



BORSA İSTANBUL İŞLETMELERİNİN VERİ MADENCİLİĞİ İLE KÜMELENMESİ

CLUSTERING THE COMPANIES LISTED ON STOCK EXCHANGE ISTANBUL BY DATA MINING

Meltem KARAATLI¹, Ece ALTINTAŞ²

Öz

Günümüzde birçok kurum ve kuruluş tarafından çeşitli şekillerde toplanan bilgiler veri yığınları halinde saklanmaktadır. Veri madenciliği, veri tabanlarında saklanan verileri kullanarak tahminler yapabilen bir süreçtir. Veri madenciliği, birçok piyasada olduğu gibi hisse senetleri piyasasında da işletmeleri rakiplerinden avantajlı hale getirebilmektedir. Bu çalışmada, Borsa İstanbul'da işlem gören hisse senetlerinin günlük artış, azalış ve sabit kalma durumları göz önüne alınarak bir veri seti oluşturulmuştur. Bu veri setinin, anlamlı bir kümeleme oluşturması ve oluşan kümelerin sektör ve işletme açısından analiz edilmesi amaçlanmıştır. Bu doğrultuda, veri madenciliği süreci kümeleme yöntemi algoritmalarından biri olan "beklenti maksimizasyonu" algoritması kullanılarak 134 işletme, hisse senedi fiyatları yükselişi "düşük", "orta düzeyde" ve "yüksek" olarak üç kümeye ayrılmıştır. Çalışmanın sonucunda işletmelerin bulunduğu kümeler sektör ve işletme açısından yorumlanmıştır.

Anahtar Kelimeler: Veri Madenciliği, Kümeleme, Borsa İstanbul, Hisse Senetleri, Beklenti Maksimizasyonu.

Abstract

Nowadays, information, collected by many institutions and organizations is stored in the form of data stacks. Data mining is a process that can make estimates using data stored in databases. Data mining in stock markets is as advantageous as many other markets in terms of gaining competitive advantage. In this study, a data set was prepared considering the daily increase, decrease and steady state of stocks traded in Istanbul Stock Exchange. By using this dataset, it is aimed to be able to create meaningful clustering and to analyze the clusters formed in terms of sector and business. In this direction, 134 firms using the "maximization of expectation" algorithm, which is one of the data mining process clustering method algorithms, are divided into three groups as "low", "medium level" and "high" stocks. As a result, the clusters, are interpreted in terms of sector and business.

Keywords: Data Mining, Clustering, Istanbul Stock, Exchange Stocks, Expectation Maximization.

¹ Doç. Dr., Süleyman Demirel Üniversitesi, meltemkaraatli@sdu.edu.tr

² ecealtintas1116@gmail.com

1. GİRİŞ

Geçmişten günümüze insanlar çevresinden bilgiler edinme ve edindiği bilgileri etrafındaki insanlarla paylaşma faaliyetlerinde bulunmuşlardır. Günümüzde bu bilgilerin dijital ortamlara aktarılmasıyla oluşan veri ambarlarının yaygınlaşması ile birlikte büyük miktarda bilginin güvenilir ve uzun süreli depolanması sağlanabilir duruma gelmiştir.

Depolanan verilerden anlamlı bilgiler çıkarmak için birçok yöntem kullanılabilir. Veri madenciliği ise, en çok tanınan ve kullanılan yöntemleri içerisinde barındırmaktadır. Bu yöntemler; sınıflama, regresyon analizi, kümeleme, birliktelik kuralları ve ilişki analizi olarak sayılabilir.

Veri madenciliği yöntemlerinin kullanımı günümüzde geniş bir alana yayılmıştır. Bu alanlardan birisi de içerisinde büyük veri tabanlarını barındıran menkul kıymet borsasıdır. Sermaye Piyasası Kurulu'nun izniyle halka arz olan hisse senetleri, sahiplerine işletmelerdeki paylarını göstermek için verilen kıymetli evraktır. Hisse senedi hem şirket açısından hem de yatırımcılar açısından önemlidir. Çünkü hisse senetleri çıkarmak şirketler için önemli bir sermaye kaynağı, yatırımcılar için ise gelir elde etme aracıdır.

Bu çalışmada, Borsa İstanbul'da işlem gören hisse senetlerinin günlük fiyat artış, azalış ve sabit kalma durumları göz önüne alınarak bir veri seti oluşturulmuştur. Bu veri setinin, anlamlı bir kümeleme oluşturması ve oluşan kümelerin sektör ve işletme açısından analiz edilmesi amaçlanmıştır.

2. LİTERATÜR

Literatürde hisse senetleri ve veri madenciliği ile ilgili yapılan çalışmalardan bazılarını aşağıda yer verilmiştir.

Aktürk çalışmasında, borsa ile ilgilenen kişilerin üzerinde oluşan risk düzeyini en aza indirmeyi amaçlamıştır. Hisse senetlerinin risk seviyelerini hesaplamak için Türkiye Cumhuriyet Merkez Bankası'ndan alınan döviz kuru bilgileri, İstanbul Menkul Kıymetler Borsası tarafından sağlanmış hisse senetleri ve endeks günlük kapanış değerleri kullanılmıştır. Çalışmada veri madenciliği, korelasyon ve volatilité hesaplamaları ile yapay zeka ve istatistiksel işlemler kullanılmıştır. Çalışmanın sonucunda, hisse senetlerinden birbirine benzer olanlar bulunmuştur (Aktürk, 2008, 1-88).

Aktürk ve Korukoğlu çalışmalarında, İstanbul Menkul Kıymetler Borsası'ndaki 2005 yılı verileri ile Türkiye Cumhuriyet Merkez Bankası'na ait verileri kullanarak fiyat değişimlerine paralellik gösteren hisse senetlerinin bulunmasına yönelik bir çalışma yapmışlardır. Çalışmada Ulusal 100 Endeksi referans olarak belirlenmiş, kural tabanlı algoritmalar kullanılmıştır. Fiyat değişkenlerine bağlı olarak değişen, benzerlik gösteren hisse senetleri belirlenmiştir (Aktürk, Korukoğlu, 2008, 113-119).

Küçüksille, İstanbul Menkul Kıymetler Borsası'nda 1995 - 2007/6 yılları arasında işlem gören 122 şirketin hisse senetleri aylık ortalama getirilerini kullanarak genetik algoritma yoluyla portföyler hesaplamıştır. Piyasanın risksiz faiz oranının altından getiri sağlamanın sebebi; Sharpe, Treynor ve Jensen performanslarının piyasanın üzerinde gerçekleşmesidir. Daha sonra ortaya çıkan portföy performansları, İMKB ulusal 100 Endeksi 2007/7 - 2008/12 dönemiyle karşılaştırmaya çalışılmıştır. İMKB'de işlem gören hisse senetlerinden oluşturulan portföylerin veri madenciliğiyle de hesaplanabileceği sonucuna varılmıştır (Küçüksille, 2009, 1-118).

Albayrak ve Yılmaz, İstanbul Menkul Kıymetler Borsası'nda işlem gören sanayi ve hizmet sektöründeki 173 işletmenin 2004 - 2006 yılları finansal göstergelerini kullanarak veri madenciliği süreci karar ağacı tekniğini uygulamışlardır. Çalışmanın sonucunda hizmet ve sanayi sektörlerindeki firmaların birbirlerinden ayrılmasındaki

en önemli etmenler; stok devir hızı, ekonomik rantabilite, işletme sermayesinin net satışlara oranı olarak belirlenmiştir (Albayrak ve Yılmaz, 2009, 31-52).

Kılıç çalışmasında, İstanbul Menkul Kıymetler Borsası'nda işlem gören 137 imalat firmasının gelir tabloları ve bilançoları ile finansal başarısızlıklarını ölçmeyi amaçlamıştır. Araştırmada, sınıflama yöntemlerinden olan karar ağacı yöntemi ve sinir ağları teknikleri kullanılmıştır (Kılıç, 2011, 1-84).

Gemici araştırmasında, İstanbul Menkul Kıymetler Borsası'nda işlem gören 10 hissenin değerlerindeki değişmelere bakarak aralarında birliktelik olup olmadığını araştırmıştır. Bu çalışmada veri madenciliğinde kullanılan Apriori algoritması kullanılmış ve sonucunda belirli birliktelik kuralları elde edilmiştir (Gemici, 2012, 1-86).

İlarslan çalışmasında, İstanbul Menkul Kıymetler Borsası'nda işlem gören 10 bankanın hisse senetlerinin günlük kapanış değerlerini kullanarak Markov Zincirleri ile gelecek fiyat hareketlerini tahmin etmiştir. Uygulama sonucunda Türkiye Sınai Katılım Bankası dışındaki diğer dokuz bankanın bir sonraki gün fiyat hareketi doğru bulunmuştur (İlarslan, 2014, 6185-6198).

3. VERİ MADENCİLİĞİ

Veri madenciliği, toplanan verilerden anlamlı bilgiler elde etmek için verilerde saklı olan bir takım örüntüleri ve eğilimleri tespit etmek isteyen, değişkenler arasındaki ilişkileri ortaya çıkaran ve böylece karar vermeyi kolaylaştıran bir süreç olarak tanımlanabilir (Seyrek ve Ata, 2010, 71).

Veri madenciliği yöntemleri tahmin edici ve tanımlayıcı modeller olarak ikiye ayrılmaktadır. Tahmin edici modeller, sonuçları bilinen verilerden yararlanılarak bir model geliştirilmesi ve kurulan bu model kullanılarak sonuçları bilinmeyen veri kümelerinin tahmin edilmesidir. Tahmin edici modeller içerisinde sınıflama ve regresyon yöntemlerini barındırmaktadır. Tanımlayıcı modeller ise, mevcut verilerdeki örüntüleri bularak karar vermede rehberlik etmektedir. Tanımlayıcı modellere örnek olarak, birliktelik kuralları ve ilişki analiz ile kümeleme yöntemleri sayılabilir (Özekes, 2012, 67 - 68).

3.1. Kümeleme Yöntemleri

Kümeleme, verilerin benzer nesnelereki gruplara bölünmesini ifade eder. Küme olarak adlandırılan her grup, kendi aralarında benzer olan ve diğer gruplara kıyasla benzer olmayan nesnelere oluşur. Küme içindeki herhangi iki nesne arasındaki mesafe, küme içindeki ve küme dışındaki herhangi iki nesne arasındaki mesafeden daha azdır. Kümeleme yöntemi, veri noktalarının (ya da nesnelere) bir koleksiyonunun, özneliklerinin değerlerine dayanarak homojen kümelere ayrılarak içsel yapısını bulmayı amaçlayan tanımlayıcı bir veri analizi görevidir. (Manikandan vd., 2016, 311).

Kümeleme modelleri, sosyal bilimler, biyoloji, astronomi, istatistik, görüntü tanıma, dijital işleme, bilgi ve pazarlama gibi çeşitli disiplinlerde uzun süredir kullanılmaktadır (Altuntaş, 2006, 6).

Veriler kümelendirken belirli bir sistematik hareket etmektedir. Veri madenciliği kümeleme analizinde izlenen sistemli aşamalar bulunmaktadır. Bu aşamalar şu şekilde ifade edilir (Hasanlı, 2014, 24):

- Birimler arası var olan benzerliklerin belirlenmesi için kullanılacak olan ölçütlerin ve değişkenlerin saptanması,
- Birimler arasında var olan benzerlikler saptandıktan sonra birimlerin kümeleneceği,
- Meydana gelen kümelere uygun olup olmadığının belirlenmesi,
- Kümelere uygun olarak oluşturulduğu varsayımı altında istatistik geçerliliğinin ortaya konması şeklinde sayılabilir.

Kümeleme yönteminde birçok algoritma bulunmaktadır. Kümeleme yönteminde kullanılan algoritmalar şu şekilde sıralanabilir:

- Bölümlemeli Yöntemler
- Yoğunluk Tabanlı Algoritmalar
- Hiyerarşik Yöntemler
- Izgara (Grid) Temelli Algoritmalar

3.1.1. Bölümlemeli Yöntemler

Bölümlemeli yöntemler, hiyerarşik olmayan kümeleme yöntemi olarak da bilinmektedir. Bölümlemeli yöntemler hiyerarşik yöntemlere göre daha büyük veri setlerine uygulanabilmektedir (Pektaş, 2013, 194).

Bölümlemeli yöntemler, “n” adet noktadan meydana gelen veritabanını giriş parametresi olarak karar verilen “k” adet bölüme (kümeye ($k \leq n$)) ayırmayı amaçlar. Veri tabanındaki her eleman belirlenen farklılıklarına göre k adet kümelerin birine dahil olur. Bölümlemeli yöntemlerin içinde K-Medoids, K-Means ve Clara-Clarans gibi algoritmalar yer almaktadır (Gülce, 2010, 22).

Bölümlemeli yöntemler hiyerarşik yöntemlere göre daha hızlı sonuç verir. Çünkü hiyerarşik yöntemlerdeki gibi benzerlik veya mesafe matrisi kullanmazlar. Bu sebeple büyük veri tabanlarının kümelenebilmesinde bölümlemeli yöntemler daha uygundur. Bununla birlikte verilen kriterlere uygun birden fazla sonuç çıkabilir ve en doğru sonucun bulunması zorlaşır. Böyle bir durumda, verilerin yerlerinin değiştirilip, dağıtılması ve tekrar sıralanması, algoritmanın tekrar koşuturulması ve yeni sonuçlarla kıyaslanması gerekir. Bu durumda da zaman açısından ek bir maliyet ortaya çıkar (Dunham, 2003, 8).

3.1.2. Yoğunluk Tabanlı Algoritmalar

Amineh Amini ve Hadi Sabouhi, gerçek zamanlı veri akışı için DMM akışı olarak da adlandırılan bir kümeleme algoritması sunmuşlardır. Bu algoritma, gelişen veri akışı için kullanılan yoğunluk tabanlı kümeleme algoritmasıdır. Daha küçük yarıçaplı mikro kümelenebilmeye benzeyen mini mikro kümelenebilmeyi tanıtmaktadır ve zaman komplikasyonu korurken kümelenebilmeye kalitesini arttıran doğru kümeyi tanımlamak için mesafe yöntemini kullanmaktadır. Bununla birlikte gerçek verilerden gelen gürültü için bir strateji sunmuşlardır ve bu deneyler gerçek ve karma veri seti üzerinde yürütülmüştür (Kazemi, 2017, 3).

Dağılmış veri tabanlarını yalnızca uzaklığı temel alan bölümleme algoritmaları açıklamak zordur. Çünkü uç noktalarda yer alan hiçbir kümeyle dahil olmayan veri tabanlarını bölümlemeli algoritmalarla kümelemek doğru sonuçlar ortaya çıkarmayabilir. Böyle durumlarda yoğunluk oluşturan verilerin kümelenebilmesine dayanan yoğunluğa dayalı algoritmalar kullanılabilir. Bu algoritmalar, DBSCAN, OPTICS ve DENCLUE algoritmaları örnek olarak gösterilebilir (Tapkan vd., 2011, 251).

3.1.3. Hiyerarşik Yöntemler

Hiyerarşik kümeleme yöntemleri, verileri ağaç yapısı şeklinde kümelemektedir. Bu metot verileri aşağıdan yukarıya veya yukarıdan aşağıya olmak üzere birleştirici ve ayrıştırıcı kümeler olarak iki gruba ayırmıştır. Hiyerarşik kümeleme yöntemleri küme sayısını belirtmekte kullanılan “k” değerine gerek duymaz ama ağaç yapısının ne zaman duracağını belirten eşik değerine gerek duyar (Erdoğan, 2004, 40).

Hiyerarşik kümeleme algoritmaları, her veri noktasını kendi başına bir kümeyle atayarak çalışmaya başlamaktadır. Daha sonrasında benzerlik ölçüsünü kullanarak, en benzer iki kümeyle tanımlar. Bunlar birlikte birleştirilir ve bu süreç gerekli sayıda kümeler oluşturuluncaya veya tüm kümeler bir kümenin parçası haline gelene kadar tekrarlanmaya devam eder. Bu sebeple tüm kümelenebilmemiş öğeler arasında bir ağaç hiyerarşisi bulunmaktadır.

Hiyerarşik algoritmalar, aglomeratif veya bölücü olur. Bu yöntem, kümeleri oluşturmak için aşağıdan yukarıya bir yaklaşım kullanırsa, doğada aglomeratif olurlar. Başka deyişle, ayrı ayrı tek öge kümeleriyle başlayarak, her adımda iki kümenin birleştirilmesiyle daha büyük kümelerin oluşturulmasıdır. Eğer başlangıçta, gerekli kümeleri üretmek için her adımda ayrıştırılan tek, geniş, bütün-elementi kapsayan bir kümenin var olması durumunda bölücü devreye girmektedir ve yukarıdan aşağıya ayrılmaktadırlar (Saxena vd., 2015, 31).

3.1.4. Izgara (Grid) Temelli Algoritmalar

Izgara temelli algoritma, çok fazla miktarda alan gerektiren verilere kümeleme işlemi yapılabilmesi için numaralandırılmış çizgilerden meydana gelen hücresel yapılar kullanılarak kümeleme yapabilen yöntemdir (Silahtaroglu, 2013: 200).

Bu yöntem, ızgara kümeleme yapmak için veri noktaları yerine değer alanı ile ilgilenir. Bu şekilde ızgara kümeleme yöntemi öncelikle bir ızgara yapısı oluşturur, daha sonra hücre yoğunluğunu hesaplar ve sonrasında küme merkezlerini tanımlar. Izgara temelli kümelemenin avantajları, özellikle çok büyük veri kümelerini kümelemek için hesaplama karmaşıklığını azaltmaktır (Bano ve Khan, 2018, 135).

Bu yöntemin diğer avantajı, hızlı tamamlanması ve nesnelerin sayısından bağımsız olmasıdır. Bu metotlara örnek olarak Sting, Clique ve WaveCluster gösterilebilir (Işık, 2006, 9).

3.2. Beklenti Maksimizasyonu Algoritması

Bu çalışmada Beklenti maksimizasyonu (expectation maximization - EM) algoritması kullanılacağı için sadece bu algoritmanın detaylarına yer verilmiştir.

Beklenti maksimizasyonu, interaktif şekilde tekrar eden iki aşama olan B (beklenti) ve M (maksimizasyonu) üzerinden atama işlemi yapan bir algoritmadır (Sayın vd., 2017, 493).

Birinci aşama olan beklenti aşamasında, gözlenen değerler kullanılarak bir dizi regresyon denklemi elde edilir. Kayıp verilere başlangıç değerleri atanır. İkinci aşama olan maksimizasyon aşamasında, atama yapılan değerler kullanılarak yeni değerlere atama yapılır. Bu adımlar birçok kez tekrar edilerek B (Beklenti) ve M (Maksimizasyon) basamaklarından her adımda yeni değerler hesaplanmış olur. Ardışık M basamakları arasında değerler artık değişme göstermediğinde maksimum olasılık tahmini gerçekleşmiş olur. Beklenti maksimizasyonu, kullanılarak elde edilen standart hataların gerçek standart hatalarla tutarlı olmaması bu yöntemin dezavantajı olarak gösterilmektedir (Işıkoğlu, 2017, 13 – 14).

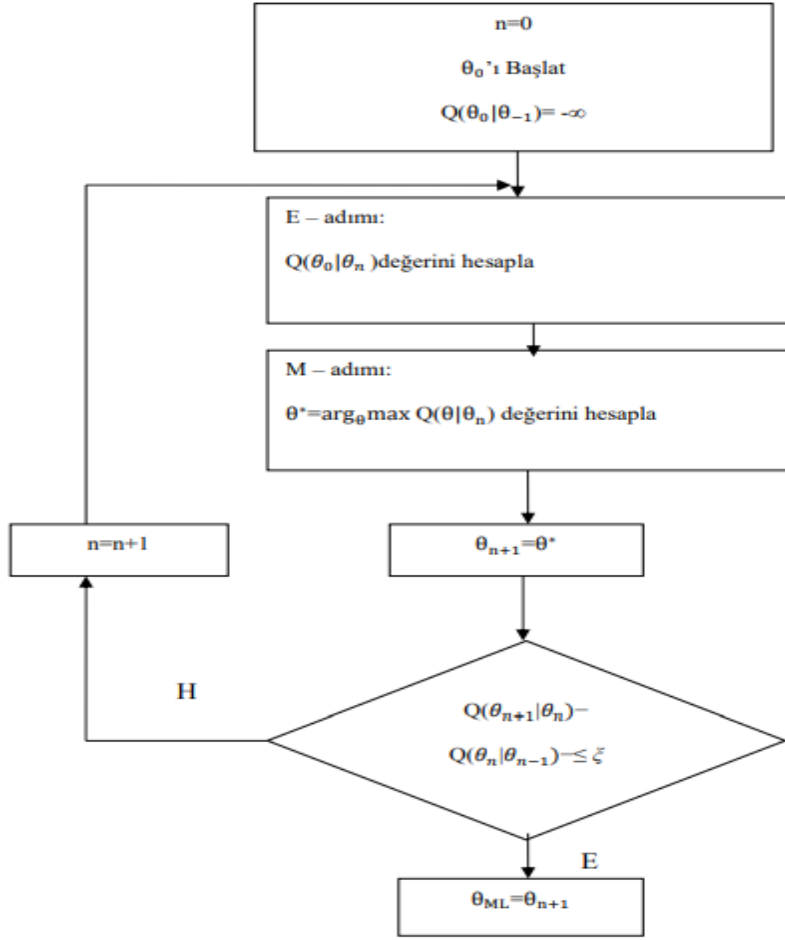
Sunulan bir $L(\theta; x, z)$ fonksiyonunda, θ : parametre vektörü, x : gözlem verisi ve z : eksik verileri temsil etmekte kullanılmaktadır. E-adımında, olasılık fonksiyonunun beklenen değeri denklem (1) kullanılarak hesaplanır. M adımında ise olasılık fonksiyonunu maksimum yapan parametreler denklem (2) yardımıyla hesaplanır.

$$Q(\theta_0|\theta_n) = E_{Z|x,\theta_n}[\log L(\theta; x, z)] \quad (1)$$

$$\theta^* = \arg_{\theta} \max Q(\theta|\theta_n) \quad (2)$$

Ortalama ve kovaryans matrisinin tahminleri beklenti maksimizasyonu yönteminde üç adımda gerçekleşmektedir. Birinci adımda, ortalama ve kovaryans matrisinin tahminleri kullanılarak eksik değerleri içeren verilerin regresyon parametreleri hesaplanır. İkinci adımda, eksik değerler, hesaplanan parametreler kullanılarak tamamlanmaktadır. Daha sonra ortalama ve kovaryans matrisi tamamlanmış veri seti kullanılarak yeniden tahmin gerçekleştirilir. Bu işlem yakınsama sağlanıncaya kadar devam eder ve sonucunda eksik veriler için en iyi tahmini değerler bulunmuş olur (Şahin, 2012, 19 - 20).

Şekil 1. Beklenti Maksimizasyonu Algoritmasının Akış Diyagramı



Kaynak: (Şahin, 2012, 20)

4. UYGULAMA

Bu bölümde çalışmanın amacı, kısıtları, yöntemi, modellemesi ve bulgularından bahsedilecektir.

4.1. Çalışmanın Amacı

Bu çalışmada, Borsa İstanbul'da 02/01/2017 – 30/06/2017 tarihleri arasında faaliyet gösteren 134 işletmenin hisse senetlerinin günlük artış, azalış ve sabit kalma durumları göz önüne alınarak bir veri seti oluşturulmuştur. Bu veri setinin, anlamlı bir kümeleme oluşturması ve oluşan kümelerin sektör ve işletme açısından analiz edilmesi amaçlanmıştır.

4.2. Çalışmanın Kısıtları

Çalışmada, Borsa İstanbul'da işlem gören tüm sektörlerdeki işletmelerin dikkate alınması hedeflenmiştir. Ancak, çalışmanın yapıldığı dönemde hisse senetlerinin bazılarının borsa tahtası kapanmış ya da farklı dönemde işleme açmış olması, bazı hisse senetlerinin de verilerinin eksik olmasından dolayı verilerine tam olarak ulaşılabilen on sektörde bulunan 134 işletme veri setine dahil edilmiştir. Uygulama aşaması için belirlenen on sektör; BIST banka, BIST gıda içecek, BIST inşaat, BIST elektrik, BIST ulaştırma, BIST turizm, BIST teknoloji, BIST tekstil deri, BIST ticaret ve BIST metal eşya makinedir.

Tablo 1'de çalışmada yararlanılan hisse senetlerinin sektörlere göre frekans dağılımı görülmektedir. Tablo 1'de görüldüğü üzere en çok işletme metal eşya makine ve gıda içecek sektöründe faaliyet göstermektedir. Daha sonra bu sektörleri içinde barındırdığı işletme sayısı bakımından ticaret, teknoloji, tekstil deri ve banka takip

etmektedir. İçerisinde en az işletme bulunan sektörler ise inşaat, ulaştırma, turizm ve elektrik olarak sayılabilmektedir.

Tablo 1. Hisse Senetlerinin Sektörlere Göre Frekans Dağılımı

Sektörler	f _i
Banka	13
Gıda İçecek	22
İnşaat	8
Elektrik	6
Ulaştırma	7
Turizm	7
Teknoloji	14
Tekstil Deri	13
Ticaret	18
Metal Eşya Makine	26
Toplam	134

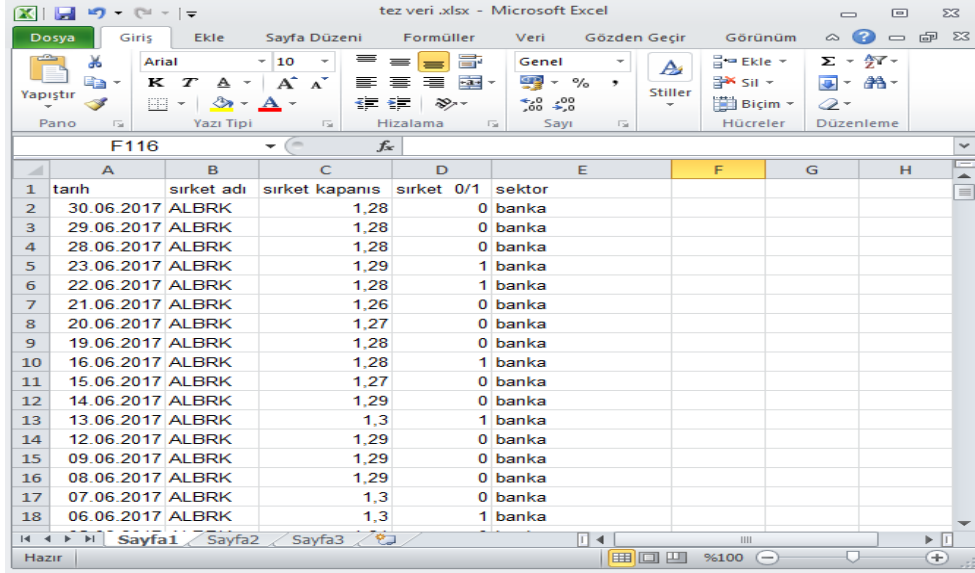
4.3. Çalışmanın Yöntemi

Uygulamada kullanılan hisse senetlerinin günlük son kapanış fiyatları, 02/01/2017 - 30/06/2017 dönemi resmi tatil ve hafta sonları çıkarılarak her işletme için 125 günlük veri seti olarak hazırlanmıştır.

Çalışmada, 134 işletmenin artış gösterdiği günlere göre kümelmesi istenildiği için hisse senetleri fiyatlarının “artış” gösterdiği günler “1”, “azalış” ve “sabit durumda” olduğu günler “0” olarak kukla değişken (dummy) kullanılarak kodlanmıştır. Hisse senetlerinin artış, azalış veya sabit durumda bulunmasının tek tek kontrolü zaman alacağından dolayı kukla değişken (dummy) kullanılarak, artış gösteren günlere ilişkin veriler Microsoft Excel (2010) programında sıralanmıştır.

Örneğin, Şekil 2 incelendiğinde ALBRK şirketi 21/06/2017 tarihinde günü 1,26 ile tamamlamıştır. 22/06/2017 tarihinde ise günü “1.28” ile tamamlayıp artış gösterdiği için kukla değişkeni “1” olarak belirlenmiştir. ALBRK şirketi 23/06/2017 tarihi günü 1,29 ile tamamlayıp 28/06/2017 tarihinde ise 1,28’e düştüğü için kukla değişkeni “0” olarak belirlenmiştir. Son olarak ALBRK şirketi 29/06/2017 tarihinde günü 1,28 ile tamamlayıp 30/06/2017 tarihinde de günü 1,28 olarak tamamladığı için kukla değişkeni “0” olarak belirlenmiştir. Bu aşama 134 işletmenin hisse senetlerine sırayla uygulanmıştır.

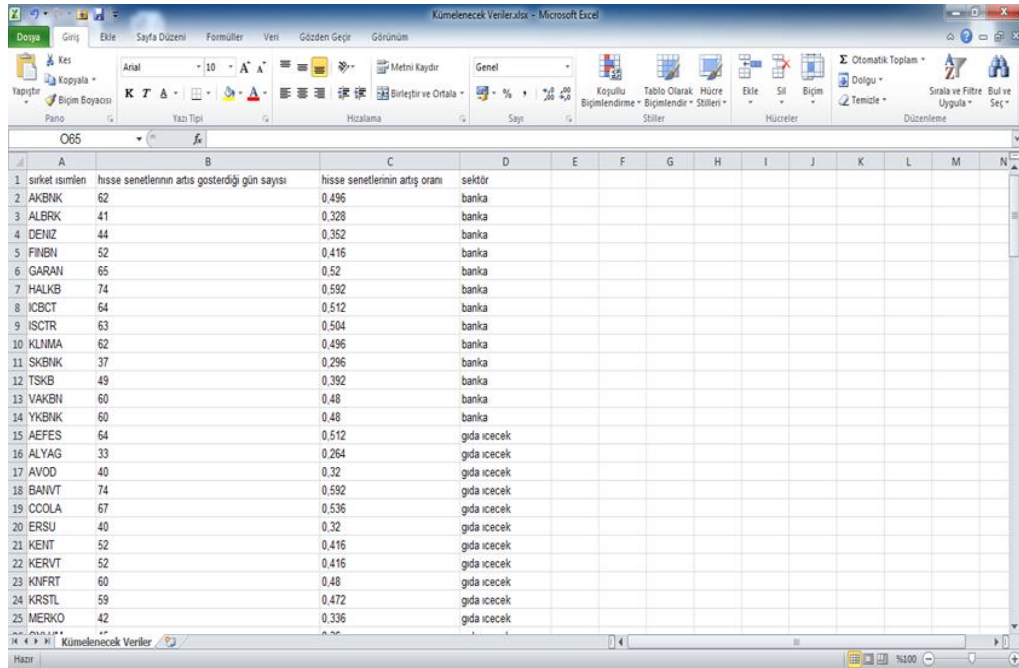
Şekil 2. Uygulama İçin Düzenlenecek Veriler (Kesit)



	A	B	C	D	E	F	G	H
1	tarih	şirket adı	şirket kapanış	şirket 0/1	sektor			
2	30.06.2017	ALBRK	1,28	0	banka			
3	29.06.2017	ALBRK	1,28	0	banka			
4	28.06.2017	ALBRK	1,28	0	banka			
5	23.06.2017	ALBRK	1,29	1	banka			
6	22.06.2017	ALBRK	1,28	1	banka			
7	21.06.2017	ALBRK	1,26	0	banka			
8	20.06.2017	ALBRK	1,27	0	banka			
9	19.06.2017	ALBRK	1,28	0	banka			
10	16.06.2017	ALBRK	1,28	1	banka			
11	15.06.2017	ALBRK	1,27	0	banka			
12	14.06.2017	ALBRK	1,29	0	banka			
13	13.06.2017	ALBRK	1,3	1	banka			
14	12.06.2017	ALBRK	1,29	0	banka			
15	09.06.2017	ALBRK	1,29	0	banka			
16	08.06.2017	ALBRK	1,29	0	banka			
17	07.06.2017	ALBRK	1,3	0	banka			
18	06.06.2017	ALBRK	1,3	1	banka			

Şekil 3'te kukla değişken kullanılarak oluşturulan tabloda her bir hisse senedinin 125 günlük durumu gözlenebilmektedir. Oluşturulan tabloda her bir hisse senedinin 125 günde kaç gün artış gösterdiği saptanmıştır. Ayrıca hisse senetlerinin artış oranları da hesaplanarak hisse senetlerinin artış gösterdiği gün sayısı ve hisse senetlerinin artış oranı uygulamada değişken olarak kullanılmıştır.

Şekil 3. Hisse Senetlerinin Artış Gösterdiği Gün Sayısı ve Artış Oranı (Kesit)



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	şirket ismi	hisse senetlerinin artış gösterdiği gün sayısı	hisse senetlerinin artış oranı	sektör										
2	AKBNK	62	0.496	banka										
3	ALBRK	41	0.328	banka										
4	DENIZ	44	0.352	banka										
5	FINBN	52	0.416	banka										
6	GARAN	65	0.52	banka										
7	HALKB	74	0.592	banka										
8	ICBCT	64	0.512	banka										
9	ISCTR	63	0.504	banka										
10	KLNMA	62	0.496	banka										
11	SKBNK	37	0.296	banka										
12	TSKB	49	0.392	banka										
13	VAKBN	60	0.48	banka										
14	YKBNK	60	0.48	banka										
15	AEFES	64	0.512	gıda içecek										
16	ALYAG	33	0.264	gıda içecek										
17	AVOD	40	0.32	gıda içecek										
18	BAIWT	74	0.592	gıda içecek										
19	COOLA	67	0.536	gıda içecek										
20	ERSU	40	0.32	gıda içecek										
21	KENT	52	0.416	gıda içecek										
22	KERVIT	52	0.416	gıda içecek										
23	KNFRT	60	0.48	gıda içecek										
24	KRSTL	59	0.472	gıda içecek										
25	MERKO	42	0.336	gıda içecek										

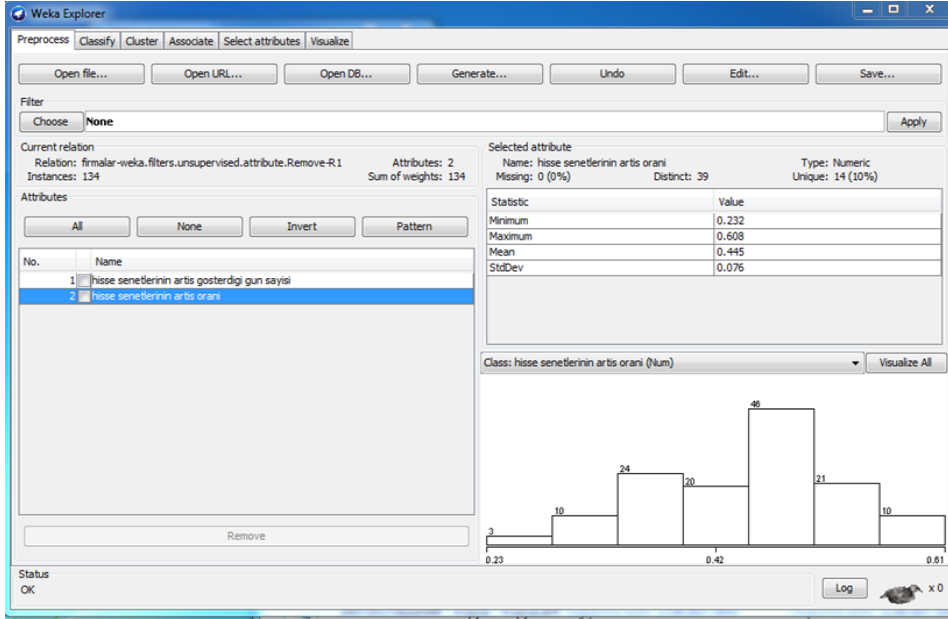
5. MODELLEME VE BULGULAR

Veri madenciliği tekniklerinden biri olan kümeleme yöntemini uygulayabilmek için WEKA programında bulunan algoritmalar tercih edilmiştir. WEKA programında kullanılan kümeleme yöntemlerinden biri olan “beklenti maksimizasyonu - expectation maximisation” algoritması kullanılmıştır. Bu algoritmada küme sayısı

kullanıcı tarafından belirlenmektedir. Çalışmada, dikkate alınan dönem içerisinde işletmelerin hisse senetleri yükselişi “düşük”, “orta düzeyde”, ve “yüksek” olarak değerlendirilmek istendiği için 3 kümeye ayrılmasına karar verilmiştir.

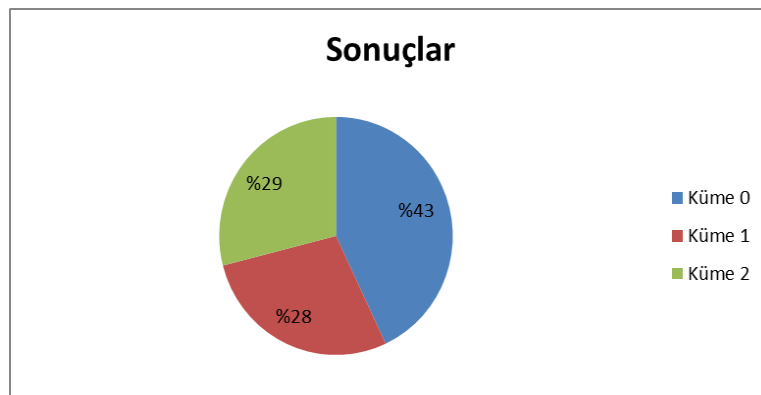
Verilerin WEKA programına yüklenmesinin ardından Şekil 4’te görüldüğü üzere temel istatistiksel analizler gerçekleştirilmiştir. Artış oranı değişkenine ait verilerin en düşük, en yüksek değerleriyle birlikte ortalama ve standart sapma gibi parametreleri ve frekans dağılımı da grafikte görülmektedir.

Şekil 4. Verilerin Düzenlendiği WEKA Arayüzü



Şekil 5’te ise WEKA programında oluşturulan modelin uygulamada elde edilen sonuçları daire grafiği üzerinde gösterilmiştir. Bu sonuçlara göre kümeleme yöntemi olan beklenti maksimizasyonu ile üç kümeye ayrılan 134 işletmenin; 57 tanesi (%43) “küme 0 - cluster 0”a, 38 tanesi (%28) “küme 1 - cluster 1”e, 39 tanesi (%29) “küme 2 - cluster 2”ye atanmıştır.

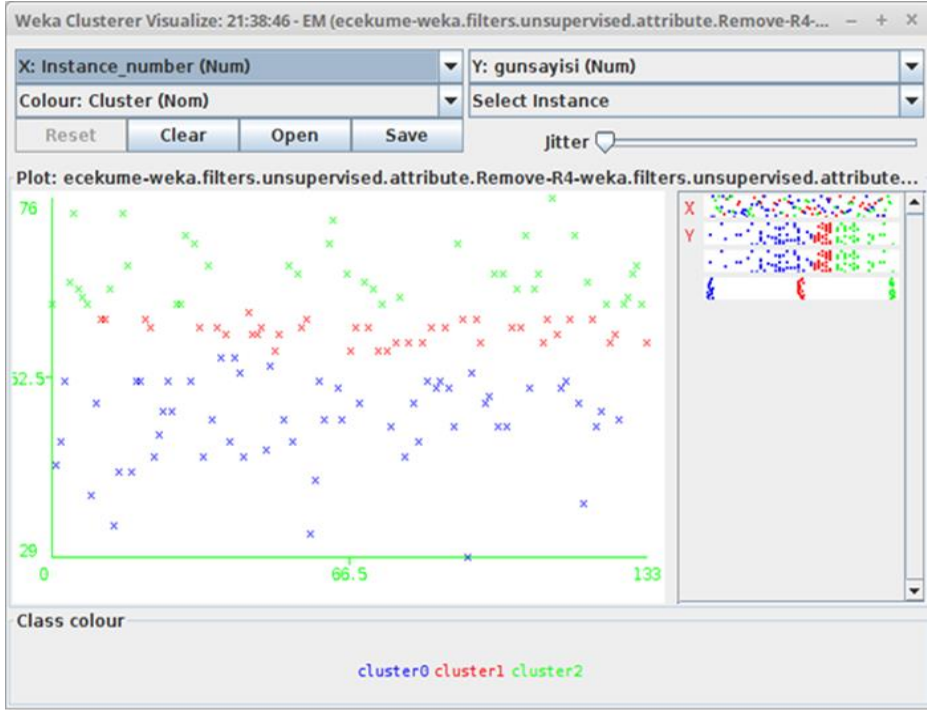
Şekil 5. Uygulama Sonuçlarının Daire Grafiğiyle Görselleştirilmiş Hali



Şekil 6’daki görselleştirme paneli incelendiğinde 3 kümenin değerleri birbirleriyle iç içe geçmemiştir. Dönem içinde hisse senetleri yükselişi düşük olan “küme 0”, diğer kümelere göre daha fazla işletmeyi

kapsamaktadır. “Küme 1”de yer alan işletmelerin değerleri birbirine daha yakındır. Fakat “küme 0” ve “küme 2” ye dahil olan işletmelerin değerleri daha dağınık bir dağılım göstermektedir.

Şekil 6. WEKA Sonuç Görselleştirme Paneli



Tablo 2’de uygulama sonuçlarına göre işletmelerin küme dağılımları görülmektedir.

Tablo 2. Şirketlerin Kümeleme Sonuçları

Küme 0	Küme 1	Küme 2
ALBRK (banka)	VAKBN (banka)	AKBNK (banka)
DENIZ (banka)	YKBNK (banka)	GARAN (banka)
FINBN (banka)	KNFRT (gıda içecek)	HALKB (banka)
SKBNK (banka)	KRSTL (gıda içecek)	ICBCT (banka)
TSKB (banka)	ULUUN (gıda içecek)	ISCTR (banka)
ALYAG (gıda içecek)	ENKAI (inşaat)	KLNMA (banka)
AVOD (gıda içecek)	ORGE (inşaat)	AEFES (gıda içecek)
ERSU (gıda içecek)	AKSEN (elektrik)	BANVT (gıda içecek)
KENT (gıda içecek)	AKSUE (elektrik)	CCOLA (gıda içecek)
KERVT (gıda içecek)	AYEN (elektrik)	PNSUT (gıda içecek)
MERKO (gıda içecek)	ODAS (elektrik)	TATGD (gıda içecek)
OYLUM (gıda içecek)	CLEBI (ulaştırma)	TBORG (gıda içecek)
PENGD (gıda içecek)	DOCO (ulaştırma)	ULKER (gıda içecek)
PETUN (gıda içecek)	AVTUR (turizm)	ANELE (inşaat)
PINSU (gıda içecek)	MAALT (turizm)	PGSUS (ulaştırma)
TUKAS (gıda içecek)	DESPC (teknoloji)	THYAO (ulaştırma)
VANGD (gıda içecek)	DGATE (teknoloji)	UTPYA (turizm)

EDIP (inşaat)	KAREL (teknoloji)	ALCTL (teknoloji)
KUYAS (inşaat)	LINK (teknoloji)	ASELS (teknoloji)
SANEL (inşaat)	NETAS (teknoloji)	INDES (teknoloji)
TURGG (inşaat)	ARSAN (tekstil deri)	KRONT (teknoloji)
YYAPI (inşaat)	BOSSA (tekstil deri)	LOGO (teknoloji)
AKENR (elektrik)	DERIM (tekstil deri)	ATEKS (tekstil deri)
ZOREN (elektrik)	KORDS (tekstil deri)	BIMAS (ticaret)
BEYAZ (ulaştırma)	YATAS (tekstil deri)	MGROS (ticaret)
GSDDE (ulaştırma)	BIZIM (ticaret)	PIMAS (ticaret)
RYSAS (ulaştırma)	DOAS (ticaret)	SELEC (ticaret)
MARTI (turizm)	INTEM (ticaret)	TKNSA (ticaret)
METUR (turizm)	SANKO (ticaret)	ALCAR (metal eşya makine)
NTTUR (turizm)	TGSAS (ticaret)	ARCLK (metal eşya makine)
TEKTU (turizm)	ASUZU (metal eşya makine)	BNTAS (metal eşya makine)
ARENA (teknoloji)	BFREN (metal eşya makine)	FROTO (metal eşya makine)
ARMDA (teknoloji)	DITAS (metal eşya makine)	JANTS (metal eşya makine)
ESCOM (teknoloji)	FMIZP (metal eşya makine)	OTKAR (metal eşya makine)
PKART (teknoloji)	KLMSN (metal eşya makine)	TMSN (metal eşya makine)
BLCYT (tekstil deri)	PARSNÂ (metal eşya makine)	TOASO (metal eşya makine)
BRMEN (tekstil deri)	PRKAB (metal eşya makine)	TTRAK (metal eşya makine)
DAGI (tekstil deri)	VESTL (metal eşya makine)	ULUSE (metal eşya makine)
HATEK (tekstil deri)		VESBE (metal eşya makine)
MNDRS (tekstil deri)		
SKTAS (tekstil deri)		
YUNSA (tekstil deri)		
ADESE (ticaret)		
BMEKS (ticaret)		
CRFSA (ticaret)		
KIPA (ticaret)		
MEPET (ticaret)		
MIPAZ (ticaret)		
PSDTC (ticaret)		
VAKKO (ticaret)		
EGEEN (metal eşya makine)		
EMKEL (metal eşya makine)		
GEREL (metal eşya makine)		
IHEVA (metal eşya makine)		
KARSN (metal eşya makine)		
MAKTK (metal eşya makine)		
SILVR (metal eşya makine)		

Küme dağılımlarına göre; belirlenen dönem içerisinde incelenen işletmelerin %43'ünün dönem içinde hisse senetleri yükselişi “düşük” olan kümede yer almış olması işletmelerin büyük bir bölümünün istikrarlı bir

artış göstermediği anlamını taşımaktadır. İşletmelerin %29'unun hisse senetleri yükselişi "yüksek" kümede yer almış olması yatırımcılarına daha yüksek kazanç sağlama imkanı sunmuştur. Ancak, farklı bir örneklem alındığında işletmelerin kümedeki yerleri değişiklik gösterebilir. Bu durumun ortaya çıkmasındaki en önemli nedenlerden biri mevsimsel etkidir. Örneğin, turizm veya ulaştırma sektörlerinde yaz aylarında diğer dönemlere göre, faaliyetlerinde artış fazla olacağı için bu sektörlerdeki işletmeler farklı dönemlerde değişik kümede yer alabilirler.

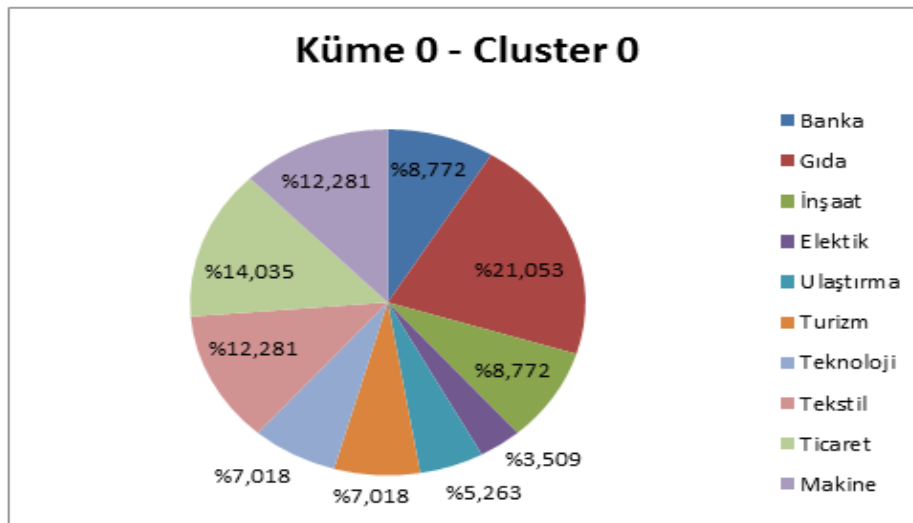
İşletmeler bağımsız olarak kümelenmesine rağmen sektörler göre incelemek de mümkündür. Çalışmada 10 farklı sektörde bulunan 134 işletmenin frekans dağılımlarına Tablo 3'te yer verilmiştir.

Tablo 3. Kümeleme Sonuçlarının Sektörlere Göre Dağılımı

	Küme 0	Küme 1	Küme 2
Banka	5	2	6
Gıda içecek	12	3	7
İnşaat	5	2	1
Elektik	2	4	0
Ulaştırma	3	2	2
Turizm	4	2	1
Teknoloji	4	5	5
Tekstil deri	7	5	1
Ticaret	8	5	5
Metal eşya makine	7	8	11
TOPLAM	57	38	39

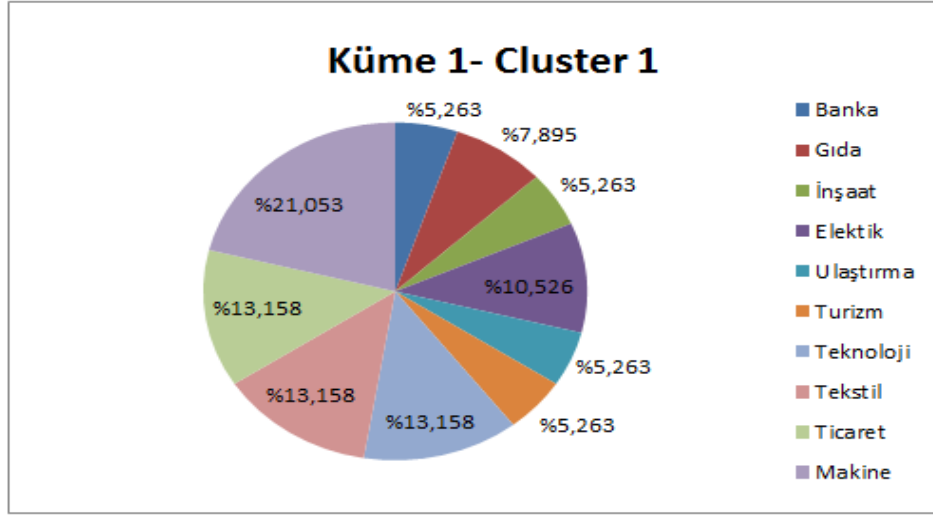
Şekil 7'de ise, her bir kümenin sektörler göre dağılımı görülmektedir. Bu sonuçlara göre, küme 0' da yer alan sektörlerden en fazla yere sahip olan %21,053 ile "gıda sektörü" ve en az yere sahip olan %3,509 ile "elektrik sektörü" dür. Hisse senedi yükselişi "düşük" olan küme 0'da yüksek yüzdeye sahip olan sektörler diğer sektörler göre daha istikrarsız olarak görülmektedir.

Şekil 7. Sektörlere Göre Küme 0 Daire Grafiği



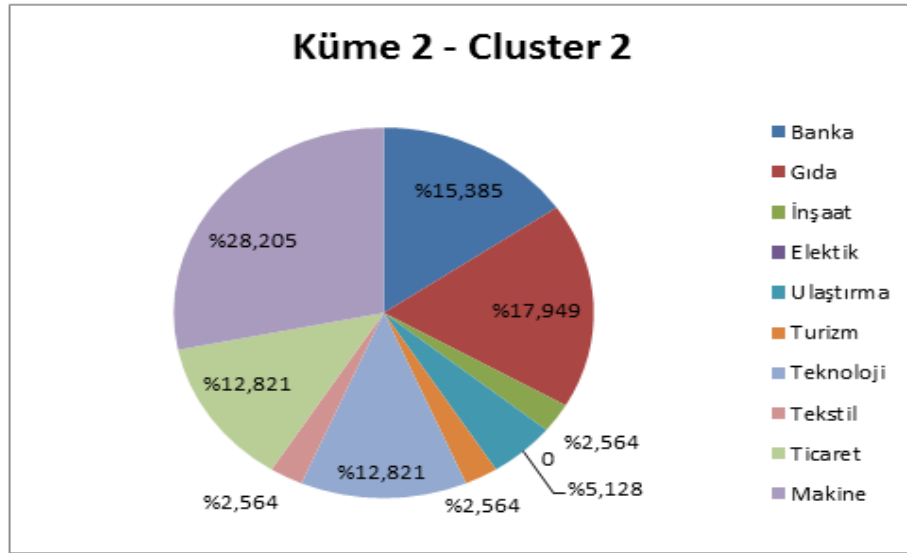
Şekil 8’de görüldüğü üzere, hisse senedi yükselişi “orta düzeyde” olan kümede makine sektörü %21,053 ile en yüksek orana sahiptir. Turizm, ulaştırma, inşaat, banka sektörü %5,263 ile en düşük oranlara sahiptir.

Şekil 8. Sektörlere Göre Küme 1 Daire Grafiği



Şekil 9 incelendiğinde, hisse senedi yükselişi “yüksek ” olan küme 2’de makine sektörü %28.205 ile en yüksek paya sahiptir. Bu kümede elektrik sektöründen hiç işletme bulunmadığı için yüzdeler diliminde yer almamıştır.

Şekil 9. Sektörlere Göre Küme 2 Daire Grafiği



Tablo 4’te kümelere yer alan işletmelerinin sektörlere göre dağılımı ve yüzdeler olarak hesaplanmasına yer verilmiştir. Üç kümenin yüzdeler dağılımı, sektördeki işletmelerin toplamının, kümelere de yer alan işletme sayısına oranıyla (*100) bulunmaktadır.

Tablo 4. Sektörlerin Bulunduğu Kümelerin Yüzdelerik Dağılımı (%)

	Küme 0	Küme 1	Küme 2	Toplam	Küme 0	Küme 1	Küme 2
Banka	5	2	6	13	38,462	15,385	46,154
Gıda içecek	12	3	7	22	54,545	13,636	31,818
İnşaat	5	2	1	8	62,500	25	12,5
Elektik	2	4	0	6	33,333	66,667	0
Ulaştırma	3	2	2	7	42,857	28,571	28,571
Turizm	4	2	1	7	57,143	28,571	14,286
Teknoloji	4	5	5	14	28,571	35,714	35,714
Tekstil deri	7	5	1	13	53,846	38,462	7,692
Ticaret	8	5	5	18	44,444	27,778	27,778
Metal eşya makine	7	8	11	26	26,923	30,769	42,308
Toplam	57	38	39	134	42,537	28,358	29,104

Şekil 10’da yer alan bilgilere göre banka sektöründeki işletmelerin %46,154’ü, metal eşya makine sektöründeki işletmelerin %42,308’i ve teknoloji sektöründeki işletmelerin %35,714’ü incelenen dönem içinde hisse senetleri yükselişi “yüksek” kümede yer almıştır. Seçili dönem içerisinde bu sektörlerdeki işletmelerin diğer sektörlerdekilere göre daha istikrarlı artış gösterdiği söylenebilir.

Elektrik sektörünün %66,667’si ve teknoloji sektörünün %35,714’ü hisse senetleri yükselişi “orta düzeyde” olan kümede yer alırken gıda sektörünün %54,545’i, inşaat sektörünün %62,5’si, ulaştırma sektörünün %42,857’si, turizm sektörünün %57,143’ü, tekstil sektörünün %53,846’sı, ticaret sektörünün %44,444’ü hisse senetleri yükselişi “düşük” kümede yer almıştır. Bu bilgilerden yola çıkarak seçili dönem içerisinde aynı sektördeki işletmelerin, farklı kümelerdeki dağılımlarından hareketle artış, azalış ve sabit kalma durumlarında işletmelerin sektör olarak birlikte hareket etmediği yorumu yapılabilmektedir.

6. SONUÇ VE ÖNERİLER

Teknolojinin gelişmesiyle birlikte veri tabanlarında saklanan verilerden anlamlı bilgiler çıkartılmaya başlanmıştır. Elde bulunan bu bilgilerin doğru şekilde kullanılması günlük yaşamda da karar almayı kolaylaştırmaktadır. Günümüzde hisse senedi almak isteyen yatırımcıların, hisse senetleriyle ilgili elde ettiği bilgileri doğru şekilde kullanması ve yorumlaması gelecekte kâr elde etme imkanı sağlayacaktır.

Bu çalışmada, Borsa İstanbul’da 02/01/2017 - 30/06/2017 tarihleri arasında işlem gören hisse senetlerinin günlük fiyat artış, azalış ve sabit kalma durumları göz önüne alınarak bir veri seti oluşturulmuştur. Bu veri setinin, anlamlı bir kümeleme oluşturması ve oluşan kümelerin sektör ve işletme açısından analiz edilmesi amaçlanmıştır.

Çalışmanın uygulama aşamasında, 134 işletmenin altı aylık hisse senedi günlük son kapanış fiyatları kullanılarak, günlük “artış”, “azalış” ve “sabit kalma” durumları kukla değişken kullanılarak düzenlenmiştir. Kukla değişken kullanılarak oluşturulan veri setinde işletmelerin 125 günde kaç gün artış gösterdiği saptanmış ve daha sonra artış oranları da hesaplanarak değişken olarak kullanılmıştır.

Çalışmada, WEKA programı tercih edilmiş ve bu programda yer alan kümeleme yöntemi algoritmalarından biri olan beklenti maksimizasyonu uygulanmıştır. Çalışmada, dikkate alınan dönem içerisinde işletmelerin, hisse senetleri yükselişi “düşük”, “orta düzeyde” ve “yüksek” olarak değerlendirilmesi istendiği için üç kümeye ayrılması uygun bulunmuştur.

Çalışma sonucunda elde edilen bulgulara göre 134 işletmenin 57 tanesi yani (%43) incelenen dönem içinde hisse senetleri yükselişi “düşük” olan kümeye, 38 tanesi yani (%28), “orta düzeyde” olan kümeye ve 39 tanesi yani (%29) “yüksek” olan kümeye dahil olmuştur. Hisse senetleri yükselişi “yüksek” olan kümede yer alan 39 işletmenin diğer kümelerde yer alan işletmelere göre daha istikrarlı artış gösterdiği görülmüştür.

Çalışmada, banka (%46,154), metal eşya makine (%42,308), teknoloji (%35,714) sektörlerinde yer alan işletmelerin büyük bir kısmı dönem içinde hisse senedi fiyat artışı “yüksek” olan kümede yer alarak diğer sektörlerdeki işletmelere göre daha istikrarlı artış göstermiştir.

Aynı sektörde bulunan işletmelerin, farklı kümelerdeki dağılımlarından hareketle artış, azalış ve sabit kalma durumlarında işletmelerin bağımsız olarak hareket ettiği, sektör olarak birlikte hareket etmediği sonucuna varılmıştır.

Gelecekte bu konuyla ilgili çalışacak olan araştırmacılar, iki farklı dönemde işletmeleri kümeleyip, işletmelerin farklı dönemdeki küme değişimlerini inceleyerek yorumlayabilir. Bunun dışında, daha fazla dönemi kapsayan bir inceleme yapılarak yükselişte, orta düzeyde ve düşüşte olan hisse senetleri analiz edilebilir.

KAYNAKÇA

- Aktürk, H. - Korukoğlu, S., (2008), Veri Madenciliği Teknolojisini Kullanarak Fiyat Değişimlerinde Paralellik Gösteren Hisse Senetlerinin Bulunması ve Risk Azaltılması, *Akademik Bilişim 2008 Konferansı Çanakkale Onsekiz Mart Üniversitesi, Çanakkale*, 113-119.
- Aktürk, H., (2008), *Borsa ve Döviz Verileri Üzerinde Veri Madenciliği Teknolojisini Kullanarak Zarar Riskini Azaltan Bir Uygulama Geliştirimi*, Ege Üniversitesi, Fen Bilimler Enstitüsü, Yüksek Lisans Tezi, İzmir.
- Albayrak, A. S. - Yılmaz, Ş .K., (2009), Veri Madenciliği: Karar Ağacı Algoritmaları ve IMKB Verileri Üzerine Bir Uygulama, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1), 31-52.
- Altuntaş, S., (2006), *Ağırlıklı İlişkilendirme Kurallarına Dayalı Veri Madenciliği Algoritmalarını Kullanarak Tesis Yerleşimi: Simülasyon ile Analiz*, Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
- Bano, S., Kahn, M. N. A., (2018), A Survey of Data Clustering Method, *International Journal of Advanced Science and Technology*, 113, 133 – 142.
- Erdoğan, Ş. Z., (2004), *Veri Madenciliği ve Veri Madenciliğinde Kullanılan K-MEANS Algoritmasının Öğrenci Veri Tabanında Uygulanması*, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İstanbul.
- Gemici, B., (2012), *Veri Madenciliği ve Bir Uygulaması*, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, İzmir.
- Gülce, A. C., (2010), *Veri Madenciliğinde Apriori Algoritması ve Apriori Algoritmasının Farklı Veri Kümelerinde Uygulanması*, Trakya Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Edirne.
- Hasanlı, H., (2014), *Çok Boyutlu Veritabanlarında Veri Madenciliği Yöntemleri Kullanılarak Bilgi Keşfi*, Ege Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İzmir.
- Işık, M., (2006), *Bölünmeli Kümeleme Yöntemleri ile Veri Madenciliği Uygulamaları*, Marmara Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.

- Işıkoğlu, M. A., (2017), *Kayıp Veri ile Baş Etme Yöntemlerinin Ölçme Değişmezliğine Etkisi Açısından Karşılaştırılması*, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Yüksek Lisans Tezi.
- İlarslan, K., (2014), *Hisse Senedi Fiyat Hareketlerinin Tahmin Edilmesinde Markov Zincirlerinin Kullanılması: İMKB 10 Bankacılık Endeksi İşletmeleri Üzerine Ampirik Bir Çalışma*, Yaşar Üniversitesi, 9(35), 6186-6198.
- Kazemi, U., (2017), Clustering Methods in Big Data, *Journal of Embedded Systems and Processing*, 2(3), 1 – 5.
- Kılıç, Y., (2011), *Finansal Başarısızlık Tahmininde Veri Madenciliğinin Kullanılması: İMKB’de Bir Uygulama*, Gaziantep Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Gaziantep.
- Küçükşille, E., (2009), *Veri Madenciliği Süreci Kullanılarak Portföy Performansının Değerlendirilmesi ve İMKB Hisse Senetleri Piyasasında Bir Uygulama*, Süleyman Demirel Üniversitesi, Sosyal Bilimler Enstitüsü, Doktora Tezi, Isparta.
- Mynet Finans, (30.06.2017), <http://finans.mynet.com/>
- Özekes, S., (2003), Veri Madenciliği Modelleri ve Uygulama Alanları, *İstanbul Ticaret Üniversitesi Dergisi*, 3(3), 65 – 68.
- Pektaş, a. O., (2013), *SPSS ile Veri Madenciliği*, (1. Baskı), Dikey Eksen Yayın Dağıtım.
- Saxena, A., Mittal, M., Goyal L. M., (2015), Comparative Analysis of Clustering Methods, *International Journal of Computer Applications*, 118, 30 – 35.
- Sayın, A., Yandı, A., Oyar, E., (2017), Kayıp Veri ile Baş Etme Yöntemlerinin Madde Parametrelerine Etkisinin İncelenmesi, *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(4), 490 - 510.
- Seyrek, İ. H. - Ata, H. A., (2010), Veri Zarflama Analizi ve Veri Madenciliği ile Mevduat Bankalarında Etkinlik Ölçümü, *BDDK Bankacılık ve Finansal Piyasalar*, 4(2), 67-84.
- Silahtaroglu, G., (2013), *Veri Madenciliği Kavram ve Algoritmaları*, (2. Baskı), İstanbul, Papatya Yayıncılık Eğitim.
- Şahin, Ş., (2012), *Büyük Menderes Nehri Üzerindeki Akım Gözlem İstasyonlarında Eksik Verilerinin Tamamlanması*, Pamukkale Üniversitesi Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
- Tapkan, P., Özbakır, L., Baykasoğlu, A., (2011), WEKA ile Veri Madenciliği Süreci ve Örnek Uygulama, *Endüstri Mühendisliği Yazılımları ve Uygulamaları Kongresi*, 247-262.