1 olarak adlandırılmıştır ve bu kümede sırasıyla PGSUS, TOASO ve TTKOM kodlu hisse senetleri bulunmaktadır. Bu üç hisse senedi dışındaki senetler, Küme 2 ile isimlendirilen kümede yer almaktadır. Her iki kümenin standartlaştırılmış ortalama getiri ve risk değerleri Tablo 2'deki gibidir. Küme 1'de yer alan hisse senetlerinin standartlaştırılmış ortalama getirisi 0.0941 iken, bu kümedeki standartlaştırılmış ortalama risk 0.7778 olarak hesaplanmıştır. Küme 2 nin standartlaştırılmış ortalama getirisi 0.5594, standartlaştırılmış ortalama riski ise 0.1214 olarak hesaplanmıştır (Tablo 2).

**Tablo 2. Kümelerin Standartlaştırılmış Ortalama Getiri ve Risk Değerleri**

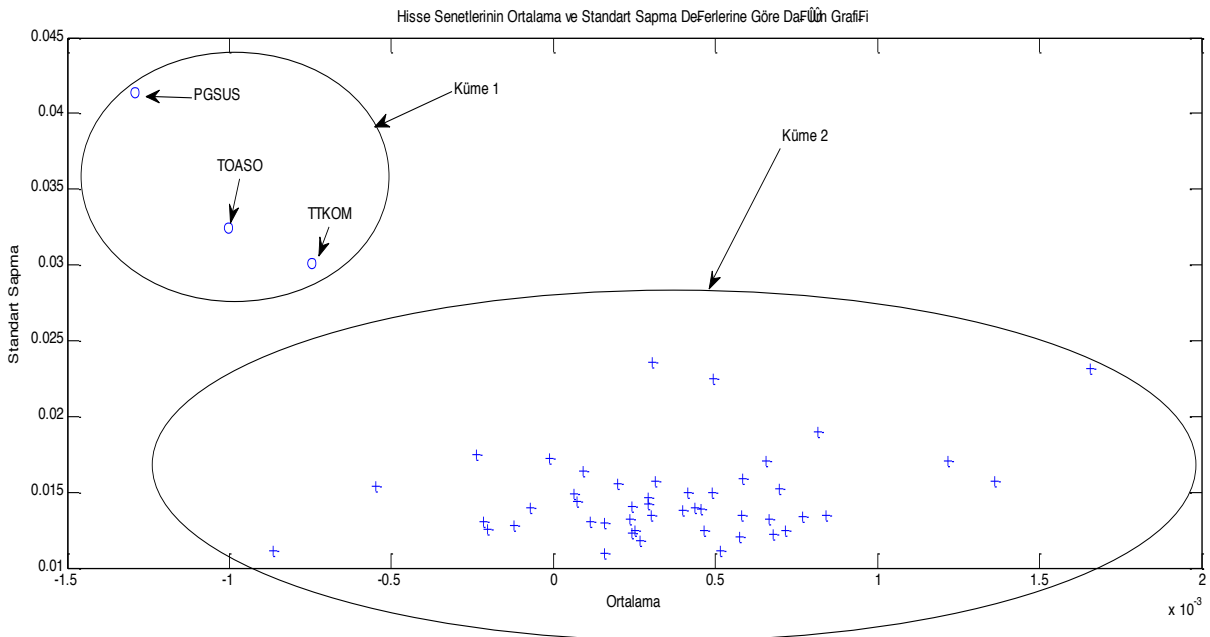
	Küme 1	Küme 2
	PGSUS TOASO TTKOM	Küme 1'in dışındaki senetler
Standartlaştırılmış Ortalama Getiri:	0,0941	0,5594
Standartlaştırılmış Ortalama Risk:	0,7778	0,1214



Şekil 4. Özdüzenleyici Haritalar (a) U Matrisi, (b) Hit Çizelgesi ve (c) Girdi Panelleri



Şekil 5. U Matriste Kümelerin Yerlerinin Belirlenmesi



Şekil 6. Hisse Senetlerinin Dağılım Grafiğinde Gösterilmesi

#### 4.4. Hisse Senetlerinin Dağılım Grafikleri

Analizi daha ileriye taşımak amacıyla, 50 adet hisse senedinin ortalama getirisi ve ortalama standart sapma değerleri dağılım grafiği üzerinde gösterilmiştir (Şekil 6). Söz konusu şekilde yatay eksende standartlaştırılmamış getiri değerleri ve dikey eksende standartlaştırılmamış standart sapma değerleri yer almaktadır. Küme 1 'e ait olan senetler "o" işareti ile gösterilirken, Küme 2 'deki senetler "+" sembolü ile

gösterilmektedir. Dağılım grafiği incelendiğinde, iki ayrı kümenin belirgin bir şekilde ortaya çıktığı görülmektedir.

#### 4.5. Benzerlik Matrisi

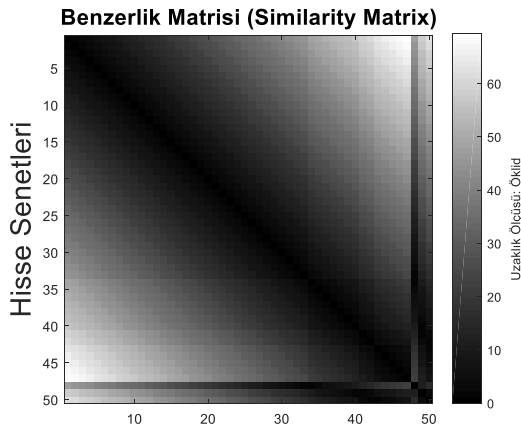
Benzerlik matrisi kümeleme performansını görsel bir şekilde değerlendirmeye olanak sağlayan yöntemlerden bir tanesidir. Bu matris çizilirken, her bir gözlemin



birbiri ile uzaklığı ölçülmektedir. Söz konusu bu uzaklıklar, gözlemlerin birbirlerine ne kadar yakın ya da uzak olduğunu göstermektedir (Kaufman ve Rousseeuw, 1990). Literatürde farklı benzerlik ölçülerinin olmasına rağmen en sık kullanılan benzerlik ölçüsü Öklit uzaklığıdır. Öklit uzaklık formülü şu şekildedir:

$$d(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{ip} - x_{jp})^2} \quad (5)$$

Formülde  $d(i, j)$  ifadesi  $i$  ve  $j$  gözlemleri arasındaki uzaklığı,  $p$  ifadesi ise kümelemeye konu olan değişken sayısını temsil etmektedir. Gözlemlerin birbirlerine yakın olma durumları bu şekilde belirlendikten sonra uzaklık değerleri kare forma getirilmiştir. Küme indeksleri sıralandıktan sonra elde edilen kare matris aşağıdaki gibi görselleştirilmiştir. Şekil 7’de açık renkli hisse senetleri birbirleri ile yüksek düzeyde benzerlik gösterirken, koyu renkli senetler birbirleri ile benzemeyen senetleri göstermektedir.



Şekil 7. Benzerlik Matrisinin Grafik ile Gösterimi

#### 4.6. Zaman Serisi Grafiği

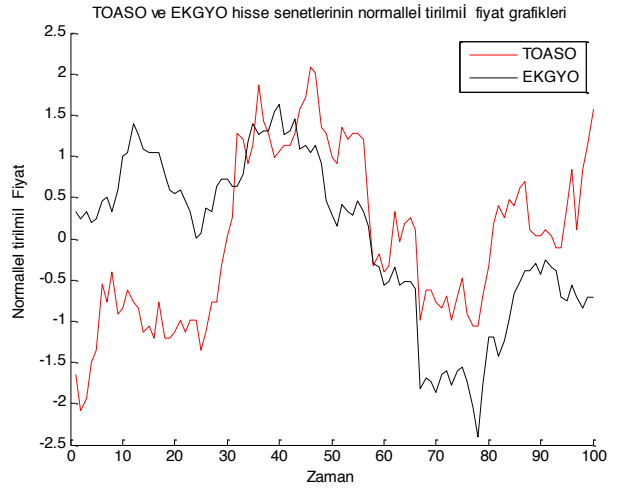
Kümeleme yönteminin hisse senetlerinin kümelemede ne kadar başarılı olduğunu değerlendirebilmek için aynı kümelerdeki (Şekil 8) ve farklı kümelerdeki (Şekil 9) bazı hisse senetlerinin son 100 güne ilişkin fiyat hareketleri aynı grafik üzerinde gösterilmiştir. Hisse senetleri kapanış fiyatları arasında ölçek farkı olduğundan, kapanış fiyatları oldukları gibi değil, normalleştirmek suretiyle grafikte yer almıştır. Normalleştirmede kullanılan formül şu şekildedir;

$$z = \frac{x - \mu}{\sigma} \quad (6)$$

Formülde,  $z$  standartlaştırılmış değeri,  $x$  standartlaştırılmamış değeri,  $\mu$  değişkenin ortalamasını,

$\sigma$  ise değişkenin standart sapmasını ifade etmektedir. Bu standartlaştırma işlemi sonrasında yeni elde edilen  $z$  serisinin ortalaması 0 (sıfır) ve standart sapması 1 (bir) dir.

Aynı kümede yer alan senetlerin normalleştirilmiş zaman serisi grafiklerinin birbirleri ile benzer özellik taşıdığı (Şekil 8) ve farklı kümelerden seçilmiş senetlerin normalleştirilmiş zaman serisi grafiklerinin (Şekil 9) farklılık arz ettiği görülmektedir. Bu grafikler, özdüzenleyici haritaların benzer senetleri aynı kümede toplayabildiğini ve kümelerin birbirleri ile farklılık arz ettiğini kanıtlar niteliktedir.

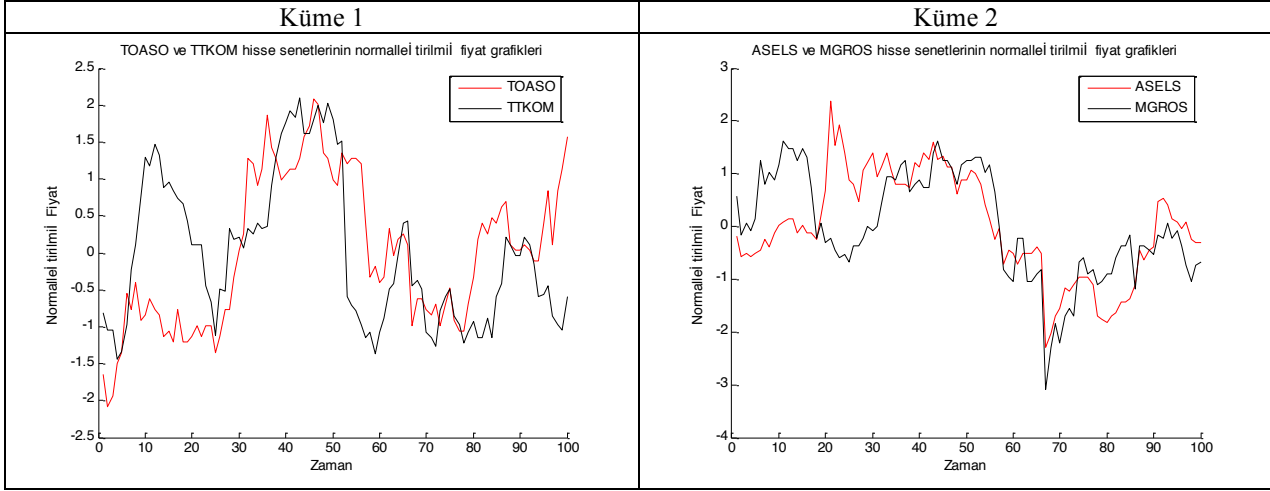


Şekil 9. Farklı Kümedeki Hisse Senetlerinin Normalleştirilmiş Fiyat Grafikleri

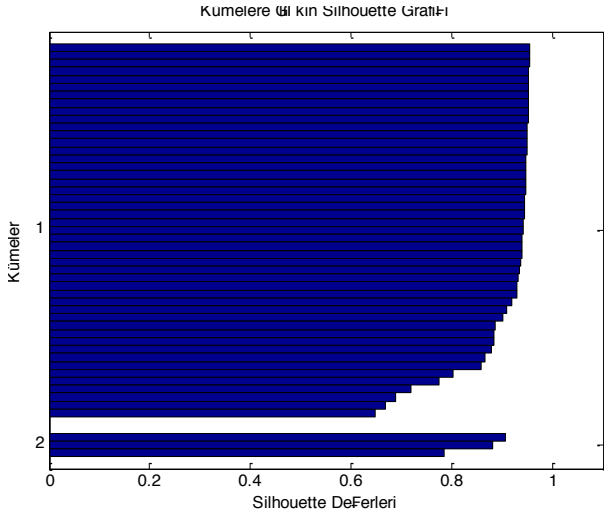
#### 4.7. Silhouette Grafiği

Silhouette grafiği, kümeleme analizinin kalitesini görsel bir şekilde değerlendirme imkânı sağlamaktadır. Silhouette grafiğinde, her bir küme bir silhouette değeri ile gösterilmektedir. Bu grafiğin çizilebilmesi için ilk önce kümelemeye konu olan her bir gözlem için silhouette değerleri hesaplanmaktadır. Silhouette değerlerinin hesaplanmasına değinmek çalışmanın hacmini arttıracaktır. Bu değerlerin nasıl hesaplanacağına ilişkin detaylı bilgiler (Kaufman ve Rousseeuw, 1990) kaynağında yer almaktadır. Silhouette değerleri -1 ile 1 arasında değerler almaktadır. Silhouette değerinin 1 olması, elemanın doğru yerde kümelendiğini göstermektedir. Değerin -1 olması ise elemanın yanlış kümede yer aldığını göstermektedir.

Şekil 10’da çalışmada kullanılan kümeleme yöntemine ilişkin Silhouette grafiği yer almaktadır. Çalışmada hesaplanan en düşük silhouette değeri 0.6481 dir ve bu değer SAHOL kodlu hisse senedine aittir. Silhouette değerlerinin yüksek çıkması, senetlerin doğru yerlerde kümelendirildiğini göstermektedir.



Şekil 8. Aynı Kümedeki Hisse Senetlerinin Normalleştirilmiş Grafikleri



Şekil 10. Kümelere İlişkin Silhouette Grafiği

## 5. Sonuç ve Tartışma

Bu çalışmanın amacı BIST50 Endeksinde yer alan hisse senetlerini kümelendirmektir. Çalışmada, denetimsiz öğrenme yöntemini uygulayan ve kümeleme için ideal bir yöntem olarak adlandırılan özdüzenleyici haritalar kullanılmıştır. Girdi olarak, hisse senetlerinin risk ve getiri değerleri kullanılmıştır. Yapılan literatür taraması, özdüzenleyici haritalar yönteminin Türkiye’de yer alan hisse senedi piyasasında işlem gören hisse senetlerinin kümelendirilmesinde kullanılmadığını ortaya çıkarmıştır. Çalışma bu açıdan bir ilk niteliği taşımaktadır.

Hisse senedi kümelemenin en önemli avantajlarından biri, hisse senedi piyasalarında varlık tahsisi (asset allocation) ve risk yönetimidir. Bu çalışmanın amaçlarından birisi de Borsa İstanbul’da işlem yapan yatırımcılara ve Borsa üyelerine bir karar destek sistemi sunmaktır. Çalışmada kullanılan yöntem birbirine benzeyen ve benzemeyen hisse senetlerini ortaya çıkarmaktadır. Bu bilgi ise yatırımcıların yüksek getiri sağlayan senetleri belirleyip, benzer özellik sergileyen diğer hisse senetlerini varlık tahsisi yaparken göz önünde bulundurmalarına olanak sağlayacaktır.

Portföy oluşturulurken, beklenen getirinin maksimizasyonu veya riskin minimizasyonu amaçlanmaktadır. Bu çalışmada uygulanan yöntem yardımıyla birbirlerine benzer hisse senetleri aynı grupta toplanmıştır. Dolayısıyla hisse senedi seçimi için yoğun çaba harcamak yerine kümelere en iyi performans gösteren hisse senetleri portföy oluşturmak için seçilebilir. Bununla birlikte özdüzenleyici haritalar ile hisse senetlerinin kümelendirilmesi hisse senedi fiyat tahmin analizinin performansını arttırmak için kullanılabilir. Başka bir ifade ile her kümeden bir adet hisse senedinin (tercihen en iyi (en yüksek getiri veya en düşük riske sahip)) seçilmesi ile birlikte etkin portföyler oluşturulabilir.

Özdüzenleyici haritaların en önemli avantajlarından biri de çok boyutlu veri setinin iki boyuta bilgi kaybetmeden indirgeyebilmesidir. Kullanılan yöntem sadece hisse senetlerini kümelere ayırmakla kalmamış aynı zamanda hisse senetlerinin risk ve getiri değerlerini bir arada görsel bir şekilde değerlendirme imkânı sunmuştur.

Çalışmada uygulanan yöntemde kümelerin belirlenmesi ve hangi gözlemlerin hangi kümelere ait olduğunu belirlemek için, kümelemeyi kullanan kişinin müdahalesi gerekmektedir. Bu bir dezavantaj olarak düşünülebilir. Fakat bu dezavantaj, istenilen küme sayısı kadar nöron kullanmak suretiyle ortadan kaldırılabilir.

### Kaynakça

- Aghabozorgi, S., Teh, Y. W. (2014). Stock market co-movement assessment using a three-phase clustering method. *Expert Systems with Applications*, 41(4 PART 1), 1301–1314. <http://doi.org/10.1016/j.eswa.2013.08.028>
- Badran, F., Yacoub, M., Thiria, S. (2005). Self-Organizing Maps and Unsupervised Classification. In G. Dreyfus (Ed.), *Neural Networks Methodology and Applications* (pp. 379–442). Berlin/Heidelberg: Springer-Verlag. [http://doi.org/10.1007/3-540-28847-3\\_7](http://doi.org/10.1007/3-540-28847-3_7)
- Basalto, N., Bellotti, R., De Carlo, F., Facchi, P., Pascazio, S. (2005). Clustering stock market companies via chaotic map synchronization. *Physica A: Statistical Mechanics and Its Applications*, 345(1-2), 196–206. <http://doi.org/10.1016/j.physa.2004.07.034>
- Brealey, R. A., Myers, S. C., Marcus, A. J. (2012). *Fundamentals of Corporate Finance*. McGrawHill.
- Brigham, E. F., Ehrhardt, M. C., (2011). *Financial Management: Theory and Practice*. South-Western Cengage Learning.
- Cabanes, G., Bennani, Y. (2010). Learning the Number of Clusters in Self Organizing Maps. Self Organizing Maps. Matsopoulos, G. (Ed.). InTech Open Access Publisher, India.
- Folguera, L., Zupan, J., Cicerone, D., Magallanes, J. F. (2015). Self-organizing maps for imputation of missing data in incomplete data matrices. *Chemometrics and Intelligent Laboratory Systems*, 143, 146–151. <http://doi.org/10.1016/j.chemolab.2015.03.002>
- Gan, G., Ma, C., Wu, J. (2007). *Data Clustering Theory, Algorithms, and Applications*. American Statistical Association.
- Gavrilov, M., Anguelov, D., Indyk, P., Motwani, R. (2000, August). Mining the stock market (extended abstract): which measure is best?. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 487-496). ACM. <http://doi.org/10.1145/347090.347189>
- Haga, J., Siekkinen, J., Sundvik, D. (2015). Initial stage clustering when estimating accounting quality measures with self-organizing maps. *Expert Systems with Applications*, 42(21), 8327–8336. <http://doi.org/10.1016/j.eswa.2015.06.049>
- Hanafizadeh, P., Mirzazadeh, M. (2011). Visualizing market segmentation using self-organizing maps and Fuzzy Delphi method - ADSL market of a telecommunication company. *Expert Systems with Applications*, 38(1), 198–205. <http://doi.org/10.1016/j.eswa.2010.06.045>
- Hsu, C. M. (2011). A hybrid procedure for stock price prediction by integrating self-organizing map and genetic programming. *Expert Systems with Applications*, 38(11), 14026–14036. <http://doi.org/10.1016/j.eswa.2011.04.210>
- Hsu, S. H., Hsieh, J. P. A., Chih, T. C., & Hsu, K. C. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications*, 36(4), 7947–7951. <http://doi.org/10.1016/j.eswa.2008.10.065>
- Huang, C. L., Tsai, C. Y. (2009). A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting. *Expert Systems with Applications*, 36(2 PART1), 1529–1539. <http://doi.org/10.1016/j.eswa.2007.11.062>
- Irmak, S., Çetin, K. (2009). Hisse senetlerinin korelasyon uzaklıklarına dayalı olarak kümelmesi. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1), 395–406.
- Karabayır, M. E., Doğanay, M. (2016). Kümeleme Analizi İle Portföy Seçimi: İMKB-100 Endeksi Üzerine Bir Çalışma (Portfolio Selection with Cluster Analysis: A Study on Istanbul Stock Exchange-100 Index).
- Kaufman, L., Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis* (Wiley Series in Probability and Statistics).
- Kim, K. jae, Ahn, H. (2008). A recommender system using GA K-means clustering in an online shopping market. *Expert Systems with Applications*, 34(2), 1200–1209. <http://doi.org/10.1016/j.eswa.2006.12.025>
- Kohonen, T. (2001). Self-Organizing Maps. In *Springer Series in Information Sciences* (Vol. 30, p. 501). <http://doi.org/10.1007/978-3-642-56927-2>
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59–69. <http://doi.org/10.1007/BF00337288>
- Kuo, R. J., Ho, L. M., Hu, C. M. (2002). Cluster analysis in industrial market segmentation through artificial neural network. *Computers & Industrial Engineering*, 42(2-4), 391–399.
- Liu, J.N.K., Hu, Y. (2012). Application of feature-weighted Support Vector regression using grey correlation degree to stock price forecasting. *Neural Computing and Applications*, 22(1), 143–152. <http://doi.org/10.1007/s00521-012-0969-3>
- Liu, R., Cai, H., Luo, C. (2012). Clustering Analysis of Stocks of CSI 300 Index Based on Manifold Learning. *Journal of Intelligent Learning Systems and Applications*, 04(02), 120–126. <http://doi.org/10.4236/jilsa.2012.42011>
- López-Iturriaga, F. J., Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of US commercial banks. *Expert Systems with Applications*, 42(6), 2857–2869. <http://doi.org/10.1016/j.eswa.2014.11.025>

- Madura, J. (2010). *International Corporate Finance*. South-Western Cengage Learning.
- Markowitz, H. M. (1952). Portfolio selection. *Journal of Finance*, 7(1), 77–91. <http://doi.org/10.2307/2329297>
- Nanda, S. R., Mahanty, B., Tiwari, M. K. (2010). Clustering Indian stock market data for portfolio management. *Expert Systems with Applications*, 37(12), 8793–8798. <http://doi.org/10.1016/j.eswa.2010.06.026>
- Niyagas, W., Srivihok, A., Kitisin, S. (2006). Clustering e-Banking Customer using Data Mining and Marketing Segmentation. *ECTI Transactions on Computer and Information Technology*, 2(1), 63-69.
- Özçalıcı, M. (2011). *Özdüzenleyici haritalar yöntemi ile banka müşterilerinin bölümlendirilmesi*. Gaziantep Üniversitesi.
- Rea, A., Rea, W. (2014). Visualization of a stock market correlation matrix. *Physica A: Statistical Mechanics and Its Applications*, 400, 109–123. <http://doi.org/10.1016/j.physa.2014.01.017>
- Taşkın, Ç., Emel, G. G. (2010). Veri madenciliğinde kümeleme yaklaşımları ve Kohonen Ağları İle perakendecilik sektöründe bir uygulama. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 15(3), 395–409.
- Vatansever, M., Büyüklü, A. H. (2009). Görsel veri madenciliği tekniklerinin kümeleme analizlerinde kullanımı ve uygulanması. *Sigma : Mühendislik ve Fen Bilimleri Dergisi*, 27(2), 83–104.
- Xiu, Z., Hong, P., Zhen, Z. (2009). Clustering in Stock Market Based on Fractal Theory. In Proceedings of the eighth international conference on machine learning and cybernetics (pp. 12–15).