

**ANKARA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

DOKTORA TEZİ

**BULANIK VERİ MADENCİLİĞİ ve SERMAYE PİYASALARINA
UYGULANMASI**

Ali Serhan KOYUNCUGİL

İSTATİSTİK ANABİLİM DALI

**ANKARA
2006**

Her hakkı saklıdır

ÖZET

Doktora Tezi

BULANIK VERİ MADENCİLİĞİ ve SERMAYE PİYASALARINA UYGULANMASI

Ali Serhan KOYUNCUGİL

Ankara Üniversitesi
Fen Bilimleri Enstitüsü
İstatistik Anabilim Dalı

Danışman: Prof. Dr. Ömer L. GEBİZLİOĞLU

Hisse senetleri arz ve talebine kasıtlı müdahale olarak tanımlanan manipülasyon, fiyat mekanizmasını bozan ve piyasanın belirsizlik yapısına zarar veren en önemli nedenlerden birisidir. Geleneksel yerinde denetim gibi tespitler zaman almakta ve piyasadaki zararı telafi etmek mümkün olmamaktadır. Bilgi teknolojilerindeki gelişmeler denetime, gözetim sistemleri ve erken uyarı sistemleri olarak yansımıştır. Erken uyarı sistemlerinin çalışma mantığı, menkul kıymetlerde beklenmeyen, sıra dışı davranışların tespitine dayanmaktadır. Veri madenciliğinin tanımı ise önceden bilinmeyen, geçerli ve kullanılabilir enformasyonun büyük veri tabanlarından açığa çıkarılması ve bu enformasyonun önemli kararlar vermek için kullanılmasıdır. Bu tanıma bakıldığında, manipülasyon tespitine yönelik erken uyarı sistemlerinin, sıradışı davranış tespiti tanımıyla büyük bir örtüşmeye sahiptir.

Bu tez çalışmasında, hisse senetleri piyasası için istatistiksel öğrenme bakış açısından bulanık veri madenciliğine dayalı bir erken uyarı sistemi önerilmektedir.

Önerilen Sistem, Hisse Senetleri Piyasası (HSP) işlem akışıyla örtüşen üç basamaklı bir erken uyarı sistemi olarak tasarlanmıştır. Erken uyarı sisteminin tasarımında veri madenciliği yöntemlerinden, K-ortalamlar Kümeleme Analizi, FANNY Bulanık Kümeleme Algoritması, Bulanık Hedefli CHAID Karar Ağaçları algoritması ve Önsel Birliktelik Kuralları algoritmaları ardışık olarak kullanılmıştır.

Tasarlanan sistem, sırasıyla manipülasyon gerçekleşen hisse senedinin, sözkonusu işlemleri gerçekleştiren aracı kuruluşların ve manipülasyonu gerçekleştiren yatırımcıların ardışık olarak tespitini esas almaktadır

Finansal erken uyarı sistemlerinde genellikle bilançolara dayalı ekonometrik modellemeler esas alınırken, tasarlanan erken uyarı sisteminde İstatistiksel Öğrenme bakışıyla Bulanık Veri Madenciliği esas alınarak veri içerisindeki anomaliler belirlenmektedir. Sistem olabildiğince otomatikleştirilerek, karar süreçleri insan faktörünün subjektivesinden olabildiğince arındırılmış, verinin ne söylediği dikkate alınmıştır. Ayrıca, Bulanık Hedefli CHAID Karar Ağacı algoritması da tez kapsamında geliştirilmiştir.

Tasarlanan sistemin başarısı, manipülasyon gerçekleştirilmiş bir hisse senedinin gerçek işlem verileriyle test edilmiş ve sistemin başarıyla çalıştığı tespit edilmiştir.

2006, 186 sayfa

Anahtar Kelimeler: Bulanık veri madenciliği, istatistiksel öğrenme, bilgi keşfi, erken uyarı sistemi, hisse senetleri piyasası, manipülasyon, Bulanık CHAID Karar Ağacı algoritması, Önsel Birliktelik Kuralları algoritması, K-ortalamlar Kümeleme Analizi, FANNY algoritması.

ABSTRACT

Ph.D. Thesis

FUZZY DATA MINING AND ITS APPLICATION TO CAPITAL MARKETS

Ali Serhan KOYUNCUGİL

Ankara University
Graduate School of Natural and Applied Sciences
Department of Statistics

Supervisor: Prof. Dr. Ömer L. GEBİZLİOĞLU

Manipulation which is described as deliberate interference to supply and demand of securities is one of the most important reasons of price mechanism disorder and makes damages to ambiguity structure of market. Detection such as conventional on site supervision takes time and the compensation of damage in market becomes impossible. Improvement of information technologies reflects supervision as surveillance systems and early warning systems (EWS). Operational logic of early warning systems is based on finding unexpected and extraordinary behaviors of securities. Definition of data mining is the process of extracting previously unknown, valid and actionable information from large databases and then using the information to make crucial business decisions. Therefore, the definitions of EWS and data mining given above lead an interesting similarity.

In this thesis an EWS for securities markets is suggested that is based on a model which comprises fuzzy data mining with a statistical learning point of view.

The proposed system is designed as a three stage EWS which is similar in essence to the process of the Stock Exchange Market. The methods of data mining, K-means Clustering Analysis, FANNY Fuzzy Clustering Algorithm, Fuzzy Targeted CHAID Decision Tree algorithm and A Priori Association Rules algorithm are used sequentially for the design of the three stage EWS.

The designed system is based on the sequential detection in the order of the share that has a manipulative action, the intermediaries that realize the mentioned trading operations, and the investors that may lead to the manipulation.

While the financial early warning systems are generally based on the econometric modelling that are founded on balance sheets, in our model the designed early warning system that is established on fuzzy data mining with a statistical learning point of view identifies the anomalies. System of the model is made automatic and the decision procedures are clarified from the subjectivity of human factor as possible as it can be done, and the main message of the data is taken into consideration. The Fuzzy Targeted CHAID Decision Tree algorithm is constructed in the thesis as an integral part of the model.

The success of the designed system is tested with real transactions of a share that is manipulated, and it is confirmed that the model and its system works properly.

2006, 186 pages

Key Words: Fuzzy data mining, statistical learning, knowledge discovery, early warning system, securities market, manipulation, Fuzzy CHAID Decision Tree algorithm, A Priori Association Rules algorithm, K-means Cluster Analysis, FANNY algorithm.

TEŞEKKÜR

Bilim, öncelikle insanın doğaya, sonra da değişimi kabullenemeyen insanlara, karşı duruşudur. Bilim insanı özünde, bilineni ve öğretileni tekrar etmeyi değil; bilinmeyi araştırmalı, dogmaların, aklın üstünlüğüne galip gelmesine seyirci kalmamalıdır. Dolayısıyla, her bilimsel araştırma özünde bir karşı duruş ve bir anlamda var olana, bilinene karşı, bilinmeyi keşif amaçlı bir itaatsizliktir.

Bu tez de, Klasik İstatistik söyleminden farklı bir yaklaşıma sahiptir ve yeni bir söz söylemek adına bir anlamda karşı duruştur. Zaten yeni söylenecek bir söz olmaması halinde, bir tez yazmanın da bir anlamı kalmamaktadır.

Alışıl gelmişten farklı birşeyler ortaya koymak, doğru yaptığını inanmayı ve inanılan yolda tereddütsüz ilerlemeyi gerektirmektedir. Dolayısıyla, aslında bu tez; beni doğru bildiğim yolda yürümekten asla vazgeçmeyecek biçimde yetiştiren Annem Ziyet KOYUNCUGİL ile Babam Süleyman KOYUNCUGİL ve yürüdüğüm yolda her zaman yanımda olan Kardeşlerim Hüseyin Tolga KOYUNCUGİL ile Ferhan Yılmaz KOYUNCUGİL'in verdiği desteğin bir ürünüdür ve Onlara aittir. Bu nedenle, herkesten ve her şeyden önce, yalnız 7 yıl süren bu macerada değil, beni daima; koşulsuz, çıkarsız sonsuz sevgiyle ve anlayışla karşılayan, destek olan; yarattığım herşeyin özünü hamurumda yoğurarak beni bu güne getiren ve yaşadığım müddetçe de varlıklarından aldığım gücü bileceğim ve hissedeceğim; emeklerinin karşılığını asla veremeyeceğim Aileme, sonsuz minnetlerimi sunarım. Yalnız bu tez için değil, Ailem oldukları için minnettarım. Bu tez bana onların hediyesidir. Çünkü bu tez onlardan esirgemek zorunda kaldığım zamanlarda yazılmıştır.

Maalesef, bugün birçok akademisyen özgür bilimsel aklı etkin kılacak eleştirel bakışa sahip öğrenciler yerine, kendilerine itaat ve hatta biat edecek öğrencilerle akademik çalışma yapmayı tercih etmektedir. Her tez, söylenecek yeni bir söz demektir. Dolayısıyla, her tez, adından da anlaşılacağı gibi var olana ve değişime ayak uyduramayanlara bir karşı duruştur. Danışmanlık yapmak üzere, kendisine biat edecek bir öğrenci yerine, doğru bildiği konuda itaat etmeyen benim gibi bir öğrenciyi tercih etmiş ve bütün yoğunluğuna rağmen bana zaman yaratıp, İstatistik Bilimi'ndeki engin bilgisiyle vizyon vererek tezin en iyi biçimde sonuçlanmasını sağlamış olan Sayın Prof. Dr. Ömer L. GEBİZLİOĞLU'na; Tezin İstatistik dışında bir diğer bileşeni olan Bilişim konusunda adeta ikinci tez danışmanı gibi emek vermiş, yaptığı eleştiri ve önerilerle tezin oluşumunda büyük katkı sağlamış olan Atılım Üniversitesi Yazılım Mühendisliği Bölümü Öğretim Üyesi Sayın Prof. Dr. K. İbrahim AKMAN'a; Hisse Senetleri Piyasası uygulamaları konusunda sadece Türkiye örneği değil, uluslararası çalışmalarla da açılım sağlamış olan Ankara Üniversitesi İstatistik Bölümü Öğretim Üyesi Sayın Prof. Dr. Dolun ÖKSOY'a yalnızca Bilimsel katkılarından dolayı değil, gösterdikleri sabır ve anlayıştan dolayı da teşekkürlerimi sunarım. Ayrıca, tüm dünyada Türkiye'yi temsil etmek için sürekli bir koşuşturmaya içerisinde olan ve gene bir yolculuk sonrası, tüm zaman kısıtlarına rağmen lütfedip, ayağının tozuyla tez savunmamı dinlemek için her

zamanki sonsuz nezaketiyle teŖrif ederek, beni mahup bırakan Bilkent niversitesi
Dnya Sistemleri, Ekonomileri ve Stratejik AraŖtırmalar Enstits đretim yesi Sayın
Prof. Dr. Orhan GVENEN'e de teŖekkrlerimi sunarım.

Ali Serhan KOYUNCUGİL
Ankara, Aralık 2006

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ŞEKİLLER DİZİNİ	viii
ÇİZELGELER DİZİNİ	ix
1. GİRİŞ	1
1.1 Giriş	1
1.2 Önceki Çalışmalar	6
2. BULANIK KÜME KURAMI	10
2.1 Genel Tanımlar	10
2.2 Bulanık Küme Kuramı	14
2.2.1 α - kesme	16
2.2.2 Normallik	18
2.2.3 Dışbükeylik ve içbükeylik	19
2.2.4 Bulanık kümelerde cebirsel işlemler	20
2.2.5 Bulanık sayı kuramı	21
2.2.6 Üyelik fonksiyonları	30
2.2.7 Üyelik fonksiyonu biçimleri	31
3. VERİ MADENCİLİĞİ	38
3.1 Veri Madenciliği ve Veri Tabanlarında Bilgi Keşfi	38
3.1.1 Veri tabanlarında bilgi keşfinin tanımı	38
3.1.2 Veri tabanlarında bilgi keşfi süreci	39
3.1.3 Veri tabanlarında bilgi keşfi sürecinin veri madenciliği adımı	41
3.2 Veri Ambarı	42
3.2.1 Veri temizleme	43
3.2.2 Veri erişimi	43
3.3 Veri Madenciliğinin Tanımı	44
3.4 İstatistiksel Öğrenme	45
3.5 Veri Madenciliği İle Diğer Analitik Yöntemlerin Karşılaştırılması	47
3.5.1 İstatistiksel analiz ile veri madenciliğinin karşılaştırılması	47
3.5.2 Veri madenciliği, OLAP ve veri sorgusunun kıyaslanması	48
3.6 Veri Madenciliği ve İş Zekası	49
3.7 Veri Madenciliği Metodolojisi	50
3.7.1 Veri madenciliği için sanayilerarası standart süreç	51
3.7.2 SEMMA	53
3.7.3 Diğer veri madenciliği metodolojileri	54
3.8 İstatistiksel Öğrenme Yaklaşımlı Veri Madenciliği Yöntemlerinin Sınıflandırılması	56
3.9 Başlıca Veri Madenciliği Yöntemleri	58
3.9.1 Regresyon analizi	59
3.9.2 K-en yakın komşuluk	64
3.9.3 K-ortalamlar kümeleme analizi	65
3.9.4 Aşamalı (Hierarchical) kümeleme yöntemleri	68
3.9.5 Karar ağaçları	70
3.9.6 Sinir ağları	76

3.9.7 Birliktelik (İlişki) kuralları.....	80
3.9.8 Önemli bileşenler analizi	87
3.9.9 Diskriminant analizi.....	88
3.9.10 Kendi kendini düzenleyen haritalar	91
4. BULANIK VERİ MADENCİLİĞİ.....	93
4.1 Bulanık Sonuç Çıkarım Kuralları	94
4.2 Veri Madenciliğinde Bulanık Mantık	96
4.3 Bulanık Veri Madenciliği Yöntemleri	97
4.3.1 Bulanık kümeleme analizi.....	98
4.3.2 Hard ve bulanık c- bölünmeler.....	98
4.3.3 Objektif fonksiyon kümeleme ve hard c-ortalamlar kümeleme.....	100
4.3.4 Bulanık c-ortalamlar.....	103
4.3.5 FANNY (Fuzzy Analysis) algoritması	105
5. SERMAYE PİYASALARI.....	107
5.1 Kavram ve Tanımlar	107
5.1.1 Mali sistem.....	107
5.1.2 Mali piyasalar.....	108
5.1.2.1 Para piyasası.....	108
5.1.2.2 Sermaye piyasası.....	109
5.1.3 Mali araçlar	109
5.1.4 Birincil ve ikincil piyasalar	109
5.1.5 İkincil piyasa türleri	110
5.2 Bankaların ve Bankacılık Sisteminin Türkiye'deki Yeri.....	111
5.3 Hisse Senetleri Piyasası	112
5.3.1 Hisse senetleri piyasasının gelişimi	112
5.3.2 Borsa üyeleri ve üye temsilcileri.....	113
5.3.3 Hisse senetlerinin borsada işlem görmesi	114
5.3.4 Hisse senetleri piyasası'nda fiyatların oluşum yöntemi.....	115
5.3.5 Hisse senetleri piyasası'nın işleyişi ile ilgili tanımlar.....	115
5.3.6 Hisse senetleri piyasası'nın işleyişi.....	118
6. BULANIK VERİ MADENCİLİĞİ ve SERMAYE PİYASALARINA UYGULANMASI: HİSSE SENETLERİ PİYASASI İÇİN BİR ERKEN UYARI SİSTEMİ ÖNERİSİ.....	120
6.1 Manipülasyon.....	120
6.1.1 Spekülasyonun tanımı ve özellikleri	120
6.1.2 Manipülasyonun tanımı ve özellikleri.....	122
6.1.3 Manipülasyon çeşitleri	124
6.1.4 İşlem bazlı manipülasyonda yaygın olarak kullanılan manipülasyon yöntemleri ve davranış kalıpları.....	125
6.1.5 Manipülasyon suçunu önlemeye yönelik olarak alınabilecek tedbirler	131
6.2 Erken Uyarı Sistemleri.....	132
6.3 Dünyada ve Türkiye'de Erken Uyarı Sistemi Örnekleri.....	133
6.4 Bulanık Veri Madenciliğine Dayalı İşlem Manipülasyonu Tespitine Yönelik Erken Uyarı Sisteminin Tanımı ve İş Akışı.....	137
6.5 Bulanık Veri Madenciliğine Dayalı İşlem Manipülasyonu Tespitine Yönelik Erken Uyarı Sisteminde Kullanılan Değişkenler ve Tanımları	140
6.6 Bulanık Veri Madenciliğine Dayalı İşlem Manipülasyonu Tespitine Yönelik Erken Uyarı Sisteminin Modellenmesi	142

7. BULANIK VERİ MADENCİLİĞİNE DAYALI İŞLEM MANİPÜLASYONU TESPİTİNE YÖNELİK ERKEN UYARI SİSTEMİNİN UYGULAMASI.....	151
8. SONUÇ ve ÖNERİLER.....	165
8.1 Ardıl Değerlendirme	165
8.2 Sonraki Çalışmalara Öneriler.....	171
KAYNAKLAR	174
EK 1 Şüpheli 218 gözlemin fiyat değişim oranı (fdo_i), üyelik dereceleri ($\mu_{1,i}$ ve $\mu_{2,i}$), en büyük aidiyet derecesi (w_i) ve bulanık fiyat değişim oranı ($bfdo_i$) çizelgesi.....	178
ÖZGEÇMİŞ	184

ŞEKİLLER DİZİNİ

Şekil 2.1 Gerçek bir sayı ve deterministik bir aralık ile bulanık bir sayı ve bulanık bir aralığın karşılaştırması	22
Şekil 2.2 Bulanık sayıların temel biçimleri.....	23
Şekil 2.3 Üçgensel bulanık sayı	24
Şekil 2.4 Yamuksal bulanık sayı.....	28
Şekil 3.1 VTBK sürecinin adımları	40
Şekil 3.2 Veri Madenciliği ve İş Zekası.....	50
Şekil 3.3 VMSSS (CRISP-DM) Metodoloji Döngüsü	52
Şekil 3.4 İki tahmin ediciyle en küçük kareler regresyonunun N-boyutlu geometrisi ...	62
Şekil 3.5 χ_0 noktasının 7 – en yakın komşuluğu	64
Şekil 3.6 Dendogram.....	69
Şekil 3.7 Bir karar ağacı örneği	70
Şekil 3.8 Bir gizli katmanlı, ileri beslemeli sinir ağı şeması.....	77
Şekil 3.9 Sigmoid fonksiyonunun grafiksel gösterimi.....	79
Şekil 3.10 Her biri Gausyen Dağılımdan gelen örnekler için Doğrusal Diskriminant Analizi'ne uydurulmuş modelin karar sınırları	90
Şekil 4.1 Veri madenciliği görevleri, belirsiz muhakeme teknikleri ve bunların entegrasyonu	93
Şekil 4.2 K-ortalamlar kümeleme ve bulanık kümeleme.....	98
Şekil 5.1 Mali Sistem.....	107
Şekil 6.1 Erken uyarı sistemi iş akışı diagramı	148
Şekil 7.1 Hedef değeri bulanık CHAID karar ağacı	155

ÇİZELGELER DİZİNİ

Çizelge 3.1 İstatistiksel analiz ve veri madenciliğinin karşılaştırılması	48
Çizelge 3.2 Veri madenciliği için kullanılan temel metodolojiler	51
Çizelge 4.1 Sonuç Çıkarım Kuralı Çeşitleri.....	95
Çizelge 6.1 Erken uyarı sistemi iş akışının işlem ve yöntem değerlendirmesi	149
Çizelge 6.2 Erken uyarı sisteminde uygulanan analizler ve kullanılan yazılımlar	150
Çizelge 7.1 3-ortalamalar kümeleme analizi sonucunda elde edilen son küme merkezleri.....	152
Çizelge 7.2 3-ortalamalar kümeleme analizi sonucunda elde edilen her bir kümedeki gözlem sayıları	152
Çizelge 7.3 Birliktelik kuralları.....	160
Çizelge 7.4 3-ortalamalar kümeleme analizi sonucunda elde edilen son küme merkezleri.....	162
Çizelge 7.5 3-ortalamalar kümeleme analizi sonucunda elde edilen her bir kümedeki gözlem sayıları	163

1. GİRİŞ

Tezin gerekçelerine yönelik olarak hazırlanan bu bölüm iki başlık halinde incelenecektir. Tezin amacına yönelik açıklamalar Giriş Bölümü olan Bölüm 1.1’de, çalışmaya yön veren önceki çalışmalar ise Bölüm 1.2’de Önceki Çalışmalar başlığı altında incelenecektir.

1.1 Giriş

Sadece sayım gibi büyük ölçekli çalışmalarda kullanılan 1950’lerin devasa boyutlu ilk bilgisayarları; günümüze gelindiğinde artık avuç içine sığacak derecede küçülmüş ve günlük hayatımızın bir parçası haline gelmiştir. Neredeyse günlük yaşamın akışındaki tüm iş ve işlemler bilgisayarla veya bilgisayar destekli olarak gerçekleştirilmeye doğru ilerlemektedir. Bu açıdan bakıldığında, küçük bir işletmenin dahi günlük işlem verileri zaman içerisinde katlanarak büyümekte ve fiyatları anlamlı derecede düşen sabit diskler sayesinde saklanabilmektedir.

1990’lara kadar, yaygın olarak ‘Dijital Hafıza’ çağrışımı yapan bilgisayarlar, geçmiş işlemlere ait verileri gerektiğinde sunmakta ve bir takım hesaplamalara izin vermekteydi. 1990’lardan itibaren ise bilgisayarlarda, çoğunlukla da veri tabanlarında, depolanan veri yığınlarının nasıl kullanılabilceği sorgulanmaya başlamış; nihayetinde yığınla verinin içerisinde gözden kaçan ve keşfedilmeyi bekleyen bilgiler olduğu düşüncesi hakim olmuştur. Bu noktadan itibaren veri yığınları arasında sıkışıp kalmış bilgilerin nasıl açığa çıkarılabileceği üzerine yoğunlaşmış ve Veri Tabanlarında Bilgi Keşfi (VTBK, Knowledge Discovery in Databases - KDD) kavramı ortaya çıkmıştır (Piatetsky-Shapiro 1991).

VTBK kavramının oluşumunda, parça parça, farklı format ve ortamlarda verinin ayrı ayrı değerlendirilmesi yerine; biraraya getirilerek, aralarında doğrudan ilişki bulunması düşünülmeyen değişkenlerin (alanların), dolaylı birtakım ilişkileri ima edebileceği mantığından hareket edilmiştir. Veri madenciliği ise VTBK’nin özünü oluşturan keşif

kısının gerekleřtiđi bir adım olarak alınabileceđi gibi bađımsız bir sre olarak da deđerlendirilebilmektedir.

Veri madenciliđinin yeni geliřen bir alan olması nedeniyle, farklı kaynaklarda farklı tanımlar yapıldıđı grlmektedir (Cabena *et al.* 1997). Buna karřın, veri madenciliđi genel anlamda; byk miktarda veri ierisinden, gizli kalmıř, deđerli, kullanılabilir bilgilerin aıđa ıkarılması biiminde ifade edilebilir. Veri ierisinde keřfi bekleyen bilgilerin varlıđı dřncesi sayesinde, İstatistik farklı bir bakıř aısıyla yeniden keřfedilmiř; zaman ierisinde Makine đrenimi ve Yapay Zeka desteđiyle neredeyse dřnebilen Akıllı Algoritmalar ile İstatistiđin birleřimi, Veri Madenciliđi’ni meydana getirmiřtir. řphesiz ki veri madenciliđini biliřimden ayrı tutarak salt istatistik olarak dřnmek ok dođru bir yaklařım olmayacaktır. Temel anlamda, verinin yer aldıđı veri tabanları ve zel bir řekli olan veri ambarları, yapay zeka ve makine đrenimine dayalı yntemler de sz konusudur. Veri madenciliđi ile Klasik İstatistik arasında belli noktalarda ayrıřmalar sz konusu olsa da, ierisinde istatistiđi telaffuz etmeden veri madenciliđinden bahsetmek mmkn olmamaktadır. yleki, bu konu en iyi biimde ‘İstatistikiler, Veri Madenciliđini elle yaparlar.’ ifadesiyle zetlenebilir.

Veri madenciliđinden daha eski bir gemiře sahip olsa da, 1990’lardan itibaren yaygın kullanıma sahip teorilerden birisi de Bulanık Teori’dir. 1930’lu yıllarda temel dřnce yapıtařları atılan ve Zadeh (1965) tarafından geliřtirilen Bulanık Teori, belirsizliđin kullanımı aısından kaınılacak bir durum deđil; gerek hayat problemlerine uygun zmler sađlamada gl bir done olduđu dřncesini yerleřtirmiřtir. Sadece olasılıđın, belirsizliđi aıklamada yetersiz kaldıđı durumlarda, Bulanık Mantıđın etkin zmler sunduđu grlmektedir. Bulanık teori Aristo’nun iki deđerli ‘Kesin Kme Kuramı’na, ok deđerli mantıđa tařıyarak bir kmeye aidiyeti kesin olarak ‘ait’ veya ‘ait deđil’ biiminde ifade etmek yerine aidiyet derecelerini gsteren ‘yelik Fonksiyonu’na yer vermiřtir.

Trkiye’de finansal sistem irdelendiđinde, bankacılık sisteminin esas alındıđı ve byk ađırlıđa sahip olduđu bir yapı gze arpmaktadır. Finansal sistemin diđer bileřeni olan

sermaye piyasalarının ise ekonomik sistemde görece olarak henüz yeni bir aktör olması nedeniyle fazla bir etkinliği olmadığı görülmektedir. Kasım 2000 ve Şubat 2001’de ard arda yaşanan iki ekonomik krizin ülke ekonomisinde yarattığı yıkım Bankacılık Sistemi’nin finansal yükü taşıyamadığı yönünde eleştirileri artırmıştır. Bu eleştirilerin ardından, ülkemizdeki düzenlemeler nedeniyle riskin devlete ait olduğu bankacılık sisteminden, riskin yatırımcıya ait olduğu sermaye piyasalarına gözler çevrilmiş ve krizleri önlemek açısından güçlü bir sermaye piyasasının varlığının önemi bir kez daha belirlenmiştir. Ancak, Türkiye’de sermaye piyasalarının toplam finansal sistem içerisindeki payı %5’ler düzeyindedir. Türkiye’nin güçlü bir finansal sisteme sahip olabilmesi ve kriz riskinin azaltılması için sermaye piyasalarının toplam finansal sistem içerisindeki payının artırılması yönünde çalışmalara ağırlık verilmiştir. Bu noktada, sermaye piyasalarının Türkiye’deki gelişiminin önündeki engeller sorgulanmaya başlamıştır. Sermaye piyasaları için şeffaflık, güven ve bilginin eşit olarak dağıtıldığı bir ortam sağlanmasının yalnız Türkiye için değil dünyadaki tüm sermaye piyasaları için temel gereklilik olduğu gerçeğinin bir kez daha altı çizilmiştir.

Güven duyulabilecek etkin bir piyasanın varolabilmesi için,

- Manipülasyon,
- İçerden öğrenenlerin ticareti,
- Muhasebe hileleri,

gibi sermaye piyasası suçlarının en alt düzeyde gerçekleşmesi gerekmektedir.

Sermaye piyasalarının işlevini yerine getirmesi, etkin bir fiyat mekanizmasını gerektirmektedir. Bu noktada fiyat mekanizmasını bozan ve işlevini yerine getirmesine engel olan unsur olarak manipülasyon dikkat çekmektedir. Manipülasyon en basit ifadesiyle, menkul kıymetlerin arz ve talebine kasıtlı müdahaleler olarak tanımlanabilir.

Çeşitli uluslararası geçerlilikte davranış kalıplarına sahip olan manipülasyon genel anlamda, denetim bazlı olarak araştırılmakta ve araştırmalar piyasalara verilen zararın telafisine imkan vermeyecek ölçüde zaman almaktadır. Zaman içerisinde teknolojinin gelişimiyle birlikte, manipülasyon davranış kalıplarının, bilgi teknolojileri yardımıyla

yakalanabileceği düşüncesi Gözetim Sistemleri’ni veya daha özelde Erken Uyarı Sistemleri’ni gündeme getirmiştir. Sözkonusu sistemler, mantık olarak, insan gözüyle, manipülasyon gerçekleşikten sonra yapılan denetimlerin yerine, bilgi teknolojileri ile tespitini esas almaktadır. Günümüzde başta ABD olmak üzere, Fransa, Kore gibi pek çok ülkede kullanılan Gözetim Sistemleri, ülkemizde de kullanılmasına karşın; erken uyarı mekanizmasına sahip etkin bir sistem sözkonusu değildir.

ABD’nin sermaye piyasaları açısından en gelişmiş ülke olması nedeniyle en çarpıcı örnek New York Borsası’ndan verilebilir. New York Borsası Piyasa Gözetim Birimi, borsaya kote şirketlerin tahta işlemlerini gerçek-zamanlı ve işlem sonrasında olmak üzere görüntülemek ile sorumlu birimdir. Birim işlemlerin piyasa kurallarına uygun olup olmadığını, kötü amaçlı veya manipülatif ve içerden öğrenenlerin ticareti olup olmadığını gözetlemektedir. Birim olağan dışı ve ihlalcı alım satım örüntülerini karmaşık bilgisayar teknolojileri kullanarak tespit etmektedir.

New York Borsası piyasa gözetim amacıyla ‘Stock Watch’ isimli bir sistem kullanmaktadır. New York Borsası Piyasa Gözetim biriminin Stock Watch ile ilgili olarak, ‘Piyasa Gözetimi’nin Stock Watch birimi analistlerin insan yargısını elektronik veri madenciliği ve örüntü tespit sistemleri birleştirmekte; haberler, araştırmalar kote şirketlerin çalışanları, yöneticilerinin veri tabanı ile bağlantılar kurmakta ve olası içeriden öğrenenlerin ticaretini tespit etmektedir (NYSE 2005).’ açıklaması yer almaktadır.

Stock Watch örneği de dikkate alındığında, veriler arasında gözlenemeyen ilginç ve gizli kalmış bilgilerin keşfedilmesi tanımına sahip veri madenciliği ile işlem manipülasyonu tespitine yönelik erken uyarı sistemi iyi bir eşleşme sağlamaktadır.

Etkin Market Teorisi’ne göre hisse senetleri piyasasında fiyatların oluşumu tamamıyla rasgeledir (Kovalerchuk and Vityaev 2002). Fiyat oluşumlarının önceden tahmin edilemiyor olması, etkin bir piyasanın temel varsayımlarından birisidir. Bunun bir adım ötesinde ise Kaotik Piyasa Hipotezi’ne göre hisse senetleri piyasası kaotik bir yapıya

sahiptir (Kovalerchuk and Vityaev 2002). Dolayısıyla, hisse senetleri piyasası için belirsizlik esastır. Bu mantıktan hareketle, hisse senetleri piyasasını açıklamakta veya hisse senetleri piyasası üzerine çalışmakta Kesin Küme Kuramı yetersiz kalabilmektedir. Kesin Küme Kuramı yerine Bulanık Küme Kuramı ile çalışmak hisse senetleri piyasasının belirsizlik içeren yapısına daha uygun düşmektedir.

Yukarıdaki çerçeve incelendiğinde Sermaye Piyasaları üzerine veri madenciliği ile ilgili bir uygulama yapmaya en uygun alanın, hisse senetleri piyasasında manipülasyon tespitine yönelik bir erken uyarı sistemi olduğu görülmektedir. Hisse senetleri piyasasının kesin küme kuramı yerine bulanık küme kuramı ile daha iyi açıklanabileceği dikkate alındığında Bulanık Küme Kuramı'na dayalı Veri Madenciliği yöntemlerinin hisse senetleri piyasası için bir erken uyarı sistemi geliştirmek için oldukça uygun olduğu görülmektedir.

Veri madenciliğinin içeriğine bakıldığında, veri organizasyonu, işlem hızı gibi veri tabanı özelinde bilişim bakış açısı ve keşif aşamasının gerçekleştirildiği analiz yöntemlerini esas alan istatistik temelli bakış açısı olmak üzere iki yaklaşım üzerine yoğunlaşıldığı görülmektedir. Bu çalışmada, İstatistik Bilimi'nin veri madenciliğine uyarlanmış yöntemleri irdelenecektir. Dolayısıyla, veri madenciliği bilişim bakış açısıyla değil istatistik bakış açısıyla irdelenecektir. Bir başka deyişle, İstatistiksel Öğrenme esas alınacaktır.

Bu tez çalışmasında, Bulanık Veri Madenciliği'ne dayalı olarak Hisse Senetleri Piyasasında manipülasyon tespitine yönelik bir erken uyarı sistemi geliştirilecektir.

Tez çalışmasının 1. Bölümünde, çalışmanın amacı ve gerekçeleri önceki çalışmalarla açıklanmaktadır.

2. Bölümde, bulanık küme kuramı anlatılmaktadır.

3. Bölümde, veri madenciliği, genel bilgi keşfi süreci içerisindeki yeri, tarihsel gelişimi, metodolojisi ve kullanılan yöntemler açısından değerlendirilecektir.

4. Bölümde, bulanık küme kuramı ile veri madenciliği arasındaki bağlantı kurulup, bulanık veri madenciliği aktarılacaktır.

5. Bölümde, sermaye piyasalarına değinilecektir. Bu bölümde sermaye piyasalarının tanım ve kavramları verilir; ekonomik sistem içerisindeki yeri aktarılacaktır. Tezin uygulama alanı olan Hisse Senetleri Piyasası da bu bölümde incelenecektir.

6. Bölümde, bulanık veri madenciliğinin sermaye piyasaları uygulaması olan Hisse Senetleri Piyasaları İçin İşlem Manipülasyonu Tespitine Yönelik Erken Uyarı Sisteminin tanımı ve modellemesi aktarılacaktır. Ayrıca, tezin uygulamasına temel olan manipülasyon ve erken uyarı sistemleri de incelenecektir.

7. Bölümde, Erken Uyarı Sistemi gerçek verilerle işletilip, elde edilen sonuçlarla sistemin başarısı irdelenecektir.

8. Bölümde sonuç ve öneriler yer alacaktır.

1.2 Önceki Çalışmalar

Tez çalışması, metodolojik açıdan istatistiksel öğrenme yaklaşımı olarak veri madenciliği ve bulanık küme teorisini temel alırken; uygulama alanı olarak sermaye piyasaları özelinde hisse senetleri piyasasında işlem manipülasyonu tespitini esas almaktadır. Tez çalışmasına yön veren çalışmalar bu kapsamda aşağıda sunulmaktadır.

Uygulama alanı olarak tercih edilen Hisse Senetleri Piyasası'nın doğası belirsizliğin ötesinde bulanık bir yapı sergilemektedir. Zadeh (1965), kesin küme kuramının çözüm sunamadığı durumlara bir alternatif olarak önerdiği bulanık kümelere yönelik ilk

çalışmayı yapmıştır. Zadeh'in (1965) yaptığı çalışmalar 1990'lardan itibaren yeniden gündeme gelerek, Terano *et al.* (1991) bulanık sistemlerin teorisi ve uygulamaları hakkında derleyici nitelikteki bir çalışma yapmışlardır ve bulanık kümelerin teori ve uygulamaları konusunda ilk çalışmalardan, son gelişmelere kadar tamamı Klir and Yuan (1995) tarafından derlenmiştir. İlk başta matematiksel programlama alanında uygulama yaygınlığı kazanan Bulanık Küme Teorisi zaman içerisinde istatistik başta olmak üzere bulanıklığın esas olduğu gerçek dünya problemlerine başarıyla uygulanmış ve Koyuncugil (1999) stokastik hedeflere sahip bir hedef programlama modeline bulanık yaklaşımla yatırım problemi için çözüm sunmuştur.

Bulanık Küme Teorisi teknolojiye paralel olarak yaygın kullanıma kavuşurken, teknolojinin bir sonucu olan bilgi sistemlerinin etkinliği ve özelden veri tabanlarının yaygın kullanımı sonucunda ulaşılan büyük hacimde veri, Veri Tabanlarında Bilgi Keşfi (VTBK) kavramını ortaya çıkarmıştır. VTBK ifadesi ilk defa Piatetsky-Shapiro (1991) tarafından 1989 yılında gerçekleştirilen ilk VTBK çalışma grubu toplantısında kullanılmış, konuyla ilgili kavram ve tanımlar ortaya konulmuştur. Ayrıca, Veri Madenciliği (VM) terimi de VTBK'nin bir bileşeni olarak tanımlanmıştır. Tanımlar bir taraftan tartışıldukurken, veri madenciliği yöntemleri geliştirmeye yönelik çalışmalar da devam etmiş, Agrawal *et al.* (1996) nicel birliktelik kurallarının madenciliği için hızlı bir algoritma olan Apriori'yi önermişlerdir. VTBK'nin temel süreçleri üzerinde hiyerarşi arayışının sonucunda, Fayyad *et al.* (1996) veri tabanlarında bilgi keşfinin süreçlerine ve veri madenciliğinin bu süreçteki yerine yönelik bir akış sunmuşlardır. Ayrıca, veri madenciliğinin temel özelliklerini irdelemişlerdir. İş zekasının en yeni ve popüler bileşeni olan veri madenciliğinin yerini belirlemek amacıyla, Cabena *et al.* (1997) veri madenciliğinin diğer iş zekası çözümleriyle mukayesesini yapıp, aralarındaki hiyerarşiyi ortaya koymuşlardır. VM'nin uygulama yaygınlığı kazanmasıyla birlikte uygulamaya yönelik çalışmalar ağırlık kazanmaya başlamış ve Berson *et al.* (1999) VM'nin en yaygın kullanıma sahip alanı olan müşteri ilişkileri yönetimi kapsamında veri madenciliği yöntemleri ve uygulamalarına yer vermişlerdir. Ayrıca, yeni nesil yöntemlerden olan Karar ağacı yöntemlerinin başlıcalarından CART ve CHAID'i incelemişlerdir. Şüphesizki veri madenciliğinin vazgeçilmez bileşenlerinden birisi de kullanılan yazılımlardır. VM'nin ticari anlamda kazandığı

yaygınlığa paralel olarak bilimsel gelişimini de sağlayabilmesi için akademik amaçlı çeşitli VM yazılımları geliştirilmiş ve Witten and Frank (2000) veri madenciliği yazılımı WEKA üzerinden veri madenciliği yöntemlerinin uygulanmasına değinmişlerdir.

VM'yi bilişim ve istatistik olmak üzere iki bakış açısıyla incelemek mümkündür. VM'yi istatistiksel açıdan etüd etmek bu tezin bakış açısını ortaya koyuyor olup; Hastie *et al.* (2001), veri madenciliğine istatistiksel perspektiften yaklaşmış ve veri madenciliği ile istatistik arasındaki ilişkiyi 'İstatistiksel Öğrenme' kavramı ile kurmuşlardır. 1990'lar veri madenciliğinin bilişim açısından gelişim dönemi olurken 2000'li yıllardan itibaren VM'nin analitik tarafı ağırlık kazanmaya başlamış; istatistiksel yöntemlerin büyük veri setlerine uygulamaları üzerine çalışmalar ağırlık kazanmıştır. Rao (2001), istatistiğin geçmişi ve gelecek vizyonuna değinip, veri madenciliğine yönelimi istatistiksel açıdan değerlendirmiştir. Veri madenciliğinin gelişimini sürdüren bir alan olması nedeniyle; diğer analitik yöntemlerle arasındaki ayrımın belirginleştirilmesi önem kazanmıştır. Bu noktadan hareketle, Roiger and Geatz (2002), veri madenciliği ile SQL sorgusu ve OLAP arasındaki ayrımı ortaya koyup; bilgi ihtiyacının düzeyine göre uygulanması gereken yöntemlere yönelik bir çerçeve çizmişlerdir. Moss and Atre (2003) veri madenciliği ile istatistiksel analiz yöntemlerini karşılaştırmış ve aralarındaki farklılıkları ortaya koymuşlardır.

Ayrıca, Chen (2001) veri madenciliği ile belirsiz muhakeme teknikleri arasındaki ilişkiyi ve bağlantıları ortaya koymuş, bulanık teorinin veri madenciliği ile kesişimine yer vermiştir. Bunun yanısıra, Kovalerchuk and Vityaev (2002), veri madenciliğinin finansdaki yeri ve kullanımını, veri madenciliği yöntemlerinin yanı sıra hibrit yöntemlerin kullanımını sunmuştur.

Tezin uygulama alanını sermaye piyasalarında hisse senedi işlem manipülasyonu oluşturmaktadır. Sermaye Piyasası Özel İhtisas Komisyonu'nun, Anonim (1989) olarak hazırladığı raporla sermaye piyasalarının Türkiye ve dünyadaki yeri ve önemine değinilmiş, sermaye piyasalarının Türkiye'de gelişimi için önerilerde bulunmuşlardır.

Özgün uygulamaya konu olan manipülasyon ve spekülasyon sıklıkla karıştırılan ve üzerinde durulması gereken bir konu olarak karşımıza çıkmaktadır. Özbay (1990) borsalarda spekülasyon ve manipülasyonu incelemiştir. Şüphesiz ki, özgün uygulamayı oluşturan erken uyarı sistemlerini etüd etmek de gerekmektedir. Tezcanlı (1993), içerden öğrenenlerin ticareti ve piyasa manipülasyonu kavramlarını tartışmış ve bir gözetim modeli önerisinde bulunmuştur. Ataman (1999), gelişmiş piyasalarda uygulanan manipülasyon inceleme yöntemlerini sorgulamış ve gözetim ile erken uyarı sistemlerine değinerek uluslararası örneklere yer vermiştir. Güngör (2001) içerden öğrenenlerin ticareti ve manipülasyonu yasal yönden incelemiştir. Karan ve Karacabey (2003), sermaye piyasalarının Türkiye'deki seyrini ve geleceğini uluslararası karşılaştırmalarla tartışmış; sorunlarını tespit edip, geleceğe yönelik çözüm önerilerinde bulunmuşlardır.

2. BULANIK KÜME KURAMI

Hisse senetleri piyasasının içerdiği kaotik sistemi modellemek için belirsizliğin ötesinde bulanıklıktan yararlanmak gerekmektedir. Bu nedenle, Bölüm 6'da verilecek olan Erken Uyarı Sistemi'nin modellenmesinde Bulanık Küme Kuramı'ndan yararlanmak gerekmektedir. Dolayısıyla, Sistem'de yer verilecek bulanık yaklaşım için Bulanık Küme Kuramı'nın irdelenmesi gerekmektedir. Bu noktadan hareketle, öncelikle, Bulanık Küme Kuramı hakkında genel tanımların verilmesi gerekmektedir.

2.1 Genel Tanımlar

Klasik yöntemler, yalıtılmış, basit ve anlaşılır olgulardan oluşan problemlerin çözümünde iyi sonuçlar vermesine karşın, karmaşık ve etkileşimli problemlerin çözümünde yetersiz kalmaktadır. Teknoloji ve buna bağlı olarak da bilimdeki hızlı gelişim, karar süreçlerinde karmaşayı artırmasının yanısıra belirsizliğe de yol açmaktadır. Bu durum klasik yöntemleri, problemlerin çözümünde çıkmaza sokmaktadır. Gerçek dünyaya ilişkin problemlere bakıldığında, kesin (deterministik) olmasından çok belirsizlik içermektedir. Bulanık küme kavramı, klasik matematiğin yetersiz kaldığı, belirsiz veya kesin olmayan karar durumlarına, kesinlik kazandırıp çözümdeki açmazı ortadan kaldıran, kavramlar ve yöntemler sunmaktadır.

Bulanık teriminin temelde iki anlamı vardır. Çok özel olarak dar anlamda, bulanık mantık FL_n , klasik çok değerli (multivalued) mantığın genelleştirilmiş uzantısının görüntüsü olan bir mantıksal sistemdir. Daha geniş anlamda, bulanık mantık FL_w , bulanık setler teorisi ile hemen hemen eşanlamlıdır. Dar anlama oranla daha geniş olarak düşünüldüğünde, etkin olarak, bulanık mantık ile FL_w sık sık eşit olarak kullanılır.

Bu yüzyılda matematik ve bilimde herkes tarafından genel anlamda kabul gören birçok değişiklik olmuştur. Bunlardan birisi de belirsizlik düşüncesiyle ilgilidir. Belirsizlik

bilimde istenmeyen ve mümkün olduğunca kaçınılan bir durumdur. Geleneksel görüşe göre bilim sadece belirli durumlar için çalışmalarını ortaya koymalıdır. Bu nedenle belirsizlik bilimsel olmayan bir durum olarak alınmalıdır. Alternatif veya modern görüşe göre ise belirsizliğin bilim için gerekli olduğu gözönüne alınmalıdır. Sadece kaçınılması gereken bir dert olarak görülmemeli, büyük yarar sağlayan bir olgu olarak değerlendirilmelidir (Klir and Yuan 1995).

Bir model kurulurken bunun en kullanılabilir biçimde oluşturulmasına dikkat edilir. Burada amaç, bütün model sistemlerinde üç anahtar karakteristik olan; karmaşıklık, güvenilirlik ve belirsizlik arasındaki ilişkiler için mümkün olduğunca sıkı bağlantı kurmaktır. Bu ilişkiler tam anlamıyla anlaşılır olmayabilir. Yalnızca şu biliniyor ki belirsizlik, modeller sisteminin mümkün olduğunca kullanışlı olmasında ana eksen olan bir role sahiptir. Genel olarak çok fazla belirsizlik, karmaşanın azalmasına ve modelin sonuçlandırılmasında güvenilirliği artırmaktadır. Burada temel sav, sistemleri modellemede, belirsizlikten optimal düzeyde faydalanıp, herbir model problemini tahmin ederek, metodlar geliştirmektir. Böylece belirsizlik, modelleme işinde önemli rol oynamakta, amaca uygun modelin diğer temel karakteristiklerini elde etmektedir. Bu önemli rolü tahmin etmek için bazı araştırmacılar tarafından çalışmalar yapılmış, 1960'ların ikinci yarısından itibaren belirsizlik hakkında geleneksel görüş yerini modern görüşe bırakmaya başlamıştır. Yeni teoriler belirsizlik ile olasılık teorisi arasındaki belirgin farkı ortaya koymuştur. Daha önceden belirsizlik ile olasılık teorisi birbiriyle bağlantılı alınmasına karşın görülmüştür ki, olasılık teorisi, belirsizliğin birçok farklı durumundan sadece bir durumu için geçerlidir (Klir and Yuan 1995).

Zadeh (1965), bu önemli noktadaki genel kabul üzerine modern görüş hakkında bir makale hazırlamıştır. Fakat bu konudaki ilk çalışma Zadeh'ten yaklaşık 30 yıl önce Amerikan filozofu Black tarafından yapılmıştır. Zadeh (1965), bulanık kümelerdeki üyelik fonksiyonlarının doğrulayıcı veya yoksayıcı olgular değil ama, daha doğrusu derece veren olgular olduğunu belirtmiştir. Klasik yöntemlere alternatif olarak Zadeh (1965) tarafından temeli oluşturulan bulanık küme kuramı, ilk başta yöneylem alanında uygulanmasına karşın, daha sonraları işletme, yapay zeka, kontrol kuramı ve istatistik başta olmak üzere pekçok alanda kullanılmaktadır (Klir and Yuan 1995).

Zadeh (1965)'in çalışmasında belirsizliğin sadece olasılık teorisi olmadığı ortaya koyulmuş fakat, birçok oluşumda olasılık teorisi temel alınmış, Aristo'nun iki değerli mantığı (Aristotalian two-valued logic) kullanılmıştır. A bir bulanık küme ve x bu kümenin elemanı ise , 'x , A'nın üyesidir' denilmiştir. Bu ifadenin kullanımı için iki değerli mantıkta olduğu gibi doğru veya yanlıştan ikisinden birinin tamamen geçerli olması gerekmemektedir. Sadece x'in A'nın üyesi olduğu derece kadar, doğru olmalıdır. Bulanık kümelerdeki [0,1] aralığında yer alan üyelik derecelerini doğruluk derecesi olarak almak çok kullanılan, fakat istenmeyen bir görüştür (Klir and Yuan 1995).

Matematiksel anlamda incelendiğinde kümeler, klasik teoride ifade edilen kesin kümeler ve bulanık teoride ifade edilen bulanık kümeler olarak ikiye ayrılabilir.

Klasik teoriye bakıldığında, X bütün mümkün sonuçların kümesi yani Evrensel Küme; A Evrensel kümenin alt kümesi olan herhangi bir küme olmak üzere; x, A kümesinin elemanı ise

$$x \in A \quad (2.1)$$

biçiminde gösterilir. Aksi taktirde; x , A kümesinin elemanı olmadığında ise

$$x \notin A \quad (2.2)$$

biçiminde gösterilir.

Daha açıklayıcı olması için evrensel küme X;

$$X = \{a, b, c, d, e, f\} \quad (2.3)$$

olsun. A kümesi ise;

$$A = \{a, c, e\} \quad (2.4)$$

biçiminde gösterilsin.

A kümesi çok sık kullanılan ‘karakteristik fonksiyon’ adı verilen bir fonksiyon olsun. Öyleki kümenin üyeleri X kümesinin elemanıdır ya da değildir. A kümesinin karakteristik fonksiyonu X_A olmak üzere,

$$X_A(x) = \begin{cases} 1, & x \in A \\ 0, & x \notin A \end{cases} \quad (2.5)$$

biçiminde gösterilir. Karakteristik fonksiyon kümesinin elemanları ise $\{0,1\}$ kümesinin elemanıdır.

Karakteristik fonksiyonun gösterimi ise,

$$X_A : X \rightarrow \{0,1\} \quad (2.6)$$

biçiminde olmaktadır. Herbir $x \in X$ için, $X_A(x) = 1$ olduğu zaman ‘x, A’nın üyesi’ olarak kabul edilir; $X_A(x) = 0$ olduğu zaman ise ‘x, A’nın üyesi değil’ olarak kabul edilir. Bu tanımlamalara göre X kümesi ,

$$X = \begin{matrix} a & b & c & d & e & f \\ 1 & 1 & 1 & 1 & 1 & 1 \end{matrix} \quad (2.7)$$

olmak üzere, A kümesi ise

$$A = \begin{matrix} a b c d e f \\ 1 0 1 0 1 0 \end{matrix} \quad (2.8)$$

olacaktır.

Herhangi bir kümede evrensel kümenin elemanı olma veya olmama, karakteristik fonksiyonun aldığı 1 veya 0 değeri göz önünde bulundurularak değerlendirilir. Buna göre A kesin bir küme olmak üzere A'nın üyelik fonksiyonu μ_A ;

$$\mu_A : X \rightarrow \{ 0,1 \} \quad (2.9)$$

biçiminde ifade edilir (Klir and Yuan 1995).

2.2 Bulanık Küme Kuramı

İşlem manipülasyonu tespitini gerçekleştirmek üzere tasarlanan Erken Uyarı Sistemi'nde, Hisse Senetleri Piyasası (HSP) işlemleri dikkate alınmaktadır. HSP'nin kaotik yapısı dikkate alındığında, işlemlere yönelik bir modellemede Kesin Küme Kuramı yerine Bulanık Küme Kuramı'ndan faydalanmak daha gerçekçi olmaktadır. Bu noktadan hareketle, şüpheli işlemlerin manipülatif olma durumunun sorgulanmasında Bulanık Küme Kuramı'ndan yararlanılmış ve Bulanık Küme Kuramı'na yönelik detaylar aşağıda tartışılmıştır.

Evrensel kümenin tanım aralığında belli üyelik dereceleri tanımlanarak eşitlik (2.9)'daki fonksiyon genelleştirilebilir. Daha büyük değerler, üyelik kümesinde daha yüksek dereceleri gösterir. Bulanık küme, üyeleri kesin olarak belli olmayan ama aday üyelerin bu kümeye ait üyelik derecelerinin bulunduğu bir kümedir (Tuncel 1996).

$$\forall x \in X : \mu_A(x) \in [0,1]$$

için A bulanık kümesi ;

$$A = \{ (x, \mu_A(x)) \mid x \in X \} \quad (2.10)$$

olmak üzere, X evrensel kümesi sonlu ise,

$$\forall x \in X : A = \sum_{i=1}^n \frac{\mu_A(x_i)}{x_i} \quad (2.11)$$

biçiminde, X evrensel kümesi sonlu değil ise,

$$\forall x \in X : A = \int_X \frac{\mu_A(x_i)}{x_i} \quad (2.12)$$

biçiminde ifade edilir.

Bulanık kümelerde genellikle üyelik fonksiyonunun aldığı değerler $[0,1]$ aralığında yer almasına karşın farklı aralıklarda da üyelik fonksiyonları tanımlamak mümkündür. Herbir üyelik fonksiyonunun elemanı X evrensel kümesinin içinde yer almalı ve tanımlanan aralıkta gerçek sayılar olmalıdır.

A bulanık kümesinin üyelik fonksiyonu μ_A ile gösterilip,

$$\mu_A : X \rightarrow [0,1] \quad (2.13)$$

biçiminde ifade edilir.

Görüldüğü gibi kesin ve bulanık kümeler arasındaki temel farklılık; elemanların aldığı üyelik derecelerinden kaynaklanmaktadır. Kesin bir kümenin dereceleri yalnız 0 veya 1 değerlerini alırken, bulanık kümenin elemanlarının üyelik dereceleri $[0,1]$ aralığındaki herhangi bir değeri alabilmektedir.

Birçok bulanık küme; az, orta, çok veya çok düşük, düşük, orta, yüksek, çok yüksek gibi sözel ifadelerle tanımlanır ve bunlar büyük harflerle gösterilen değişkenlerle ifade edilir. Bu değişkenlere ise bulanık değişkenler adı verilir (Klir and Yuan 1995, Tuncel 1996).

A bulanık kümesinin desteği, X evrensel kümesinin bir altkümesi olup,

$${}^0 A = \{ x \in X \mid \mu_A(x) > 0 \} \quad (2.14)$$

biçiminde tanımlanır.

Eğer destek kümesi gerçel sayılardan oluşup, üyelik fonksiyonu da sürekli ise, dışbükey bulanık bir kümenin α - kesmesi kapalı bir aralıktır (Klir and Yuan 1995, Tuncel 1996).

Alt kesimde, α - kesmesi daha ayrıntılı olarak verilecektir. Daha sonraki kesimlerde ise bulanık teorideki temel özellikler anahatları ile tanımlanacaktır.

2.2.1 α - kesme

A bulanık kümesinin $\alpha \in (0,1]$ - kesmesi, X evrensel kümesinin kesin bir altkümesi olup matematiksel olarak,

$${}^{\alpha}A = \{ x \in X \mid \mu_A(x) > \alpha \} \quad (2.15)$$

biçiminde tanımlanır. Bunun yanısıra,

$${}^{\alpha+}A = \{ x \in X \mid \mu_A(x) \geq \alpha \} \quad (2.16)$$

olarak tanımlanan ${}^{\alpha+}A$ kümesine A 'nın zayıf α - kesmesi denir. Bu nedenle, A 'nın α - kesmesi bazen A 'nın güçlü α - kesmesi olarak da adlandırılır. A bulanık kümesinin güçlü α - kesmesi veya zayıf α - kesmesi, sırasıyla ${}^{\alpha}A$ veya ${}^{\alpha+}A$ deterministik kümeleri olmak üzere, bu kümeler üyelik dereceleri A olan ve belli bir α değerinden büyük veya büyük eşit olan X evrensel kümesinin bütün elemanlarını içerir. Verilen A bulanık kümesinin farklı α - kesmelerle gösterilen, bütün düzeylerinin $\alpha \in (0,1]$ kümesine, A 'nın düzey kümesi denir. Λ , X üzerinde tanımlı A bulanık kümesinin düzey kümesi olmak üzere;

$$\Lambda(A) = \{ \alpha \mid A(x) = \alpha, \exists x \in X \} \quad (2.17)$$

biçiminde gösterilir.

α - kesme ve güçlü α - kesmelerin her ikisi içinde önemli bir özellik, $\alpha \in [0,1]$, A bulanık bir küme ve $\alpha_1, \alpha_2 \in [0,1]$, $\alpha_1 < \alpha_2$ biçiminde farklı iki değer olmak üzere,

$${}^{\alpha_1}A \supseteq {}^{\alpha_2}A \text{ ve } {}^{\alpha_1+}A \supseteq {}^{\alpha_2+}A \quad (2.18)$$

biçiminde gösterilir. Bu özellik

$$\alpha_1 A \cap \alpha_2 A = \alpha_2 A, \alpha_1 A \cup \alpha_2 A = \alpha_1 A$$

ve

$$\alpha_1^+ A \cap \alpha_2^+ A = \alpha_2^+ A, \alpha_1^+ A \cup \alpha_2^+ A = \alpha_1^+ A$$

eşitlikleriyle de ifade edilebilir.

Bu özelliğin önemli bir sonucu olarak, herhangi bir bulanık kümenin bütün α - kesme ve bütün güçlü α - kesmelerinin sınırlandırılmış deterministik kümenin (nested crisp set) iki farklı biçimde gösterimi olduğu apaçık ortadadır (Klir and Yuan 1995).

Eğer kümenin üyelik fonksiyonu sürekli ise, α - kesme ile zayıf α - kesme arasındaki farklılık ortadan kalkar. Zayıf α - kesmeler ile hesaplama yapmak daha kolaydır. Eğer destek kümesi gerçel sayılardan oluşuyorsa ve üyelik fonksiyonu sürekliyse, dışbükey bulanık bir kümenin zayıf α - kesmesi kapalı bir aralıktır (Terano *et al.* 1991, Tuncel 1996).

2.2.2 Normallik

Bir üyelik derecesinin alabileceği en küçük değer 0'dır. Eğer A bulanık kümesinin elemanlarının aldığı en büyük üyelik derecesinin değeri 1 ise, A bulanık kümesi normal olarak adlandırılıp, matematiksel olarak,

$$\forall x \in X : \sup_x \mu_A(x) = 1 \quad (2.19)$$

biçiminde ifade edilir. Bulanık küme normal değil ise, normalaltı olarak adlandırılır. Boş olmayan her normalaltı bulanık küme, üyelik derecelerinin her biri en büyük üyelik derecesine bölünerek normalleştirilebilir (Terano *et al.* 1991, Tuncel 1996).

2.2.3 Dışbükeylik ve içbükeylik

Bu kesimde genel anlamda dışbükeylik ve içbükeyliğe ilişkin açıklamalara yer verilmiş ve bulanık küme kavramı açısından içbükeylik ve dışbükeylik kavramlarına değinilmiştir.

$f : S \rightarrow X_1$ ve S, X_n ' de tanımlı, boş olmayan dışbükey bir küme olmak üzere $\forall x_1, x_2 \in S$ ve $\forall \lambda \in (0,1)$ için

$$f(\lambda x_1 + (1 - \lambda) x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \quad (2.20)$$

koşulu sağlanıyorsa, f fonksiyonu S kümesinde dışbükey bir fonksiyondur. S ' deki her x_1 ve x_2 noktası ve her $\lambda \in (0,1)$ için (2.20) ifadesi ' \leq ' yerine '<' biçiminde kesin bir eşitsizlik ise, f fonksiyonu S kümesinde tanımlı kesin dışbükey bir fonksiyondur. $-f$ fonksiyonu S kümesinde dışbükey (veya kesin dışbükey) bir küme ise, $f : S \rightarrow x_1$ fonksiyonu içbükey (veya kesin içbükey) bir fonksiyondur.

R üzerindeki A bulanık kümesinin dışbükey olabilmesi için gerek ve yeter koşul; X evrensel kümesindeki $\forall x, y \in R$ çifti ve $\forall \lambda \in [0,1]$ için,

$$\mu_A(\lambda x + (1 - \lambda)y) \geq \min(\mu_A(x), \mu_A(y)) \quad (2.21)$$

ifadesinin sağlanmasıdır.

Her $\alpha \in (0,1]$ için α -kesmelerin tümü dışbükey ise, bulanık küme de dışbükeydir. Eğer A bulanık kümesi dışbükey ise, bu kümenin tümleyeni olan A^c bulanık kümesi içbükeydir. A ve B dışbükey kümeler ise, $A \cap B$ kümesi de dışbükeydir. A ve B içbükey kümeler ise, $A \cup B$ kümesi de içbükeydir.

Dışbükey bulanık bir kümenin üyelik fonksiyonları dışbükey olmak zorunda değildir. Aslında, dışbükey bulanık kümelerin üyelik fonksiyonları (standart tanımlara göre) dışbükey değil içbükeydir (Klir and Yuan 1995, Tuncel 1996).

Dışbükey ve içbükey fonksiyonların geometrik yorumu şöyle yapılabilir: x_1 ve x_2 , f alanında iki farklı nokta olmak üzere, $\lambda \in (0,1)$ için $\lambda x_1 + (1-\lambda) x_2$ noktası düşünülün. Burada $f(\lambda x_1 + (1-\lambda) x_2)$, $\lambda x_1 + (1-\lambda) x_2$ noktasının f fonksiyon değerini verirken, $\lambda f(x_1) + (1-\lambda) f(x_2)$, $f(x_1)$ ve $f(x_2)$ değerlerinin ağırlıklı ortalamasını verir. Dolayısıyla bir f dışbükey fonksiyonu için, $\lambda x_1 + (1-\lambda) x_2$ doğru parçası üzerindeki noktalarındaki f değeri, $[x_1, f(x_1)]$ ve $[x_2, f(x_2)]$ noktalarını birleştiren eğrinin yüksekliğinden küçük ya da eşit olur. İçbükey bir fonksiyon için ise, eğrinin yüksekliğinden büyük ya da eşit olur (Tuncel 1996).

2.2.4 Bulanık kümelerde cebirsel işlemler

Bulanık kümelerde cebirsel işlemler, kümelerin üyelik fonksiyonları yardımıyla tanımlanır. $A, B \subset X$ olmak üzere sırasıyla temel işlemler;

$$\text{Eşitlik: } \forall x \in X : \mu_A(x) = \mu_B(x) \Leftrightarrow A = B,$$

$$\text{Kapsama: } \forall x \in X : \mu_A(x) \leq \mu_B(x) \Rightarrow A \subset B,$$

Tümleme: $\forall x \in X : \mu_A(x) = 1 - \mu_B(x) \Rightarrow A = B^c$ veya $B = A^c$ (A^c ve B^c sırasıyla A ve B'nin tümleyenidir.),

Kesişim: $\forall x \in X : \mu_{A \cap B}(x) = \text{Min}(\mu_A(x), \mu_B(x)) = \mu_A(x) \wedge \mu_B(x)$,

Birleşim: $\forall x \in X : \mu_{A \cup B}(x) = \text{Max}(\mu_A(x), \mu_B(x)) = \mu_A(x) \vee \mu_B(x)$,

Çarpım: $\forall x \in X : \mu_{AB}(x) = \mu_A(x) \cdot \mu_B(x)$,

Toplam: $\forall x \in X : \mu_{A \oplus B}(x) = \mu_A(x) + \mu_B(x) - \mu_A(x) \cdot \mu_B(x)$,

biçiminde yapılmaktadır (Tuncel 1996).

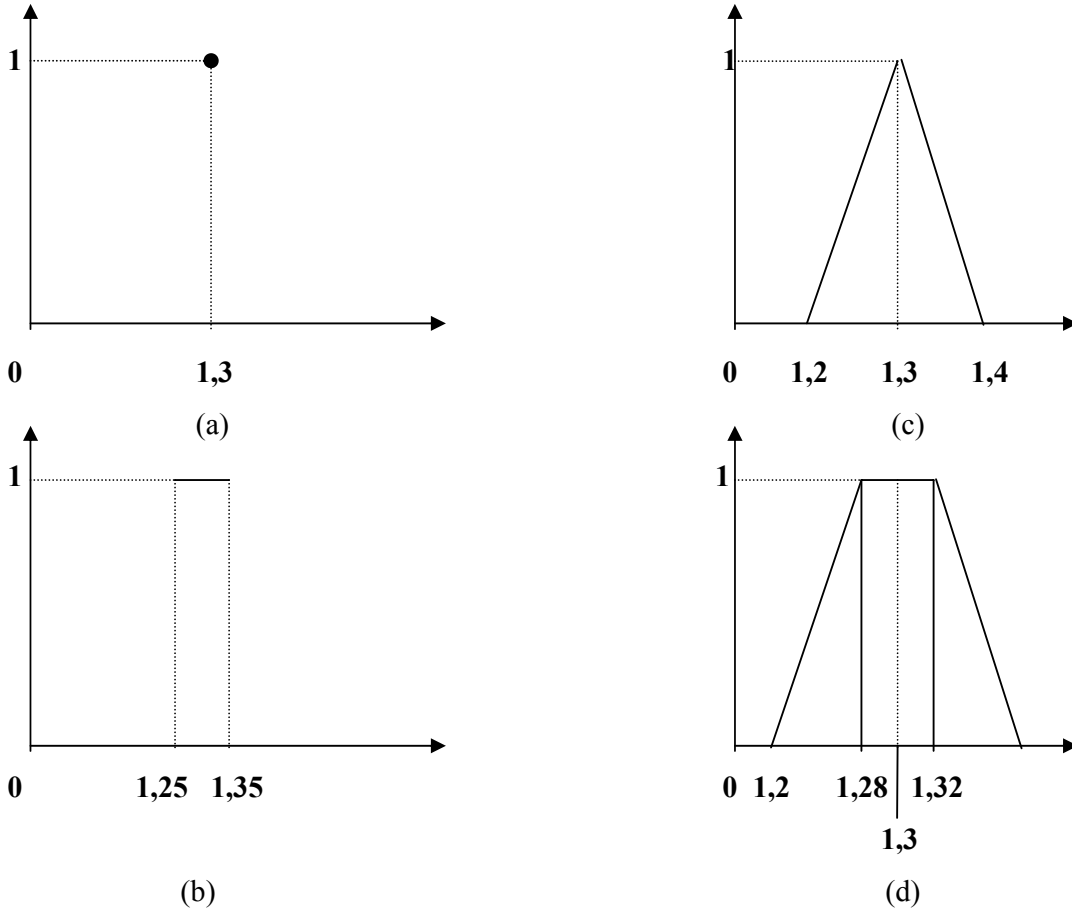
2.2.5 Bulanık sayı kuramı

Normal ve dışbükey bir bulanık kümenin zayıf bir α -kesmesi kapalı bir küme ise, bulanık sayı olarak adlandırılır.

Üyelik fonksiyonu sürekli ise, zayıf α -kesme kapalı bir aralık olur. Ancak kapalı bir aralık için üyelik fonksiyonunun mutlaka sürekli olması gerekmez.

R üzerindeki A bulanık kümesinde bir bulanık sayı tanımlayabilmek için, aşağıdaki üç özelliğin sağlanması gerekir:

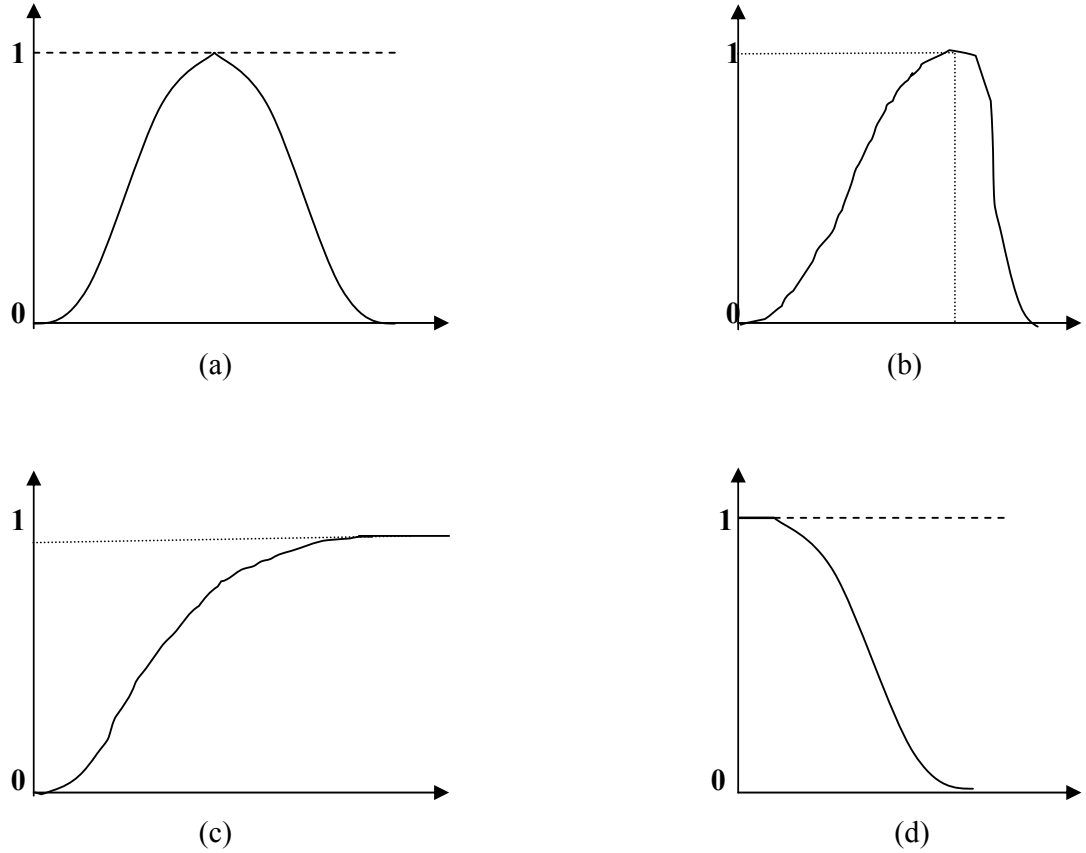
- i) A normal bulanık bir küme olmalıdır.
- ii) α_A her $\alpha \in (0,1]$ için kapalı bir aralık olmalıdır.
- iii) A'nın desteği 0^+ A sınırlı olmalıdır.



Şekil 2.1 Gerçek bir sayı ve deterministik bir aralık ile bulanık bir sayı ve bulanık bir aralığın karşılaştırması

Bulanık küme normal olduğu için; r, R' de bir sayı olmak üzere, küme ' r 'ye yakın gerçek sayıların ' r ' kümesi olacak ve r tarafından tamamen tatmin edileceği için r 'nin üyelik derecesi 1 olacaktır. Bulanık sayının sınırlı desteği ve $\alpha \neq 0$ için bütün α -kesmelerin kapalı aralık olması, kapalı aralıklarda standard aritmetik işlemlere ait terimlerle, bulanık sayılar üzerinde aritmetiksel işlemler, klasik aralık analizleriyle oluşturulabilir. Dolayısıyla herhangi bir bulanık sayının α -kesmesinin her $\alpha \in (0,1]$ için kapalı aralık olması gerekmektedir. Her bulanık sayı, dışbükey bir bulanık bir kümedir. Tersisi her zaman doğru değildir. Bazı dışbükey bulanık kümelerin α -kesmeleri, kapalı veya yarı kapalı aralıklar olabilir.

Bulanık sayıların özel durumları, sıradan gerçek sayıları ve gerçek sayıların aralıklarını içerir. Şekil 2.1’ de: (a), sıradan 1.3 gerçek sayısını; (b), sıradan $[1.25, 1.35]$ kapalı aralığını; (c) 1.3’e yakın bir bulanık sayıyı; (d), düz bölge(bulanık aralık) ile bulanık sayıyı göstermektedir.



Şekil 2.2 Bulanık sayıların temel biçimleri

Şekil 2.1’de gösterilen üçgensel ve yamuksal üyelik fonksiyonları, bulanık sayıları tanımlamakta çok sık kullanılan biçimler olmasına rağmen, farklı uygulamalarda farklı biçimler de tercih edilebilir. Bunun yanısıra, bulanık sayıların üyelik fonksiyonları Şekil 2.1’de olduğu gibi simetrik olmak zorunda değildir. Oldukça tipik bir örnek olan ‘çan şekilli ’ üyelik fonksiyonları adı verilen fonksiyonlara ilişkin biçimler Şekil 2.2 verilmiştir. Şekil 2.2.a’da simetrik; Şekil 2.2.b’de asimetrik, üyelik fonksiyonları gösterilmiştir. Dikkat çekilmesi gereken başka bir nokta da Şekil 2.2.c’de gösterilen sadece artan üyelik fonksiyonları ve Şekil 2.2.d’de gösterilen sadece azalan üyelik

fonksiyonları da, bulanık sayıları tanımlamak için yeterlidir. Herbir uygulamaya özgü olarak bunlardan uygun olanı kullanılacaktır (Klir and Yuan 1995).

Üçgensel ve yamuksal bulanık sayılar

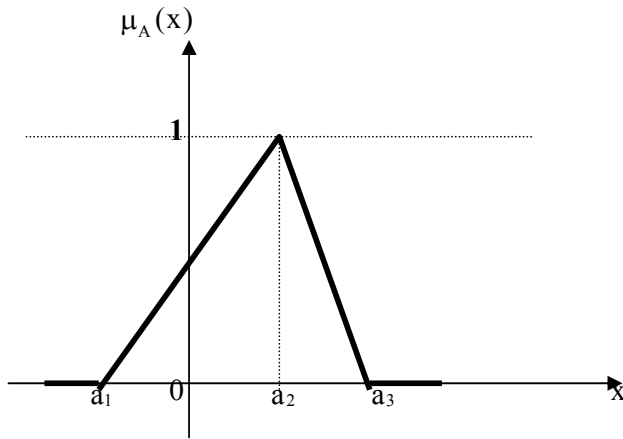
Bulanık sayılar kümesinin eleman sayısı sonsuzdur. Çeşitli bulanık sayı biçimleri olmasına karşın kullanımı en çok tercih edilen üçgensel ve yamuksal üyelik fonksiyonlardır. Özellikle olasılıksal matematiksel programlama problemlerini çözmede bu tip bulanık sayılar kullanılır (Lai and Hwang 1992, Tuncel 1996).

Bu bölümde üçgensel ve yamuksal bulanık sayılar ile temel özellikleri verilecektir.

I. Üçgensel bulanık sayılar

Üçgensel bulanık sayılar, özellikle sistem modelleme de kullanılır (Kauffman and Gupta 1988).

Şekil 2.3' de de görüldüğü gibi üçgensel bulanık sayıların üyelik fonksiyonları normal ve dışbükey fonksiyonlardır.



Şekil 2.3 Üçgensel bulanık sayı

Üçgensel bulanık bir sayı, (a_1, a_2, a_3) gibi bir üçlüyle tanımlanabilir. Üçgensel bulanık bir sayının üyelik fonksiyonu,

$$\mu_A(x) = \begin{cases} 0 & , x < a_1 \\ \frac{x - a_1}{a_2 - a_1} & , a_1 \leq x \leq a_2 \\ \frac{a_3 - x}{a_3 - a_2} & , a_2 \leq x \leq a_3 \\ 0 & , x > a_3 \end{cases} \quad (2.22)$$

biçiminde tanımlanır.

Üçgensel bulanık bir sayı, α -seviyesinde bir güven aralığı ile

$$\forall \alpha \in [0,1] \text{ ve } A = (a_1, a_2, a_3): \quad (2.23)$$

$$A_\alpha = [a_1^{(\alpha)}, a_3^{(\alpha)}] = [(a_2 - a_1)\alpha + a_1, -(a_3 - a_2)\alpha + a_3]$$

şeklinde tanımlanabilir. (Kaufmann and Gupta 1988, Tuncel 1996)

Üçgensel bulanık sayıların bazı önemli cebirsel özellikleri aşağıdaki gibi sıralanabilir:

(a) İki üçgensel bulanık sayının toplanması ya da çıkarılması işlemleri sonucunda yine üçgensel bulanık bir sayı elde edilir.

(b) Üçgensel bulanık sayıların çarpılması, bölünmesi ya da tersinin alınması işlemleri sonucunda her zaman üçgensel bulanık bir sayı elde edilmeyebilir.

(c) Üçgensel bulanık sayıların maksimum ya da minimum işlemleri sonucunda her zaman üçgensel bulanık sayı elde edilmeyebilir.

Cebirsel işlemlerin uygulanması sonucunda üçgensel bulanık sayı elde edilemezse, işlemlerden elde edilen bulanık sayı, üçgensel bulanık sayılara yakınsatılabilir.

Üçgensel bulanık sayıların cebirsel işlemlerinden bazıları aşağıda yer almaktadır:

(a) Toplama:

$$A(+)B=(a_1, a_2, a_3)(+)(b_1, b_2, b_3) = (a_1 + b_1, a_2 + b_2, a_3 + b_3) \quad (2.24)$$

işlemi sonucunda elde edilen üçgensel bulanık bir sayıdır.

(b) Çıkarma:

$$A(-)B=(a_1, a_2, a_3)(-)(b_1, b_2, b_3) = (a_1 - b_3, a_2 - b_2, a_3 - b_1) \quad (2.25)$$

işlemi sonucunda elde edilen üçgensel bulanık bir sayıdır.

(c) Simetri:

$$-A=-(a_1, a_2, a_3) = (-a_3, -a_2, -a_1) \quad (2.26)$$

işlemi sonucunda elde edilen üçgensel bulanık bir sayıdır.

(d) Çarpma, ters, bölme

Çarpma, ters ve bölme işlemlerini yapmak için üçlüler kullanılamaz. Fakat bu hesaplamalar, güven aralığını her α - seviyesi için tanımlayarak yapılabilir. R' de tanımlı üçgensel bulanık sayılar için α - seviyeler öyle ayrıştırılmalıdır ki, α , 0'dan 1'e doğru arttığında pozitif ve negatif değerlerle birlikte minimum ve maksimumu değerlerin hesaplamadaki etkisi incelenebilmelidir (Kaufmann and Gupta 1988, Tuncel 1996).

II. Yamuksal bulanık sayılar

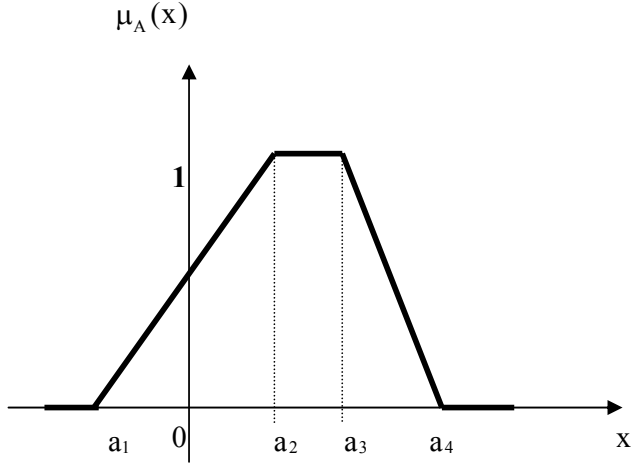
Üçgensel bulanık sayılar, yamuksal bulanık sayıların $a_2 = a_3$ olmak üzere özel bir tipidir. Şekil 2.4.'de görüldüğü gibi $\alpha=1$ durumunda bir nokta değil, (a_2, a_3) aralığında tanımlı bir doğru söz konusudur.

Yamuksal bulanık sayıların cebirsel özellikleri, üçgensel bulanık sayıların cebirsel özelliklerine benzer. Yamuksal bulanık bir sayı, $A = (a_1, a_2, a_3, a_4)$ gibi bir dörtlüyle tanımlanır.

Üyelik fonksiyonu ise,

$$\mu_A(x) = \begin{cases} 0 & , x < a_1 \\ \frac{x - a_1}{a_2 - a_1} & , a_1 \leq x \leq a_2 \\ 1 & , a_2 \leq x \leq a_3 \\ \frac{a_4 - x}{a_4 - a_3} & , a_3 \leq x \leq a_4 \\ 0 & , x > a_4 \end{cases} \quad (2.27)$$

biçimindedir.



Şekil 2.4 Yamuksal bulanık sayı

Yamuksal bulanık bir sayı, α -seviyesinde bir güven aralığı ile

$$\forall \alpha \in [0,1] \text{ ve } A = (a_1, a_2, a_3, a_4): \quad (2.28)$$

$$A_\alpha = [a_1^{(\alpha)}, a_4^{(\alpha)}] = [(a_2 - a_1)\alpha + a_1, -(a_4 - a_3)\alpha + a_4]$$

biçiminde tanımlanabilir (Kaufmann and Gupta 1988, Tuncel 1996).

Üçgensel bulanık sayıların cebirsel işlemleriyle ilgili tüm sonuçlar yamuksal bulanık sayılar için de geçerlidir.

Yamuksal bulanık sayıların cebirsel işlemlerinden bazıları aşağıda tanımlanmıştır:

(a) Toplama:

$$A(+)B=(a_1, a_2, a_3, a_4)(+)(b_1, b_2, b_3, b_4)=(a_1 + b_1, a_2 + b_2, a_3 + b_3, a_4 + b_4) \quad (2.29)$$

işlemi sonucunda elde edilen yamuksal bulanık bir sayıdır.

(b) Çıkarma:

$$A(-)B=(a_1, a_2, a_3, a_4)(-)(b_1, b_2, b_3, b_4)=(a_1 - b_4, a_2 - b_3, a_3 - b_2, a_4 - b_1) \quad (2.30)$$

işlemi sonucunda elde edilen yamuksal bulanık bir sayıdır.

(c) Simetri:

$$-A=- (a_1, a_2, a_3, a_4) = (-a_4, -a_3, -a_2, -a_1) \quad (2.31)$$

işlemi sonucunda elde edilen gene yamuksal bulanık bir sayıdır.

(d) Çarpma, ters, bölme

Çarpma, ters ve bölme işlemlerini yapmak için dörtlüler kullanılamaz. Fakat bu hesaplamalar, güven aralığını her α - seviyesi için tanımlayarak yapılabilir (Kaufmann and Gupta 1988, Tuncel 1996).

2.2.6 Üyelik fonksiyonları

Erken Uyarı Sistemi'nin modellendiği Bölüm 6'da, manipülatif işlemlerin şüpheli işlem kümelerine aidiyetleri, Bulanık Küme Kuramı'ndan yararlanarak belirlenmektedir. Bu noktada, aidiyet derecelerini belirlemek için üyelik fonksiyonlarından faydalanılmıştır.

Bulanık küme kuramında belirsizlik, kesin olmama ve subjektiflik içeren durumlar incelenir. Kesin küme kuramına belirsizlik dahil edildiğinde ortaya çıkan bulanık küme kuramı, orijinal klasik küme kuramından daha esneklerdir.

Bulanık küme kuramının iki önemli özelliği vardır. Bu özellikler aşağıdaki gibi özetlenebilir:

- (a) Bulanık kümelerin ve işlemcilerin üyelik fonksiyonları, bulanık küme kuramında çok önemli bir rol oynar.
- (b) Bulanık küme kuramı aslında çok genel, esnek ve kurallı bir kuramdır. Gerçek bir probleme uygulanırken çok dikkatli olmak gerekir. Çünkü ne üyelik fonksiyonu ne de işlemci kavramının tek bir anlamsal yorumu vardır. Problemin içeriğine bağlı olarak değişebilen anlamsal yorum için farklı matematiksel tanımlamalar ve işlemciler gerekir. Bu nedenle üyelik fonksiyonları ve işlemciler bulanık küme kuramının temel taşlarıdır (Tuncel 1996).

Üyelik fonksiyonu, X evrensel kümesine ait bir x ögesinin A altkümeye ait olma derecesini veren bir fonksiyondur.

Kesin ve bulanık kümelerin üyelik fonksiyonları, eşitlik (2.9) ve (2.13)'e eşdeğer olarak sırasıyla

$$\forall x \in X : \mu_A(x) \in \{0,1\} \quad (2.32)$$

ve

$$\forall x \in X : \mu_A(x) \in [0,1] \quad (2.33)$$

biçiminde gösterilirler. Görüldüğü gibi kesin kümeler ile bulanık kümeler arasındaki en önemli farklılık, üyelik fonksiyonlarının aldığı değerlerden kaynaklanmaktadır. Kesin bir kümenin üyeleri kesin olarak bilinirken, bulanık bir kümenin üyesi olmaya aday öğeler ve bu öğelerin üyelik dereceleri kesin olarak bilinmemektedir.

Bazı kaynaklarda ‘üyelik fonksiyonu’ deyimini yerine ‘karakteristik fonksiyon’ deyimini kullanılmaktadır.

2.2.7 Üyelik fonksiyonu biçimleri

Bu bölümde üyelik fonksiyonlarının fonksiyonel biçimleri üzerinde durulacaktır. Hem tercihe dayalı üyelik fonksiyonları hem de olanak dağılımları incelenecektir. Bu inceleme, bulanık matematiksel programlama problemlerine yön verebilecek niteliktedir.

Üyelik fonksiyonları, aşağıda verilen dört ana grupta ele alınabilir (Dombi 1990, Tuncel 1996).

I. Deneysel karar vermeye dayalı üyelik fonksiyonları:

(a) Zadeh'in unimodal fonksiyonları;

$$\mu_{\text{genç}}(x) = \begin{cases} 1/\{1 + [(x-25)/5]^2\} & x > 25 \\ 1 & , x \leq 25 \end{cases} \quad (2.34)$$

ve

$$\mu_{\text{ihtiyar}}(x) = \begin{cases} 1/\{1 + [(x-50)/5]^2\} & x > 50 \\ 1 & , x \leq 50 \end{cases} \quad (2.35)$$

biçimindedir.

(b) Dimitru ve Luban'ın kuvvet fonksiyonları;

$$\mu(x) = x^2/a^2 + 1 \quad , x \in [0, a] \quad (2.36)$$

ve

$$\mu(x) = -x^2/a^2 - 2x/a + 1 \quad , x \in [0, a] \quad (2.37)$$

biçimindedir.

(c) Svarowski'nin sinüs fonksiyonu;

$$\mu(x) = 0.5 + 0.5[\sin\{\pi/(b-a)[x - (a+b)/2]\}] \quad , x \in [a, b] \quad (2.38)$$

biçimindedir.

II. Güvenilirlik kavramına dayalı üyelik fonksiyonları

(a) Zimmermann'ın doğrusal fonksiyonu;

$$\mu(x) = 1 - x/a, \quad x \in [0, a] \quad (2.39)$$

biçimindedir.

(b) Tanaka, Uejima ve Asai'nin simetrik üçgensel fonksiyonu;

$$\mu_x(x) = \begin{cases} 1 - |b - x|/a, & b - a \leq x \leq b + a \\ 0 & \text{ö.d.} \end{cases} \quad (2.40)$$

biçimindedir.

(c) Hannan'ın parçalı doğrusal fonksiyonu;

$$\mu(x) = \sum_j \alpha_j |x - a_j| + \beta x + r, \quad j = 1, 2, \dots, N$$

$$\alpha_j = (t_{j+1} - t_j)/2 \quad (2.41)$$

$$\beta_j = (t_{N+1} + t_1)/2$$

$$r = (s_{N+1} + s_1)/2$$

biçiminde olmak üzere her $a_{i-1} \leq x \leq a_i$ aralığı için $\mu(x) = t_i x + s_i$ 'dir. t_i , eğim ve s_i , a_{i-1} 'de başlayıp a_i 'de biten eğim bölgesi için y-bileşenidir.

(d) Leberling'in hiperbolik fonksiyonu;

$$\mu(x) = 0.5 + 0.5 \tanh(a(x - b)) , -\infty \leq x \leq +\infty \quad (2.42)$$

biçiminde olmak üzere a bir parametredir.

(e) Sakawa ve Yumine'nin üstel ve ters hiperbolik fonksiyonları;

$$\mu(x) = c(1 - e^{(b-x)(b-a)}), x \in [a, b] \quad (2.43)$$

ve

$$\mu(x) = 0.5 + c \tan h^{-1}(d(x - b)) \quad (2.44)$$

biçiminde olmak üzere c ve d parametrelerdir.

(f) Dimitru ve Luban'ın fonksiyonu;

$$\mu(x) = 1/(1 + x/a) \quad (2.45)$$

biçiminde olmak üzere a bir parametredir.

(g) Dubois ve Prade'in L-R bulanık sayısı;

$$\mu(x) = \begin{cases} L((a-x)/\alpha), & x < a \\ R((x-b)/\beta), & x > b \\ 1 & , a \leq x \leq b \end{cases} \quad (2.46)$$

biçiminde olmak üzere L(.) ve R(.) referans fonksiyonlarıdır.

III. Teorik isteğe dayalı üyelik fonksiyonları:

(a) Civanlar ve Trussel'in fonksiyonu;

$$\mu(x) = \begin{cases} ap(x) & , ap(x) \leq 1 \\ 0 & , \text{ö.d.} \end{cases} \quad (2.47)$$

biçiminde olmak üzere $a \in [0,1]$ bir parametre ve $p(x)$ olasılık yoğunluk fonksiyonudur.

(b) Savarovski'nin fonksiyonu;

$$\mu(x) = \begin{cases} 0 & , x < a \\ K(x-a)^2 & , a \leq x \leq b \\ K_2 x^2 + K_1 x + K_0 & , b < x \leq c \\ 1 & , x > c \end{cases} \quad (2.48)$$

biçiminde olmak üzere K, K_0, K_1 ve K_2 parametrelerdir.

IV. Kişilere özgü kavramlar için bir model oluşturan üyelik fonksiyonları:

(a) Hersh ve Caramaza'nın fonksiyonu;

$$\mu(x) = 0.5 + d(r/10) \quad (2.49)$$

biçiminde olmak üzere 'evet' cevapları için $d(x)=1$ ve 'hayır' cevapları için $d(x)=-1$ olmak üzere r bir güven değeridir.

(b) Zimmermann ve Zysno'nun fonksiyonu;

$$\mu(x) = 0.5 + (1/d) \left[1 / (1 + e^{-a(x-b)}) - c \right] \quad (2.50)$$

(c) Dombi'nin fonksiyonu

$$\mu(x) = (1-s) x^2 / \left\{ (1-s) x^2 + s(1-x)^2 \right\} \quad (2.51)$$

biçiminde olmak üzere s , biçimin karakteristik değeri; $y=\mu(x)$ ve $y=x$ değerlerinin kesişimidir.

Daha önce de belirtildiği gibi üyelik fonksiyonları, tercihe dayalı üyelik fonksiyonları ve olanak dağılımları olmak üzere iki gruba ayrılabilir. Tercihe dayalı bir üyelik fonksiyonu, tercih bilgisini karar vericiden alarak oluşturabilir. Diğer yandan olasılık dağılımının bazı yönlerden aynısı olan olanak dağılımı, olayların olası ortaya çıkışları düşünülmektedir (Lai and Hwang 1992, Tuncel 1996).

Bulanık ve olanaksal (possibilistic) matematiksel programlama problemlerinde (A, b, ve c) girdisinin bulanıklığını ve belirsizliğini modellemek için tercihe dayalı üyelik

fonksiyonlarının ve olanak dağılımlarının gerçekçi biçimde verildiği ya da oluşturulduğu varsayılır. Bu nedenle, tercihe dayalı üyelik fonksiyonlarının ya da olanak dağılımlarının verilmiş ya da oluşturulmuş biçimi çok önemlidir.

Üyelik fonksiyonlarını oluşturmada kullanılan iki yaklaşım mevcuttur; belitsel ve anlamsal yaklaşım (Giles 1988). Yarar kuramındaki yaklaşımlara benzeyen belitsel yaklaşım, matematiksel yapı üzerinde odaklanmıştır. Anlamsal yaklaşım ise, matematiksel yapı üzerinde değil terimlerin pratik yorumu üzerinde odaklanmıştır (Lai and Hwang 1992, Tuncel 1996).

Tercihe dayalı üyelik fonksiyonlarını ve olanak dağılımlarını oluşturmak için geliştirilmiş birçok yaklaşım mevcuttur. Uzaklık yaklaşımı, doğru değerlendirilmiş yaklaşım ve hesap yaklaşımı bu yaklaşımlardan sadece birkaçıdır (Lai and Hwang 1992, Tuncel 1996).

3. VERİ MADENCİLİĞİ

Tezin özgün uygulama kısmı olan Erken Uyarı Sistemi'nin temel metodolojisini oluşturan Veri Madenciliği (VM) bu bölümde tartışılmaktadır. VM'nin bir parçası olduğu Veri Tabanlarında Bilgi Keşfi Süreci'nin (VTBK) tanımı ve bileşenleri, VM'nin temel tanım ve kavramları, Bölüm 6'da modelleme kısmında kullanılan metodolojik yaklaşımın temelleri ve veri madenciliğine istatistiksel perspektiften bakışı sağlayan İstatistiksel Öğrenme'nin tanımı bu bölümde verildiği gibi; modellemede kullanılan veya kullanılmaya elverişli yöntemler de gene bu bölümde işlenmektedir.

3.1 Veri Madenciliği ve Veri Tabanlarında Bilgi Keşfi

Veri madenciliğinden bahsetmeden önce aktarılması gereken temel bir kavram Veri Tabanlarında Bilgi Keşfi (VTBK)'dir. Yaygın olarak KDD (Knowledge Discovery in Databases) kısaltmasıyla bilinen VTBK ve Veri Madenciliği (VM) ile aralarındaki ilişki aşağıda başlıklar halinde incelenmektedir.

3.1.1 Veri tabanlarında bilgi keşfinin tanımı

Tarihsel olarak, veri içerisinde faydalı örüntüleri bulma kavramına veri madenciliği, bilgi aktarımı, enformasyon keşfi, enformasyon hasadı, veri arkeolojisi ve veri örüntü işleme gibi pek çok farklı isim verilmiştir. Veri madenciliği terimi daha çok istatistikçiler, veri analistleri ve yönetim bilgi sistemleri toplulukları tarafından kullanılmaktadır. Aynı zamanda, veri tabanı alanında da popülerliğe ulaşmıştır. 'Veri tabanlarında bilgi keşfi' ifadesi 1989 yılında ilk KDD çalışma toplantısında ortaya atılmış (Piatetsky-Shapiro 1991) ve bilginin veri keşfi sürecinin nihai ürünü olduğuna vurgu yapılmak istenmiştir. Yapay zeka ve makine öğrenimi alanlarında popülerleşmiştir.

Yukarıdaki perspektiften bakıldığında, Veri Tabanlarında Bilgi Keşfi, veriden faydalı bilginin keşfedilmesi sürecinin tamamına atıfta bulunmakta ve veri madenciliği bu sürecin bir adımına karşılık gelmektedir. Veri madenciliği, veriden örüntülerin aktarımı için özel algoritmaların uygulanmasıdır.

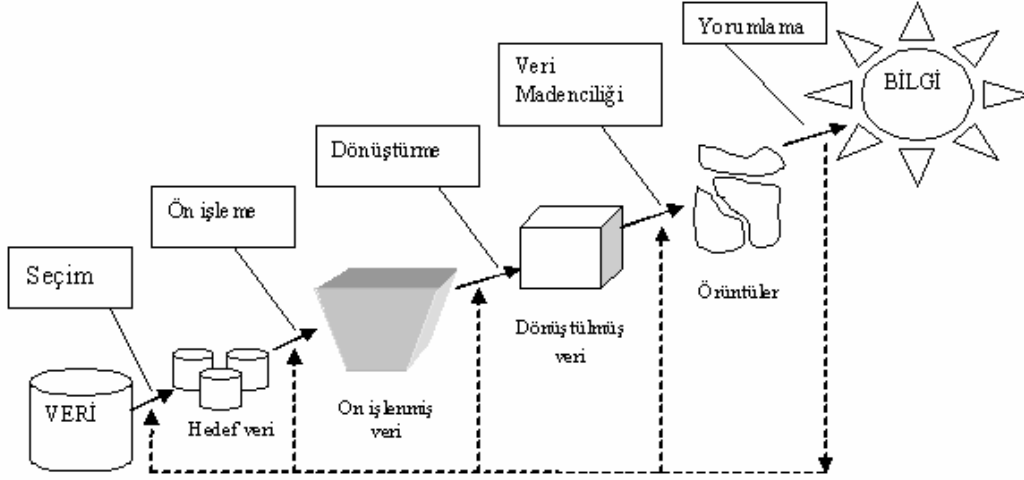
Veri tabanlarında bilgi keşfi (VTBK) makine öğrenimi, örüntü tanıma, veri tabanları, istatistik, yapay zeka, uzman sistemler, veri görselleştirme ve yüksek performanslı hesaplama gibi araştırma alanlarının kesişimi olarak gelişmiş ve gelişimine devam etmektedir. Tek hedef, büyük veri setleri kapsamında, düşük düzeyde veriden, yüksek düzeyde bilgi aktarmaktır.

VTBK'nin veri madenciliği bileşeni, VTBK'nin veri madenciliği sürecinde veri içerisinde örüntüleri bulmada ağırlıklı olarak makine öğrenimi, istatistik ve örüntü tanıma gibi bilinen tekniklere güvenmektedir. Bu konuda doğal bir soru 'VTBK, örüntü tanıma veya makine öğrenimi veya ilgili alanlardan nasıl farklı olmaktadır?' olabilir. Cevap ise, bu alanların, VTBK'nın VM adımında bazı VM metodlarını sağlamalarıdır. Oysa VTBK, verinin nasıl depolanıp erişileceğinden, algoritmaların devasa veri setlerine nasıl ölçeklenebileceğine ve hala etkin olarak çalışmalarına, sonuçların nasıl yorumlanabileceği ve görselleştirilebileceğine ve bütün insan-makine interaksyonunun kullanışlı olarak nasıl modellenip, desteklenebileceğine olmak üzere veriden bilginin keşfinin tüm süreçleri üzerine odaklanır. VTBK süreci, örneğin makine öğrenimi gibi herhangi bir tekniğin ilgi alanı içerisinde yer almanın ötesinde çok disiplinli bir faaliyet olarak görülmelidir. Bu kapsamda, (makine öğreniminin yanısıra) yapay zekanın diğer alanları için de, VTBK'ye katkı sağlayacak açık fırsatlar vardır (Fayyad *et al.* 1996).

3.1.2 Veri tabanlarında bilgi keşfi süreci

VTBK süreci, veritabanlarını kullanarak veritabanlarında istenilen seçim, ön işleme, alt örnekleme, dönüşüm, örüntülerin açığa çıkarılması için veri madenciliği yöntemlerinin (algoritmalarının) uygulanması ve açığa çıkarılan örüntülerin tanımlanması için veri madenciliği ürünlerinin yorumlanmasını ihtiva eder. VTBK sürecinin, VM bileşeni,

veriden hangi örüntülerin aktarılıp, dikkate alınacağıın algoritmik anlamda ifadesi olarak değerlendirilmelidir. VTBK sürecinin bütünü, Şekil 3.1’de de görüldüğü gibi, değerlendirme ve madenlenmiş örüntülerin hangilerinin yeni bilgi olarak değerlendirileceğinin olası yorumunu da içerir.



Şekil 3.1 VTBK sürecinin adımları

VTBK süreci interaktif ve yinelemeli, kullanıcı tarafından kararların verilmesini gerektiren adımlardan oluşmaktadır. Brachman and Anand (1996), sürecin interaktif yapısına vurgu yapan pratik bir görünüm vermişlerdir. Sürecin bazı temel adımlarının çerçevesi aşağıda verilmiştir.

1. Adım: Uygulama alanı ve ilgili önsel bilgi ile ilgili bir anlayış geliştirmek ve müşterinin bakış açısından VTBK sürecinin hedefini tanımlamak.
2. Adım: Hedef veri kümesini yaratmak: Keşfin uygulanacağı veri kümesini seçmek veya değişkenlerin bir alt kümesi veya veri örnekleri üzerine odaklanmaktır.
3. Adım: Veri temizleme ve ön işleme: Eğer uygunsuz gürültünün kaldırılması, model için gerekli enformasyonun toplanması, kayıp veri alanları için stratejilere karar vermeyi içeren temel operasyonlardır.

4. Adım: Veri indirgeme ve projeksiyon: Görev hedefine bağlı veriyi temsil edecek faydalı özellikleri bulmaktır. Boyut indirgeme veya dönüşüm yöntemleriyle göz önüne alınan değişken sayısı indirgenebilir veya verinin değişmez (invariant) temsili bulunabilir.
5. Adım: VTBK sürecinin hedefleri ile (1. Adımdaki), veri madenciliği yönteminin eşleştirilmesi: Özetleme, sınıflandırma, regresyon, kümeleme vb. yöntemler uygulanmaktadır.
6. Adım: Veri madenciliği algoritma(larının)sının seçimi: Açıklayıcı analizler, model ve hipotez seçimi: Tercih edilen VM algoritmaları ve seçilen yöntemler veri örüntülerini araştırmak için kullanılır. Bu süreç, hangi modelin ve parametrelerin uygun olabileceğine ve VM yönteminin VTBK sürecinin bütün kriterleriyle eşleşip eşleşmediğine karar verilmesini içermektedir.
7. Adım: Veri Madenciliği: Özel bir temsili form veya temsili küme içerisinde ilgilenilen örüntüler; sınıflandırma kuralları ve ağaçları, regresyon ve kümelemeyi içererek araştırılır.
8. Adım: Madenden çıkarılan örüntülerin yorumlanması: Sonraki iterasyonlarda, Adım 1-7'den herhangi birine dönülmesi ihtimaliyle madenden çıkarılan örüntüler yorumlanır.
9. Adım: Keşfedilen bilgilerin birleştirilmesi: Keşfedilen bilgi sonraki çalışmalar için bir başka sistem altında toplanabilir veya basitçe dokümanite edilip, raporlanarak ilgili birimlere iletilir. Bu aynı zamanda, önceden inanılan veya aktarılan bilgilerin doğruluğunu kontrol etme ve olası farklılıkların ayrıştırılmasını da içerir (Fayyad *et al.* 1996, Zaine 1999).

3.1.3 Veri tabanlarında bilgi keşfi sürecinin veri madenciliği adımı

Bilgi keşfi hedefleri, sistemin kullanım amacına göre tanımlanır. Hedefleri ikiye ayırabiliriz:

1. Doğrulama
2. Keşif

Doğrulama ile, sistem kullanıcının hipotezlerini doğrulamak ile sınırlıdır. Keşif ile, sistem bağımsız olarak yeni örüntüler bulur. Keşif hedefi, bazı varlıkların gelecekteki davranışlarını tahmin etmek için sistemin örüntüleri bulmasında kullanıldığında tahmin ve kullanıcıya sunumda insanın anlayabileceği bir form için sistem kullanıldığında tanımlama olmak üzere iki alt gruba ayrılmaktadır.

Veri madenciliği gözlenen veriye model uydurmayı veya gözlenen verideki örüntüleri tanımlamayı gerektirmektedir. Model uydurma, bilgi çıkarımı rolünü üstlenmektedir: Modelin, kullanışlı veya ilginç keşifsel bilgiye işaret edip etmediği, tamamıyla interaktif VTBK sürecinin subjektif insan yargısına tipik olarak ihtiyaç duyduğu bir parçasıdır. Model uydurmada istatistiksel ve mantıksal olmak üzere iki temel matematiksel yapı kullanılmaktadır. Modelde, istatistiksel yaklaşım deterministik olmayan etkiye, mantıksal yaklaşım ise deterministik etkiye izin vermektedir.

Pek çok VM yöntemi, makine öğrenimi, örüntü tanıma ve istatistikten denenmiş ve test edilmiş teknikleri temel almıştır: Sınıflandırma, kümeleme, regresyon vb. (Fayyad *et al.* 1996).

3.2 Veri Ambarı

Veri Ambarları, Veri Madenciliği ile eşanlı olarak anılan ve Veri Madenciliği sürecinin gerçekleştirildiği veriyi sağlayan özel bir veri tabanıdır. Tanım olarak Veri Ambarı, pekçok farklı kaynaktan ve genellikle de farklı yapıda verinin depolandığı ve hepsinin de aynı birleşik çatı altında kullanılmasının ümit edildiği yapılardır. Ayrıca, Veri Ambarı pek çok farklı kaynaktan elde edilen veriyi aynı çatı altında analiz etme imkanı sunar (Fayyad *et al.* 1996).

Veritabanlarından geliştirilen ilgili bir alan da, işlem verilerini toplamak ve online analiz ve karar destek için uygun hale getirmek için temizlemenin adı olan ve popüler iş trendi olarak atıfta bulunan veri ambarlarıdır. Veri ambarcılığı, veri kümelerine VTBK aşaması için iki önemli yoldan yardımcı olur:

1. Veri temizleme
2. Veri erişimi (Cabena *et al.* 1997).

3.2.1 Veri temizleme

Organizasyonlar sahip oldukları geniş kapsamlı veri ve veri tabanlarını birleşik mantıksal görünümde olduğunu düşünmeye zorladıklarından, haritalanmış verinin sonuçlarını bir tek isimlendirme eğiliminde olmak, kayıp veriyi düzgün temsil etme ve ele alma ve gürültü ve hataları adres göstermek zorundadırlar (Cabena *et al.* 1997).

3.2.2 Veri erişimi

Veriye erişim ve tarihi açıdan elde edilmesi zor olan veriye erişim yolları sağlayacak, düzgün ve iyi tanımlanmış yöntemler yaratılmalıdır (Cabena *et al.* 1997).

Öncelikle organizasyon ve bireyler verilerini nasıl depolayacaklarını belirlemeli ve verilerine erişim problemini çözmelidirler. Doğal olarak bir sonraki adım ‘ Bütün veri ile ne yapacağız ?’ sorusudur. Bu soru, doğal olarak VTBK fırsatını ortaya çıkarmaktadır.

Veri ambarlarını analiz etmenin popüler yaklaşımlarından birisi Online Analytical Processing (OLAP)’tır. OLAP araçları, birçok boyutta hesaplama özetleri ve tanımlamalarında SQL’den üstün olan, çok boyutlu veri analizi sağlamaya odaklanmışlardır. OLAP araçları interaktif veri analizi sağlama ve basitleştirmeyi hedeflemişlerdir. Ama VTBK araçlarının hedefi, süreci mümkün olduğunca otomatikleştirmedir (Cabena *et al.* 1997).

3.3 Veri Madenciliğinin Tanımı

Veritabanlarında bilgi keşfi, sıklıkla, büyük hacimde veri koleksiyonundan faydalı bilgiyi keşfetmeyi hedefleyen, veri madenciliği olarak anılmaktadır. Veritabanları günümüzde terabaytlarla ifade edilmektedir. Bu büyük hacimde verinin içinde stratejik önem taşıyan gizli enformasyon yatmaktadır. Ama bu kadar çok ağaç olduğunda, ormanın tümü hakkında anlamlı sonuçlar nasıl çıkarılır?

Yukarıdaki soruya en yeni yanıt, hem geliri artırırken hem de maliyetleri indirgeyen veri madenciliğidir.

Veri madenciliği, pek çok analiz aracı kullanımıyla veri içerisinde örüntü ve ilişkileri keşfederek, bunları geçerli tahminler yapmak için kullanan bir süreçtir (TWO CROWS 1999).

Veri madenciliği, en basit tanımıyla, veri tabanlarındaki ilişkili örüntüleri otomatik olarak belirlemedir. Veri madenciliği sihir değildir. Yıllardır, istatistikçiler veri tabanlarını elle kazımakta, istatistiksel açıdan önemli ilişkiler aramaktadır.

Veri madenciliği müşteri davranışlarını tahmin etmek için model kurmak amacıyla, iyi tanımlanmış istatistiksel ve makine öğrenimi yöntemlerini kullanmaktadır.

Veri madenciliği veri kümesi içerisinde keşfedilmemiş örüntüleri bulmayı hedefleyen teknikler koleksiyonunu betimlemektedir. Veri madenciliğinin amacı, geçmiş faaliyetlerin analizini temel alarak gelecekteki davranışların tahminine yönelik karar-verme modelleri yaratmaktır. Veri madenciliği, William Frawley ve Gregory Piatetsky-Shapiro (1991) tarafından, ‘ ... verideki gizli, önceden bilinmeyen ve potansiyel olarak faydalı enformasyonun önemsiz olmayanlarının açığa çıkarılması...’ biçiminde yapılan bilgi keşfi tanımını destekler (Berson *et al.* 1999).

3.4 İstatistiksel Öğrenme

Veri Madenciliği genel anlamda incelendiğinde, sürecin işleyişi, içeriği ve bileşenleri bakımından; İstatistik ve Bilişim olmak üzere temel anlamda iki bakış açısına sahip olmak mümkündür. Bu tez, veri madenciliğine istatistiksel açıdan yaklaşmakta ve veri madenciliğine istatistiksel bakış İstatistiksel Öğrenme olarak adlandırılmaktadır. Tezin özgün uygulaması olan Erken Uyarı Sistemi'nde de İstatistiksel Öğrenme temel alınmaktadır.

İstatistik alanı bilim ve sanayideki problemlere kapı açmak için sürekli olarak onlara meydan okumaktadır. İlk zamanlarda bu problemler sık sık ziraat ve sanayi deneylerinden gelmekte ve görece olarak dar kapsamlı olmaktadır. Bilgisayarların ve bilgi çağının gelişimiyle istatistiksel problemler hem boyut hem de karmaşıklık açısından patlamıştır. Veri depolama üniteleri, organizasyon ve araştırmadaki gelişmeler yeni bir alan olan veri madenciliğine işaret etmiştir; biyolojideki istatistiksel ve hesaplama problemleri ve ilaç biyoenformatiği yaratmıştır. Pek çok alanda hala çok büyük miktarda veri üretilmekte ve istatistikçilerin işi bunların tamamı hakkında akıl yürütmektir: önemli örüntü ve eğilimleri açığa çıkartmak ve 'verinin ne söylediğini' anlamak. Buna veriden öğrenme denilmektedir.

Veriden öğrenmedeki gelişmeler, istatistiksel bilimlerde bir devrime işaret etmiştir. Hesaplamanın böyle anahtar bir rol oynamasından beri, araştırmacıların bilgisayar bilimleri ve mühendislik gibi diğer alanlarda da bu yeni gelişmeleri gerçekleştirmesi çok fazla sürpriz olmamıştır.

Öğrenme problemleri pürüzlü olarak denetimli ve denetimsiz olarak ikiye ayrılabilir. Denetimli öğrenimde hedef, girdi ölçülerinin sayısını temel alarak çıktı ölçüsünün değerini tahmin etmektir; denetimsiz öğrenimde ise çıktı ölçüsü yoktur ve hedef girdi ölçüleri kümesi arasındaki birliktelik ve örüntüleri betimlemektir.

İstatistiksel öğrenme bilim, finans ve sanayinin pek çok alanında anahtar rol oynamaktadır. Öğrenme problemlerine bazı örnekler;

- Bir hastanın kalp krizi geçirdiğinde ikinci krizi geçirmeden hastaneye yetiştirilip yetiştirilemeyeceğinin tahmini. Tahmin, bu hasta için demografik, diyet ve klinik ölçüleri temel alacaktır.
- Şirket performans ölçülerini ve ekonomik verileri temel alarak 6 ay sonraki stok miktarlarını tahmin etmek.
- Manyetik görüntüden el yazısı posta kodlarındaki rakamları tanımlamak.
- Diabetik bir hastanın kanındaki kızılötesi soğurmadan kanındaki glukoz miktarını tahmin etmek.
- Klinik ve demografik değişkenleri temel alarak prostat kanseri için risk faktörlerini tanımlamak.

olarak verilebilir.

Öğrenme bilimi istatistik, veri madenciliği ve yapay zeka, mühendislik alanı ve diğer disiplinlerin kesişiminde anahtar rol oynamaktadır.

Veriden öğrenim dikkate alındığında, tipik senaryo; genellikle nicel (stok fiyatı) veya kategorik (kalp krizi veya değil) çıktı ölçümü vardır ve içerik kümesi (diyet ve klinik ölçümler gibi) temel alınarak tahmin edilmek istenir. Nesne kümeleri için çıktı ve içerik ölçümlerinin gözlemlerinde verinin eğitim kümesi vardır. Bu verinin kullanımıyla tahmin veya öğrenme modeli kurulur, böylece yeni gözlenmemiş çıktıları tahmin etme imkanı tanır. İyi bir öğrenici böyle bir çıktıyı başarıyla tahmin eder.

Yukarıda yapılan tanımlama denetimli öğrenim problemini tasvir etmektedir. Öğrenme sürecine kılavuzluk edecek bir çıktı değişkeninin varlığı nedeniyle denetimli denilmektedir. Denetimsiz öğrenim probleminde ise sadece içerik gözlenir ve çıktı ölçümü yoktur. Görev sadece verinin nasıl organize edildiğini veya kümelendiğini betimlemektir (Hastie *et al.* 2001).

3.5 Veri Madenciliği İle Diğer Analitik Yöntemlerin Karşılaştırılması

Veri madenciliği ile yeni tanışanların özellikle de veri tabanı pazarlaması, geleneksel veri analizi ve istatistik alanında çalışmış olanların veri madenciliği ile diğer analitik yöntemler arasındaki farkın ne olduğunu sıkça sormaları muhtemeldir.

Veri madenciliği sıkça aşağıdakiler gibi düşünülmektedir:

- Büyük bir veri ambarı üzerinde SQL (Structured Query Language) sorgusu,
- Herhangi bir sayıda veritabanı veya veri ambarları üzerinde SQL sorgusu,
- İleri düzeyde enformasyon erişimi, örneğin akıllı ajanlar yoluyla,
- Çok boyutlu veritabanı analizi (Multidimensional Database Analysis-MDA),
- OLAP,
- Açıklayıcı veri analizi,
- İleri grafiksel görselleştirme,
- Veri ambarı üzerinde geleneksel istatistiksel işleme.

Bu yaklaşımların hiçbiri veri madenciliği değildir. Çünkü, her birindeki temel eksiklik, bilginin keşfinin önceden önerilmiş bir hipotez olmaksızın gerçekleştirilmesidir (Cabena *et al.* 1997).

3.5.1 İstatistiksel analiz ile veri madenciliğinin karşılaştırılması

İstatistiksel analiz ve veri madenciliğinin karşılaştırması ve farklılaştığı noktalar aşağıda yer almaktadır (Moss and Atre 2003).

Çizelge 3.1 İstatistiksel analiz ve veri madenciliğinin karşılaştırılması

<i>İstatistiksel Analiz</i>	<i>Veri Madenciliği</i>
<ul style="list-style-type: none">• İstatistikçiler genellikle bir hipotez ile başlarlar.• İstatistikçiler hipotezlerini eşleştirmek için kendi eşitliklerini geliştirmek zorundadırlar.• İstatistiksel analizler niceliksel ve niteliksel verileri kullanır.• İstatistikçiler kirli veriyi analizleri sırasında bulur ve filtre ederler.• İstatistikçiler kendi sonuçlarını yorumlar ve bu sonuçları yöneticilere iletirler.	<ul style="list-style-type: none">• Veri madenciliği hipoteze gerek duymaz.• Veri madenciliği algoritmaları eşitlikleri otomatik olarak geliştirir.• Veri madenciliği nicelik ve nitelik verileri yanında farklı tiplerde veriler (örneğin metin, ses) de kullanır.• Veri madenciliği temiz veriye dayanır.• Veri madenciliğinin sonuçlarını yorumlamak kolay değildir. Veri madenciliği sonuçlarını analiz etmede ve yorumlamada ve bulguları yöneticilere iletmede mutlaka istatistikçiye ihtiyaç duyulmaktadır.

3.5.2 Veri madenciliği, OLAP ve veri sorgusunun kıyaslanması

- Hemen hemen ne arandığı biliniyor ve büyük veri tabanı ile çalışmak isteniyorsa veri sorgusu kullanılmalıdır.
- Büyük veri tabanlarında basit ilişkiler keşfedilmek isteniyorsa OLAP kullanılmalıdır.
- Veri içerisinde açıkça gözlenemeyen örüntü ve ilişkiler bulunmak isteniyorsa veri madenciliği kullanılmalıdır. Veri madenciliği algoritmalarının görece yavaşlığı nedeniyle, genellikle veri tabanının küçük veya örneklem olması gerekmektedir. Veri madenciliği algoritmalarını büyük veri tabanlarında

çalışacak biçimde ölçekleyebilmek, veri madenciliğinin güncel araştıma konularından birisidir.

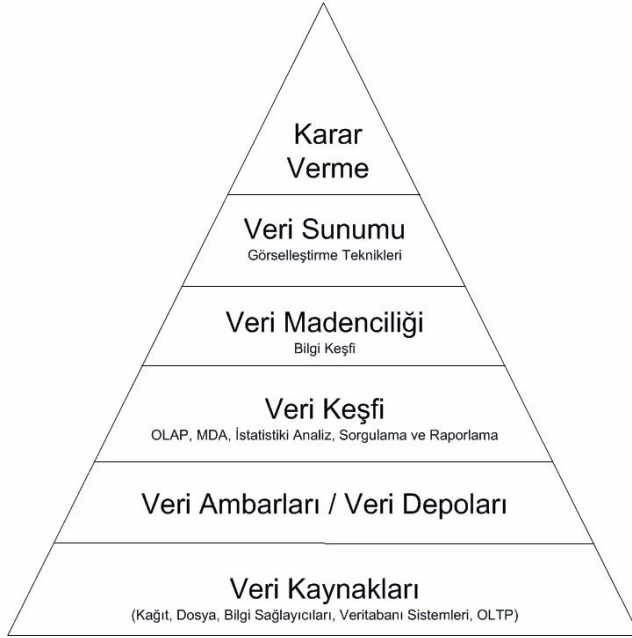
SQL, OLAP ve veri madenciliği kullanımını, keşfedilmek istenen bilgi tipine göre sınıflarsak:

- Sığ Bilgi: Seçilen kayıtlara ait ortalama ve toplam değer gibi özet bilgiler için kayıt seçmek yeterlidir ki SQL bunu yapabilir.
- Çok boyutlu bilgi: Farklı özelliklerin, ortaya çıkma sıklığı hakkında bilgi. Veri küpü üzerinde OLAP bunu yapabilir.
- Gizli bilgi: Önceden tahmin edilemeyen örüntü ve ilişkiler veri madenciliği için başlangıç olabilir.
- Derin bilgi: Sadece önsel teknik veya meta-bilginin kullanımıyla keşfedilebilecek gizli örüntüler ve ilişkiler hakkında bilgi. Bu konu veri madenciliğinin araştırma sınırları içerisindedir (Roiger and Geatz 2002).

3.6 Veri Madenciliği ve İş Zekası

İş zekası terimi, işte karar vermeyi destekleyen ve bilgi teknolojilerini temel alan bütün süreçler, teknikler ve araçlar için genel anlamda kullanılan bir ifadedir. Veri madenciliği iş zekasının yeni ve önemli bir bileşenidir. Şekil 3.2 farklı iş zekası çözümlerinin taktik ve stratejik iş kararları temelindeki potansiyel değerlerine göre mantıksal pozisyonlarını göstermektedir.

Genel olarak, piramitte aşağıdan yukarıya çıktıkça karar vermeyi destekleyen enformasyonun değeri artmaktadır (Cabena *et al.* 1997).



Şekil 3.2 Veri Madenciliği ve İş Zekası

3.7 Veri Madenciliği Metodolojisi

Bilgi keşfi kavramının ortaya çıkışında veri madenciliği sürecin bir parçası olsa da, günümüzde, veri madenciliği başlı başına bilgi keşfi ifadesi ile neredeyse eş anlamlı olarak anılmaktadır. Dolayısıyla, bu açıdan bakıldığında, veri madenciliğinin iş süreçleri aslında bilgi keşfinin iş süreçleri ile aynıdır. Ancak, zaman içerisinde veri madenciliğinin gelişimi ile beraber metodolojik bir standart belirleme ihtiyacı doğmuş ve bunun sonucunda bazı metodolojiler ortaya çıkmıştır.

Çizelge 3.2’de KDNUGGETS (2005) tarafından Nisan 2004’de gerçekleştirilen, metodoloji kullanımına yönelik en güncel araştırma sonuçları verilmektedir. Bünyesinde veri madenciliği projeleri barındıran 170 kuruluşa ‘Veri madenciliği için kullandığınız ana metodoloji nedir?’ sorusu sorulmuş ve soru neticesinde, tabloda da görüldüğü üzere yaygın metodolojilerden CRISP-DM (%42) ve SEMMA (%10) öne çıkmıştır. Araştırma sonuçlarına göre en ilgi çekici sonuç ise ‘Kendi metodolojim’ yanıtını verenlerin en yüksek ikinci orana sahip olmalarıdır.

Çizelge 3.2 Veri madenciliği için kullanılan temel metodolojiler

CRISP-DM (72)	% 42
SEMMA (17)	% 10
Kendi organizasyonumun (11)	% 6
Kendi metodolojim (48)	% 28
Diğer (10)	% 6
Yok (12)	% 7

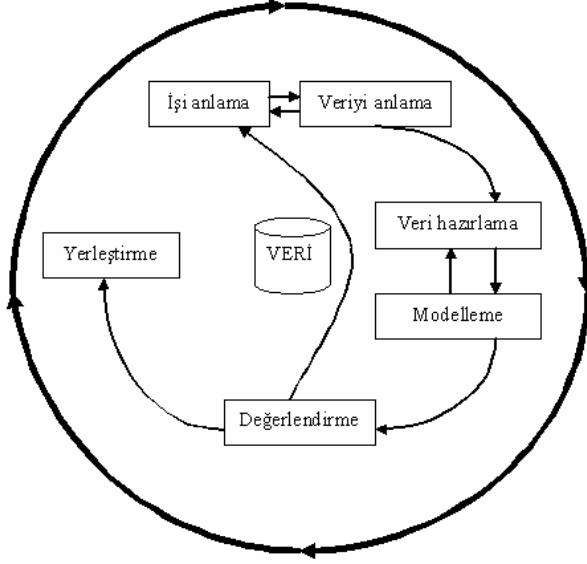
3.7.1 Veri madenciliği için sanayilerarası standart süreç

Daimler-Benz AG (Almanya), SPSS/Integral Solutions Ltd. (İngiltere) ve NCR Systems Engineering Copenhagen (Danimarka) veri madenciliğini halihazırda uygulayan ve kullanan firmalar olarak kendi başlarına uyguladıkları süreçlerin doğruluğunu ve bir standartın oluşturulup oluşturulamayacağını sorgulamışlardır. 1996 yılında, tüm uygulayıcılara yönelik, veri madenciliği için iyi düşünülmüş, kimseye ait olmayan, ücretsiz bir standart süreç modeli hazırlamak üzere yola çıkmışlardır. Yaklaşık bir yıl sonra Avrupa Komisyonu tarafından amaçlarına yönelik bir fon sağlanmış ve söz konusu gruba katılan OHRA Verzekeringen en Bank Groep B.V. (Danimarka) ile oluşturulan Özel İlgi Grubu (ÖİG) çalışmalarını yürütmeye devam etmiştir. 1999 yılı sonunda Veri Madenciliği için Sanayilerarası Standart Süreç (VMSSS) (CRoss-Industry Standard Process for Data Mining- CRISP-DM) CRISP-DM 1.0 versiyonunu hazırlamış ve veri madenciliği için standart bir süreç modeli önermişlerdir.

ÖİG'nin, VMSSS'i kendi projelerinde başarıyla uygulamasının neticesinde Ağustos 2004'te CRISP-DM 1.0 versiyonu kullanıcılara sunulmuştur.

CRISP-DM Special Interest Group raporunda da belirtildiği üzere VMSSS teknik prensiplere sahip teorik, akademik bir tarzda yaratılmamıştır. Geçmişte geliştirilen metodolojilerin pratik açıdan zayıf kalması nedeniyle, uygulamaya dönük ve veri

madenciliğine yönelik gerçek hayat tecrübelerine dayanması esas alınmıştır (CRISP-DM Special Interest Group 2004).



Şekil 3.3 VMSSS (CRISP-DM) Metodoloji Döngüsü

Şekil 3.3’de de görüldüğü CRISP-DM geri dönüşlere sahip 6 adımlık bir döngüden oluşmaktadır. Döngünün adımları;

1. Adım: İş anlama (Business understanding)
2. Adım: Veriyi anlama (Data understanding)
3. Adım: Veri hazırlama (Data preparation)
4. Adım: Modelleme (Modeling)
5. Adım: Değerlendirme (Evaluation)
6. Adım: Yerleştirme (Deployment)

biçiminde şekillenmektedir (CRISP-DM Special Interest Group 2004).

3.7.2 SEMMA

SEMMA, veri madenciliğinde yürütülen sürecin özünü gösteren; Örnekle (Sample), Araştır (Explore), Değiştir (Modify), Modelle (Model), Değerlendir (Assess) sürecinin başharflerinden oluşmaktadır.

SEMMA'nın detaylarına yer verilmezden önce; 'SEMMA'nın bir metodoloji olarak gösterilmesi gibi bir yanlış anlama sözkonusudur. SEMMA bir metodoloji değildir ama veri madenciliğinin yerine getirmesi gereken mantıksal organizasyonun ötesindedir.' ifadesi yer almaktadır. Buna karşın, uygulamada yaygın olarak veri madenciliği metodolojisi biçiminde yer alması nedeniyle aşağıda SEMMA'nın işleyişine yer verilmektedir.

Örnekle (Opsiyonel): Kolaylıkla işlem yapılabilecek kadar küçük, büyük veri setindeki önemli enformasyonu aktaracak kadar büyük bir veri setini örnek al. Optimal maliyet ve performans için SAS Enstitüsü, büyük veri setini güvenilir biçimde istatistiksel açıdan temsil edebilecek örnekleme stratejini savunmaktadır. Bütün veri yerine temsilci örnekleme madenlemek önemli iş bilgisini almak için gerekli işlem süresini indirgemektedir. Eğer genel örüntü bütün veride gözlenebiliyorsa, temsil edici örnekleme de izlenebilecektir. Eğer bir örüntü temsil edici örnekleme gözlenmiyorsa büyük resmi etkileyecek kadar önemli değildir, özetleme yöntemleri kullanılarak keşfedilebilir.

Araştır (Explore): Amaç ve uzlaşımlara ulaşmak için veri içerisindeki beklenmeyen eğilimler ve anomaliler araştırılacaktır. Araştırma, keşif sürecini arıtmaya yardım etmektedir. Eğer görsel araştırma anlaşılır eğilimleri açığa çıkarmıyorsa, veri faktör analizi, uyum analizi veya kümeleme gibi istatistik tekniklerle araştırılabilir.

Değiştir (Modify): Model seçme sürecine odaklanarak, değişkenler yaratıp, seçerek veya dönüştürerek veri değiştirilebilir.

Modelle (Model): İstenilen çıktıyı kabul edilebilir biçimde tahmin edecek veri kombinasyonu için yazılımın imkan tanıdığı bir model yaratılır. Veri madenciliğindeki modelleme teknikleri sinir ağlarını, ağaç tabanlı modelleri, lojistik modelleri ve zaman serisi analizi, hafıza tabanlı muhakeme (memory-based reasoning), temel bileşenler analizi gibi diğer istatistiksel yöntemleri içerir (SAS 2005).

Değerlendir: Veri madenciliği sürecinde elde edilen bulguları kullanışlı ve güvenilir olmasına göre değerlendirir ve ne kadar iyi düzenlediğini tahmin et (SAS 2005).

3.7.3 Diğer veri madenciliği metodolojileri

Veri madenciliğinin metodolojisi için sadece CRISP-DM ve SEMMA söz konusu değildir. Farklı veri madenciliği kaynaklarında, sözkonusu iki metodolojiden farklı birkaç metodoloji aşağıda sunulmaktadır.

Genel olarak veri madenciliği adımları:

1. Adım: İş amaçlarının tanımlanması
2. Adım: Veri hazırlanması
3. Adım: Veri dönüşümü
4. Adım: Veri madenciliği
5. Adım: Sonuçların analizi
6. Adım: Bilginin özümsemesi

olarak verilebilir (Cabena *et al.* 1997).

TWO CROWS (1999)'da ise veri madenciliği adımları;

1. Adım: İş problemini tanımla
2. Adım: Veri madenciliği veri tabanını kur
3. Adım: Veriyi keşfet
4. Adım: Modelleme için veriyi hazırla
5. Adım: Modeli kur
6. Adım: Modeli değerlendir

7. Adım: Modeli ve sonuçları yerleştir
biçiminde yer almaktadır.

Chen (2001) ise veri madenciliği sürecinin adımlarını;

1. Adım: Problemi tanımla
2. Adım: Veri biriktir/seç
3. Adım: Veri hazırla
4. Adım: Veri ön-işlemesi
- 5a. Adım: Model veya algoritma seç
- 5b. Adım: Model veya algoritmanın eğitim parametrelerini seç
6. Adım: Verinin eğitimi/testi veya algoritmanın uygulanması
7. Adım: Son değerlendirme/modelin entegrasyonu

biçiminde tanımlamaktadır.

CRISP-DM ve SEMMA'nın iki rakip yazılım firması tarafından kullanılıyor olması nedeniyle bir farklılık yaratma gayreti içerisinde oldukları sonucu çıkarılabilir. Ancak, Cabena *et al.* (1997), TWO CROWS (1999) ve Chen (2001) incelendiğinde hepsinin belirli noktalarda farklı metodolojiler önerdiği gözlenmektedir. Veri madenciliğine yönelik incelenen bütün kaynaklarda, önerilen metodolojilerin hemen her birinin bir şekilde diğerlerinden farklılaştığı gözlenmektedir. Buna karşın, bütün veri madenciliği metodolojilerinin aynı temel mantığa sahip olduğu, veri madenciliğinin çıkış noktası olan Veri Tabanlarında Bilgi Keşfi sürecinin adımlarını izledikleri görülmektedir. Bu nedenle, veri madenciliği metodolojisi olarak Veri Tabanlarında Bilgi Keşfi sürecinin adımlarının izlenmesinin doğru bir yaklaşım olduğu düşünülmektedir. Veri madenciliği metodolojisi için yapılabilecek bir diğer tespit ise veri madenciliğinin çok yeni bir alan olması nedeniyle henüz bir arayış söz konusu olduğu ve tanım birliğine dahi varılamayan bir alan için metodoloji standartlarının oluşması için henüz erken olduğu düşünülmektedir. Dolayısıyla, Bölüm 3.1.2'de verilen VTBK süreci adımları esas olmak üzere, uygulamanın yapısına göre bazı değişiklikler yapılabileceği kanaati oluşmuştur.

3.8 İstatistiksel Öğrenme Yaklaşımli Veri Madenciliği Yöntemlerinin Sınıflandırılması

Veri Madenciliği yöntemlerini,

- Denetimli,
- Denetimsiz,

olmak üzere iki ana kategoriye ayırmak mümkündür. Denetimli ve denetimsiz yöntemler için genel kabul görmüş tanımlar aşağıda verilmiştir.

Denetimli (Supervised): İyi tanımlanmış veya kesin bir hedef olduğunda denetimli (supervised) ifadesi kullanılır.

Denetimsiz (Unsupervised): Elde edilmesi istenen sonuç için özel bir tanımlama yapılmamışsa veya belirsizlik sözkonusu ise denetimsiz (unsupervised) ifadesi kullanılır (Hastie *et al.* 2001).

Denetimli ve denetimsiz ifadeleri birbirinin tersine karşılık gelmektedir. Denetimli ve denetimsiz yöntemleri sürecin bütünü açısından değerlendirmek gerekirse;

- Denetimsiz yöntemler daha çok veriyi anlamaya, tanımaya, keşfetmeye yönelik olarak kullanılan ve sonraki uygulanacak yöntemler için fikir vermeyi amaçlamaktadır,
- Denetimli yöntemler ise veriden bilgi ve sonuç çıkarmaya yönelik kullanılmaktadır,

denilebilir.

Bu nedenle denetimsiz bir yöntemle elde edilen bir bilgi veya sonucu, eğer mümkünse denetimli bir yöntemle teyit etmek, elde edilen bulguların doğruluğu ve geçerliliği açısından önem taşımaktadır.

Denetimli ve denetimsiz algoritmaların mantığını ve farklılığını en iyi aktaracak olan yöntem, Kümeleme Analizi'dir.

Aşamalı Kümelemede hem birimlerin hem de değişkenlerin birbiriyle değişik benzerlik ölçütlerine göre kümelenmesi; küme sayısı baştan verilmeden elde edilmektedir. Küme sayısı baştan belirli olmadığı için, Aşamalı Kümeleme Analizi denetimsiz (unsupervised) bir yöntemdir.

Aşamalı Olmayan Kümeleme yöntemlerinden K-Ortalamlar Kümelemede ise birimlerin uygun oldukları kümelerde toplanmaları ve 'n birimin k sayıda kümeye parçalanması' hedeflenmektedir. Küme sayısı baştan belli olduğu için, K-Ortalamlar Kümeleme Analizi denetimli (supervised) bir yöntemdir.

Başlıca Veri Madenciliği yöntemleri denetimli veya denetimsiz olmalarına göre;

- **Denetimli (Supervised)**
 - En yakın k komşuluk (k-Nearest-Neighbor)
 - K-ortalamlar kümeleme (K-means clustering)
 - Regresyon modelleri (Regression models)
 - Kural çıkarımı (Rule induction)
 - Karar ağaçları (Decision trees)
 - Sinir ağları (Neural networks)

- **Denetimsiz (Unsupervised)**
 - Aşamalı kümeleme (Hierarchical clustering)
 - Kendi kendini düzenleyen haritalar (Self organized maps)

olarak sınıflandırılabilir (Hastie *et al.* 2001 ve Thearling 2005).

3.9 Başlıca Veri Madenciliği Yöntemleri

Veri madenciliği ile ilgili kullanılan pek çok yöntemin yanına hemen her geçen gün yeni yöntem ve algoritmalar eklenmektedir. Bunlardan bir kısmı onlarca yıldır kullanılan klasik teknikler diyebileceğimiz ağırlıklı olarak istatistiksel yöntemlerdir. Diğer yöntemlerde genellikle istatistiği temel alan ama daha çok makine öğrenimi ve yapay zeka destekli yeni nesil yöntemlerdir.

Veri madenciliğinde kullanılan klasik yöntemlerin başlıcaları;

- Regresyon
- K - En Yakın Komşuluk
- Kümeleme

olarak sayılabilir.

Yeni nesil yöntemlerin başlıcaları ise;

- Karar Ağaçları,
- Birliktelik Kuralları,
- Sınır Ağları,

olarak sıralanabilir (Berson *et al.* 1999).

Ayrıca diğer veri madenciliği yöntemlerinin başlıcaları da;

- Temel Bileşenler Analizi,
- Diskriminant Analizi,
- Faktör Analizi,
- Kohonen Ağları,
- Bulanık Mantığa Dayalı Yöntemler,
- Genetik Algoritmalar,
- Bayesci Ağlar,
- Pürüzlü (Rough) Küme Teorisine Dayalı yöntemler,

olarak sıralanabilir (Chen 2001).

Yukarıda sayılan yöntemlerin dışında birden fazla tekniği içine alan hibrid yöntemler ve zaman serilerine dayalı yöntemlerden de veri madenciliği yöntemi olarak faydalanılmaktadır (Kovalerchuk and Vityaev 2002).

Özet olarak, bilgi keşfine yarayan her yöntem veri madenciliği yöntemi olarak kullanılabilir.

3.9.1 Regresyon analizi

$X = (X_1, X_2, \dots, X_p)$ girdilerinden oluşan bir vektör ve Y tahmin etmek istediğimiz gerçek değerli bir çıktı olsun. Doğrusal Regresyon modeli,

$$f(X) = \beta_0 + \sum_{j=1}^p X_j \beta_j \quad (3.1)$$

biçimindedir.

Doğrusal model, regresyon fonksiyonu $E(Y|X)$ 'in doğrusal olduğunu varsayar veya doğrusal model anlamlı bir yaklaşımdır. β_j 'ler bilinmeyen parametreler veya katsayılar olmak üzere, X_j değişkenleri farklı kaynaklardan gelebilir:

- Nicel girdiler,
- Nicel girdilerin logaritmik, karekök veya karesel gibi dönüşümler,
- $X_2 = X_1^2$, $X_3 = X_1^3$ gibi polinomial gösterimleri işaret eden temel genişletmeler,
- Nitel girdilerin düzeylerinin sayısal veya kukla kodlamaları. Örneğin, eğer G beş düzeyli fatör girdiyse, olacak biçimde $X_j = I(G = j)$, X_j , $j=1, \dots, 5$ değişkeni yaratılabilir. X_j grubu, düzey bağımlı sabitler kümesi ile birlikte, G'nin

düzeyini temsil edebilir, öyleki $\sum_{j=1}^5 X_j \beta_j$ 'de X_j 'lerden birisi bir ve diğerleri sıfırdır.

- Değişkenler arası etkileşimler, örneğin $X_3 = X_1 X_2$.

X_j 'lerin kaynağı ne olursa olsun, parametrelerde model doğrusaldır.

Tipik olarak, tahmin edilmek istenen β_j parametrelerinden $(x_1, y_1) \dots (x_N, y_N)$ eğitim verisi kümesi vardır. Herbir $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ i . olayın içerik ölçümünün vektörüdür. En popüler tahmin yöntemi en küçük karelerdir. Öyleki, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ katsayıları, artık kareler toplamı

$$\begin{aligned} RSS(\beta) &= \sum_{i=1}^N (y_i - f(x_i))^2 \\ &= \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 \end{aligned} \quad (3.2)$$

ifadesini en küçükleyecek biçimde seçilir.

Eğer eğitim gözlemleri (x_i, y_i) kendi kitlesinden bağımsız rasgele çekilişleri temsil ediyorsa bu kriter istatistiksel bakış açısından anlamlıdır. x_i 'ler rasgele çekilmemişse bile, eğer y_i 'ler verilen x_i girdilerinden koşullu bağımsız ise kriter hala geçerlidir.

Eşitlik (3.2)'nin nasıl minimize edileceğine bakılırsa; her bir satırı girdi vektörü olan $N \times (p+1)$ vektörünü (ilk sıra 1 olsun) X ile gösterelim ve benzer biçimde y eğitim verisindeki çıktılardan N -vektörü olsun. Buna göre artık kareler toplamı

$$RSS(\beta) = (y - X\beta)^T (y - X\beta) \quad (3.3)$$

biçiminde ifade edilebilir. Bu ifade $p+1$ parametrenin karesel fonksiyonudur. β 'ya göre türev alındığında;

$$\begin{aligned} \frac{\partial RSS}{\partial \beta} &= -2 X^T (y - X\beta) \\ \frac{\partial^2 RSS}{\partial \beta \partial \beta^T} &= -2 X^T X \end{aligned} \quad (3.4)$$

elde edilir.

Varsayalımki (bir an) X tekil olmayan ve böylece $X^T X$ pozitif tanımlıdır, ilk türev sifıra eşitlenirse

$$X^T (y - X\beta) = 0 \quad (3.5)$$

eşsiz sonuç

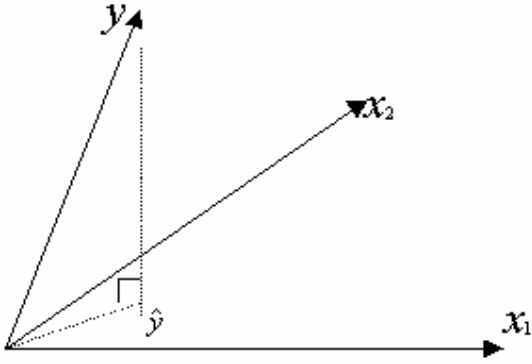
$$\hat{\beta} = (X^T X)^{-1} X^T y \quad (3.6)$$

elde edilir.

Verilmiş $\hat{f}(x_0) = (1 : x_0^T) \hat{\beta}$ için x_0 girdi vektöründeki tahmin değerleri, $\hat{y}_i = \hat{f}(x_i)$ olmak üzere,

$$\hat{y} = X\hat{\beta} = X (X^T X)^{-1} X^T y \quad (3.7)$$

eğitilmiş girdilerindeki uydurulmuş değerlerdir. (3.7) eşitliğinde görünen $H = X (X^T X)^{-1} X^T$ matrisi bazen ‘şapka’ matrisi olarak anılır, çünkü y üzerine şapka koyar.



Şekil 3.4 İki tahmin ediciyle en küçük kareler regresyonunun N-boyutlu geometrisi – Çıktı vektörü y x_1 ve x_2 girdi vektörlerince kapsanan hiper düzlem üzerinde ortogonal (dik) olarak projekte. \hat{y} projeksiyonu en küçük kareler tahminlerinin vektörlerini temsil eder.

Şekil 3.4 en küçük kareler tahmininin R^N ’deki geometrik temsilini gösterir. X ’in sütun vektörleri, $x_0 \equiv 1$ olmak üzere x_0, x_1, \dots, x_p ile gösterilmiştir. Bu ilk sütuna da diğerleri gibi muamele edilecektir. Bu vektörler R^N ’in bir alt uzayını kapsar, aynı zamanda X ’in sütun vektörlerini ifade eder. $RSS(\beta) = \|y - X\beta\|^2$, $\hat{\beta}$ ’nin seçimiyle en küçüklenir, böylece artık vektörü $y - \hat{y}$ bu alt uzaya diktir. Bu diklik (3.5)’de gösterilmiş ve sonuç tahmini \hat{y} böylece bu alt uzay üzerinde y ’nin dik projeksiyonudur. Şapka matrisi H ortogonal projeksiyonu hesaplar ve böylece aynı zamanda projeksiyon matrisi olarak da bilinir.

X'in sütunları doğrusal bağımsız olmayabilir, böylece X tam rankli değildir. Örneğin $x_2 = 3x_1$ gibi iki girdi mükemmel biçimde korele ise bu olabilir. Ardından $X^T X$ tekildir ve en küçük kareler katsayıları $\hat{\beta}$ 'lar eşsiz olarak tanımlanamaz. Bununla beraber, uydurulmuş değerler $\hat{y} = X\hat{\beta}$ X'in sütun uzayları üzerinde y'nin projeksiyonudur; X'in sütun vektörleri cinsinden bu projeksiyonu göstermenin birden fazla yolu vardır. Bir veya birden fazla nitel girdiler gereksiz olarak kodlandığı zaman, tam ranklı olmama durumu çok sık olarak meydana gelir. Eşsiz olmayan temsili çözümlerin sıklıkla doğal yolu X'deki gereksiz sütunları düşürmek ve/veya yeniden kodlamaktır. Pek çok regresyon yazılım paketleri bu fazlalıkları tespit etmekte ve bunları otomatik olarak kaldırmak için bazı stratejiler uygulamaktadır. Aynı zamanda rank eksiklikleri, girdi sayısı p'nin eğitim olayı sayısı N'yi geçtiği durumlarda, sinyal ve imaj analizlerinde de meydana gelir. Bu durumda, içerik tipik olarak filtreleme yoluyla indirgenir veya aksi taktirde düzenlemeler kontrol edilerek uydurulur.

Şu ana kadar verinin gerçek dağılımı hakkında minimal varsayımlar yapılmıştır. $\hat{\beta}$ 'nin örneklem özelliklerine bağlı kalarak, y_i gözlemlerinin korele olmadığını, sabit σ^2 varyansına sahip olduğunu ve x_i 'lerin sabit (rasgele olmayan) olduğunu varsayacağız. EKK parametre tahminlerinin varyans-kovaryans matrisi, (3.6)'dan kolayca türetilir ve

$$Var(\hat{\beta}) = (X^T X)^{-1} \sigma^2 \quad (3.8)$$

ile gösterilir.

Tipik olarak σ^2 varyansının tahmini,

$$\hat{\sigma}^2 = \frac{1}{N-p-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

ile yapılır.

Paydada N yerine N-p-1 olmasının nedeni $\hat{\sigma}^2$ 'yi σ^2 'nin yansız tahmini yapmasıdır:

$$E(\hat{\sigma}^2) = \sigma^2.$$

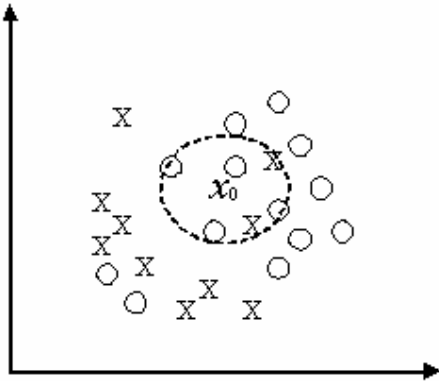
Parametreler ve model hakkında çıkarımlar yapmak için ek varsayımlara ihtiyaç duyulmaktadır. Şimdi (3.1)'in doğru model olduğunu varsayacağız; bu Y'nin koşullu bekleneninin X_1, \dots, X_p 'de doğrusal olduğu anlamına gelir.

$$\begin{aligned} Y &= E(Y | X_1, \dots, X_p) + \varepsilon \\ &= \beta_0 + \sum_{j=1}^p X_j \beta_j + \varepsilon \end{aligned}$$

ifadesinde $\varepsilon \sim N(0, \sigma^2)$ 'dir (Hastie *et al.* 2001).

3.9.2 K-en yakın komşuluk

Bu sınıflandırıcılar hafıza tabanlıdır ve uyduracak modele ihtiyaç duymazlar. x_0 sorgu noktası verilmiş olsun, x_r , $r=1, \dots, k$ eğitim noktasından mesafe olarak x_0 'a en yakın olanı bulunur ve k komşudan daha fazla oy alana göre sınıflandırılır. Şekil 3.5'e bakıldığında; sorgulanan x_0 noktası, 7 - En yakın komşuluğuna göre O sınıfının elemanıdır.



Şekil 3.5 x_0 noktasının 7 - en yakın komşuluğu

Kolaylık açısından niteliklerin reel değerli olduğu varsayılacak ve nitelik uzayında Öklid Mesafesi

$$d_i = \|\mathbf{x}_{(i)} - \mathbf{x}_0\| \quad (3.9)$$

kullanılacaktır.

Tipik olarak bütün nitelikler ortalaması 0, standart sapması 1 olacak şekilde standardize edilecek, böylece farklı birimlerden ölçümleri değerlendirmek mümkün olacaktır.

Kolaylığına rağmen, k en yakın komşuluk, elyazısı rakamları, uydu görüntüleri ve EKG örüntülerini de içeren büyük hacimli sınıflandırma problemlerine başarıyla uygulanabilmektedir.

Sadece, sorgu noktasına en yakın eğitim noktalarının kullanımı nedeniyle, 1-en yakın komşuluk tahmininin yanlılığı genellikle düşüktür ama varyansı yüksektir (Hastie *et al.* 2001).

3.9.3 K-ortalamlar kümeleme analizi

Tezin özgün uygulaması olan Erken Uyarı Sistemi'nde kullanılan algoritmalardan birisi de Denetimli İstatistiksel Öğrenme yöntemlerinden olan K-ortalamlar Kümeleme Analizi'dir. Belirli bir benzemezlik (veya benzerlik) ölçüsüne dayalı olarak N gözlemi, karar verici tarafından belirlenen K gruba ayırmayı sağlayan algoritma; küme içinde homojen kümeler arasında heterojen gruplar yaratmaktadır.

K-ortalamlar algoritması iterativ (yinelemeli) olarak azalan kümeleme yöntemlerinin en popülerlerinden birisidir. Bütün değişkenlerin nicel ve benzemezlik ölçüsü olarak

$$d(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{j=1}^p (\mathbf{x}_{ij} - \mathbf{x}_{ij'})^2 = \|\mathbf{x}_i - \mathbf{x}_{i'}\|^2$$

karesele Öklid mesafesinin olduğu durumlarda kullanılır. Belirtilmeli ki, x_{ij} değişkenlerinin yeniden tanımlanmasında ağırlıklandırılmış Öklid mesafesi de kullanılabilir.

İç noktaların saçılımı,

$$W(C) = \frac{1}{2} \sum_{k=1}^K \sum_{C(i)=k} \sum_{C(i')=k} \|x_i - x_{i'}\|^2$$

olarak yazılabilir,

$$= \sum_{k=1}^K \sum_{C(i)=k} \|x_i - \bar{x}_k\|^2 \quad (3.10)$$

olmak üzere, $\bar{x}_k = (\bar{x}_{1k}, \dots, \bar{x}_{pk})$ k. küme ile ilişkili ortalama vektörüdür. Böylece, kriter N gözlemi K kümeye atayarak minimize edilir; her bir küme için, gözlemlerin küme ortalamasından ortalama benzemezliğini minimize etme yolu izlenir.

Yinelemeli (iterativ) azalış algoritması için,

$$C^* = \min_C \sum_{k=1}^K \sum_{C(i)=k} \|x_i - \bar{x}_k\|^2$$

çözümü, herhangi bir gözlem kümesi S için

$$\bar{x}_S = \arg \min_m \sum_{i \in S} \|x_i - m\|^2 \quad (3.11)$$

ile elde edilebilir.

Böylece genişletilmiş optimizasyon problemi

$$\min_{C, \{m_k\}_k} \sum_{k=1}^K \sum_{C(i)=k} \|x_i - m_k\|^2 \quad (3.12)$$

çözümüyle C^* 'i elde edebiliriz.

İfade (3.12), algoritma 3.1’de verilen ardışık optimizasyon prosedürü ile minimize edilebilir.

Algoritma 3.1 K-ortalamlar kümeleme analizi

1. Verilmiş küme ataması C için, toplam küme varyansı (3.12), (3.11)’deki geçerli atanmış kümelerin ürünü olan $\{m_1, \dots, m_k\}$ ortalamalarıyla minimize edilir.

2. Verilmiş $\{m_1, \dots, m_k\}$ ortalamalar kümesi olmak üzere, (3.12) her bir gözlem en yakın (geçerli) kümeye atanarak minimize edilir. Bu durum,

$$C(i) = \arg \min_{1 \leq k \leq K} \|x_i - m_k\|^2 \quad (3.13)$$

ifadesiyle gösterilir.

3. Adım 1 ve 2 atamalar değişmeyene kadar yinelenir.

Adım 1 ve 2’nin her biri (3.12) kriterinin değerini indirger, böylece yakınsama gerçekleşir. Bununla beraber, sonuç, alt-optimal local minimumu temsil edebilir. Hartigan and Wong (1979)’un algoritması daha ileriye gider ve gözlemlerin bir gruptan diğerine bir tek geçiş olmadığından emin olur ki bu amacı indirgeyecektir. Buna ek olarak, algoritma başlangıç ortalamaları için pek çok rasgele seçimler yaparak başlatılabilir ve amaç fonksiyon için en küçük değere sahip çözüm seçilebilir (Hastie *et al.* 2001).

3.9.4 Aşamalı (Hierarchical) kümeleme yöntemleri

K-ortalamlar'ın uygulanmasının sonucu, seçilen küme sayısının araştırılmasına ve başlangıç konfigürasyonunun atamasına bağlıdır. Zıt olarak, aşamalı kümeleme yöntemleri böyle tanımlamalara ihtiyaç duymaz. Bunun yerine kullanıcının, iki gruptaki gözlem çiftlerinin benzemezliğini temel alan (ayrık) grupların gözlemleri arasında bir benzemezlik ölçüsü tanımlaması gerekir. Adından da anlaşıldığı gibi, hiyerarşinin her bir düzeyindeki kümeler, bir alt düzeydeki kümelerin birleştirilmesinden oluşan bir hiyerarşik gösterimi üretir. En alt düzeyde, her bir küme bir tek gözlemi içerir. En üst düzeyde, bütün verileri içeren sadece bir tek küme vardır.

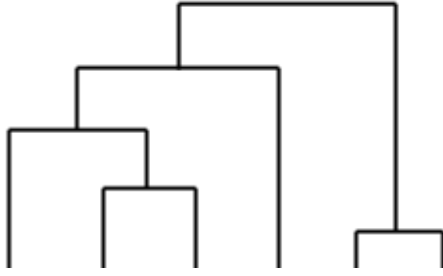
Aşamalı kümeleme için stratejiler iki temel çekim noktasına bölünmüştür: Birleştirici (Agglomerative) (aşağıdan-yukarıya) ve ayırıcı (divisive) (yukarıdan-aşağıya). Birleştirici stratejiler en alttan başlar ve her bir düzeyde seçilmiş küme çiftleri tekrarlanarak bir tek küme içinde birleştirilir. Bu, bir üst düzeyde bir eksik küme ile gruplar üretir. Birleştirme için seçilen çift en küçük grupları benzemezliğe sahip iki grubun birleştirilmesinden oluşur. Ayırıcı yöntemler en yukarıdan başlar ve her bir düzeyde mevcut kümeler tekrarlanarak iki yeni kümeye ayrılır. Ayırma, en büyük gruplar arası benzemezliğe sahip iki yeni grup üretmek üzere seçilir. Her iki çekim noktasında da, hiyerarşide N-1 düzey vardır.

Hiyerarşinin her bir düzeyi gözlemlerin ayrık kümelerini içeren, verilerin özel gruplarını ifade eder. Bütün hiyerarşi, böyle grupların ardışık gösterir. Hangi düzeydeki gözlemlerin kendi grubundaki gözlemlerle, aynı düzeyde farklı gruba atanan gözlemlere göre, yeterince daha benzer olduğu, (eğer varsa) gerçek yapısını doğal bir küme olarak sunduğu kararı kullanıcının kararına bağlıdır. Açıklık istatistiği bu amaçla önceden tanımlanır.

Tekrarlanan ikili ayırma/birleştirme, köklere sahip ikili bir ağaç olarak gösterilebilir. Ağacın düğümleri, grupları temsil eder. Kök düğümü tüm veri setini temsil eder. N sonuç düğümünün her biri, bir tek gözlemi temsil eder. Her bir sonuç düğümü olmayan

düğüm (ana düğüm) iki yavru düğüme sahiptir. Ayırıcı kümelemede, iki yavru düğüm, ana düğümün bölünmesi sonucu oluşan iki grubu temsil eder; birleştirici kümeleme için iki yavru düğüm, ana düğüm halinde birleştirilecek iki grubu temsil eder.

Bütün birleştirici ve bazı ayırıcı yöntemler (aşağıdan yukarıya bakıldığında) monotonluk özelliğine sahiptir. Öyleki, birleştirilmiş kümelerdeki benzemezlik, birleştirici düzey ile moton artandır. Böylece ikili ağaç çizilebilir, öyleki, her düğümün yüksekliği, iki yavru düğümün grup-içi benzemezlik değeriyle orantılıdır. Sonuç düğümleri sadece gözlemleri temsil eder ve hepsi sıfır yüksekliğiyle çizilir. Bu tipte grafiksel gösterime dendogram denir ve Şekil 3.6'daki gibi bir gösterime sahiptir.



Şekil 3.6 Dendogram

Bir dendogram, hiyerarşik kümeleme yönteminin grafiksel biçimde tam bir betimlemesini yüksek düzeyde yorumlanabilir biçimde sunar. Hiyerarşik kümeleme analizinin popülaritesinin temel nedenlerinden birisi de budur.

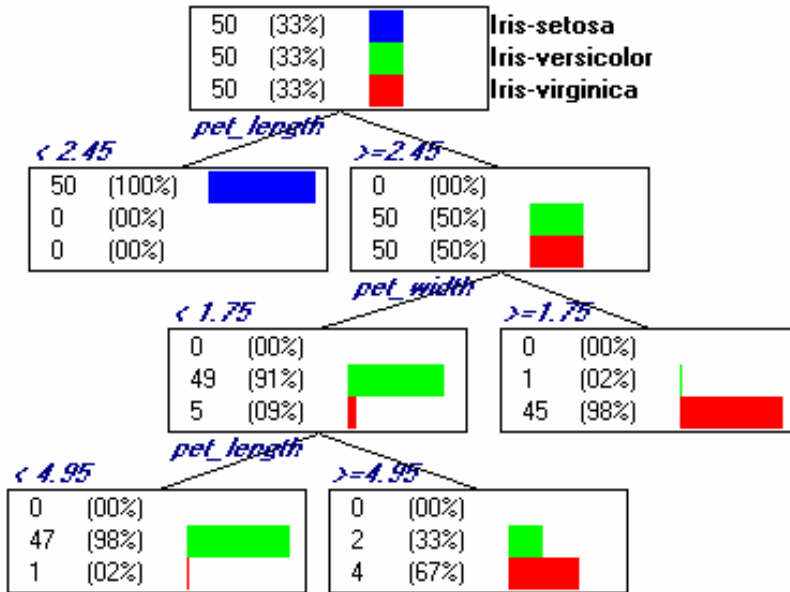
Bir dendogram, algoritmanın sonuçlarının betimlemesinin ötesinde, sık sık verinin kendisinin grafiksel özeti olarak da görülebilir. Bununla beraber, böyle yorumlar dikkatli yapılmalıdır. Birincisi, farklı hiyerarşik yöntemler, veride küçük değişikliklere yol açacak, bu da yeterince farklı dendogramlara işaret edecektir. Aynı zamanda, böyle bir özetleme de algoritma tarafından üretilen hiyerarşik yapının sahip olduğu gözlem çiftlerinin benzemezliğinin uzantısı olarak geçerli olacaktır. Hiyerarşik yöntemler,

veride gerçekte böyle bir yapı olsun veya olmasın, hiyerarşik yapıyı empoze edecektir (Hastie *et al.* 2001).

3.9.5 Karar ağaçları

Veri Madenciliği yöntemleri klasik ve yeni nesil yöntemler olmak üzere sınıflandırıldığında, yeni nesil yöntemlerden ilk akla gelenlerden birisi Karar Ağaçlarıdır. Hedef değişkeni, tahmin edici değişkenlerle arasındaki ilişkiyi dikkate alarak haritalamayı sağlayan ve ağaç dalları biçiminde segmente eden yöntem; tezin özgün uygulaması olan Erken Uyarı Sistemi'nde de bünyesinde yer alan farklı algoritmalarından birisi olan CHAID kullanılarak yer almaktadır.

Karar ağacı adından da anlaşıldığı gibi ağaç olarak görünen, tahminsel bir modeldir. Ağacın her dalı bir sınıflandırma sorusu ve yaprakları da veri setinin bu sınıflandırmaya ait parçalarıdır. Şekil 3.7'de bir karar ağacı örneği görülmektedir.



Şekil 3.7 Bir karar ağacı örneği

Ağaç ile ilgili temel özellikler aşağıda sıralanmaktadır:

- Veriyi herhangi bir kayba yol açmadan her bir dala ayırmaktadır.
- Modelin nasıl yapılandırıldığını anlamak çok kolaydır (Sınır ağları veya standart istatistik modellerinin tersine).
- Oluşturulan modeli kullanmak da çok kolaydır. Ayrıca, bazı sezgileri de modelde yapılandırmak mümkündür.
- İççe geçmiş eğer / sonra (if / then) kurallarının dizisidir. Görsel olması nedeniyle oldukça kolay anlaşılırdır. Aynı zamanda kolaylıkla SQL sorgusuna dönüştürülebilir.
- Değişken tiplerine göre, farklı yöntemler kullanılabilir.

İş perspektifinden karar ağaçları orijinal veriden segmentler yaratmak yaratmak olarak görülebilir (her bir segment ağacın bir yaprağı olacaktır). Müşterilerin, ürünlerin ve satış bölgelerinin segmentasyonu pazarlama yöneticilerinin pekçok yıldır yaptığı bir şeydir. Geçmişte bu segmentasyon büyük miktarda verinin üst düzeyde görünümünü elde etmek için düzenlenmektedir ve segmentasyonun benzer verileri bir araya toplamak dışında özel bir nedeni yoktur (Berson *et al.* 2000).

Bu durumda segmentasyon yapmanın bazı önemli enformasyon parçalarını tahmin etmek gibi özel bir nedeni vardır. Her bir segmente düşen kayıtların buraya düşme nedeni, iyi tanımlanmamış bir benzerlik yokken (halihazırda benzer değıllerdir) tahmin edilecek enformasyondaki benzerlik nedeniyledir. Karar ağacından çıkarılan bu tahminsel segmentler aynı zamanda tahmin edilen segmentlerin karakteristiklerini de betimlemektedir. Böylece karar ağaçları ve onları yaratan algoritmalar karmaşık olabilir, sunulan sonuçlar ise iş kullanıcıları için oldukça faydalı ve anlaşılması kolaydır (Berson *et al.* 2000).

Ağaç yapısından ve kolay kural çıkarımına imkan tanınması nedeniyle karar ağaçları anlaşılabilir modeller kurmak için oldukça faydalı bir tekniktir.

Karar ağacı teknolojisi veri setlerinin ve iş problemlerinin keşfi için kullanılabilir. Bu genellikle ağacın her bir bölümündeki tahmin edicilere ve değerlerine bakarak yapılabilir. Sıklıkla bu tahmin ediciler kullanılabilir içerik sağlayabilir veya cevaplanması gereken sorular önerebilir. Eğer ağaç tek bir kayıt kalana kadar büyümesine devam ederse, pek çok soru ve dal yaratılacağı tasavvur edilebilir. Ağacı bu kadar büyütme hem hesaplama açısından pahalı hem de gereksizdir. Pek çok karar ağacı algoritması aşağıdaki üç kriterden biri görüldüğünde büyümeyi durdurur:

- Segment sadece bir tek kayıt içerir.
- Segmentteki bütün kayıtlar benzer karakteristiklere sahiptir.
- Büyüme yeni bir bölüm yapmak için yeterince önemli değildir (Berson *et al.* 2000).

Son 30 yılda pek çok karar ağacı öğrenim metodu geliştirilmiştir ve kredi başvurusunda risk değerlendirmesi gibi finansal ve bankacılık uygulamalarında başarılı olarak kullanılmaktadır. Aslında, her bir karar ağacı insanın okuyabileceği eğer-ise kuralıdır. Karar ağacı temsilci kuralların keşfinin özel bir yolunu yansıtmaktadır. Bu yolla keşif karar ağacı öğreniminin önemli bir avantajını yaratmaktadır: keşfedilen kurallar tutarlıdır. Herhangi bir karar ağacından aktarılan kurallar arasında zıtlık yoktur. Diğer taraftan, bağımsız olarak keşfedilen kurallar birbirine zıt olabilir. Eğer öğrenim verisi önemli ölçüde gürültü içeriyorsa, bu tipik bir sonuçtur. Bu nedenle, okunabilirlik ve tutarlılık karar ağacı öğreniminin önemli özelliklerindedir (Kovalerchuk and Vityaev 2000).

Karar ağacı modelleri

Karar Ağaçları yöntemleri, temel anlamda hedef değişkeni, tahmin edici değişkenlere göre segmente etme mantığına dayansa da; bünyesinde değişik amaçlara hizmet eden birbirinden farklı algoritmalara sahiptir. En sık kullanıma sahip karar ağacı algoritmaları, geçmişten günümüze gelişimini de sergileyecek biçimde aşağıda sıralanmaktadır.

i. ID3 ve C4.5

1970'lerin sonuna doğru J. Ross Quinlan ID3 denilen karar ağacı algoritmasını sunmuştur. Başlangıçta ID3 tavlâ ve oyunlar için iyi oyun oynama stratejilerini öğrenmek gibi görevler için kullanılmıştır. ID3 akademi ve sanayi olmak üzere çok geniş bir alanda problemlere uygulanmaya başladığından beri oldukça değişmiş ve gelişmiştir.

ID3 tahmin edicileri ve onların bölünüm değerlerini bölünümle sağlanan enformasyon kazancına göre seçmektedir. Kazanç, bölünüm öncesi ve sonrasında doğru tahmin yapabilmek için ihtiyaç duyulan enformasyon miktarındaki farkı anlatmaktadır. Kazanç, orijinal segmetin entropisi ile sonuçta elde edilen bölünüm segmentlerinin toplam entropileri arasındaki farktır.

ID3 daha sonra C4.5 adı verilen versiyona dönüşmüştür. C4.5, ID3'e göre pek çok alanda gelişmiştir. C4.5'de,

- Kayıp veriye sahip tahmin ediciler hala kullanılabilir,dir,
- Sürekli değere sahip tahmin ediciler kullanılabilir,dir,
- Budama takdim edilmiştir,
- Kural çıkarımı sağlanmıştır.

Bu tekniklerin çoğu ve bazı artılar daha CART algoritmasında görülmektedir (Berson *et al.* 2000).

ii. CART

Görelî olarak daha yeni parametrik olmayan analiz tekniğı CART (Classification and Regression Trees) veya Sınıflandırma ve Regresyon Ağaçları popüler bir veri analiz yöntemi olmaya başlamıştır (Ribic and Miller 1998).

CART ağacını oluşturmada tahmin ediciler, farklı tahminlerin kayıtları nasıl etkilediği temeline göre seçilmektedir. Örneğin verilmiş bir tahmin edici için verilmiş bir bölünüm noktasının diğerlerine göre daha iyi olup olmadığını tanımlayan bir ölçü entropi metriğidir. Ölçü temelini Claude Shannon ve Warren Weaver'ın 1949 yılında enformasyon teorisi üzerine yaptıkları çalışmadan almaktadır ve enformasyonun telefon hatları üzerinden nasıl etkin olarak iletileceğini dikkate almışlardır. İlginç biçimde, onların sonuçları karar ağaçlarının yaratılmasında da faydalı olmuştur.

CART'ın büyük avantajlarından birisi algoritmanın modelin geçerliliğini sınaması ve genel optimal modelin keşfedilmesidir. CART çok karmaşık bir ağaç oluşturup ardından geriye doğru budayarak çapraz geçerlilik sınaması ve test kümesi sınaması sonuçlarını temel alan optimal genel ağaca ulaşarak bunu başarmaktadır. Ağaç, test veri kümesi üzerinde ağacın pek çok budanmış versiyonunun performansını temel alarak geriye doğru budanmaktadır. Ağaç, yeni ve görülmemiş veriyi, çapraz geçerlilik sınamasını kullanarak iyi biçimde seçer.

CART algoritması, kayıp veri dikkate alındığında görece olarak sağlamdır. Eğer belli bir değişkenin belli bir kayıdı için bir değer kayıpsa, ağaç oluşturulurken optimal bölünümün tanımlanmasında bu kayıt kullanılmayacaktır. Etkin bir CART'da, en iyi mümkün bölünümün seçimi kararının verilmesinde eldeki enformasyon olabildiğince fazla kullanılacaktır.

CART yeni veri üzerinde tahmin amaçlı kullanıldığında, kayıp veriler vekillerle ele alınır. Vekiller, ağaçta gerçek bölünümü taklit eden bölünüm değerleri ve tahmin edicilerdir ve tercih edilen tahmin edici için veri kayıp olduğunda kullanılacaktır. Her ne kadar ayakkabı numarası yükseklik için mükemmel bir tahmin edici olmasa da, belli bir kayıta bu enformasyon kayıp olduğunda yüksekliği temel alan bölünümü taklit etmek için vekil olarak CART modelince tahmin edilebilir (Berson *et al.* 2000).

iii. CHAID (Chi-Square Automatic Interaction Detector)

Tezin özgün uygulaması olan Erken Uyarı Sistemi'nde kullanılan yöntemlerden birisi de CHAID karar ağacı algoritmasıdır. Hedef değişkeni, tahmin edici değişkenlerle ilişki düzeyine göre sınıflandırma amacıyla; diğer karar ağacı algoritmalarından farklı olarak ikiden fazla gruba ayırarak dallanan algoritma, tüm olası alt grupları ağaç biçiminde kolay anlaşılır biçimde göstermektedir.

CART ile aynı düzeyde popüler bir başka karar ağacı algoritması CHAID (Chi-Square Automatic Interaction Detector) Otomatik Ki-kare Etkileşim Keşfedicisi'dir. CHAID karar ağacını oluşturmada CART'a benzemektedir ama bölünümüleri seçmede farklı bir yol izlemektedir. Optimal bölünümün seçiminde Entropi veya Gini metriklerinin kullanımı yerine değerlerin tahmininde hangi kategorik tahmin edicinin bağımsızlıktan en uzak olduğunu tanımlamak için olasılık tablolarında Ki-Kare testi kullanımına dayanmaktadır.

CHAID ile diğer yöntemler arasındaki en önemli farklılıklarından birisi, ağaç türetimidir. ID3, C 4.5 ve CART ikili ağaçlar türetirken, CHAID ikili olmayan çoklu ağaçlar türetir (Berson *et al.* 2000).

CHAID sürekli ve kategorik tüm değişken tipleriyle çalışabilmektedir. Bununla beraber, sürekli tahmin edici değişkenler otomatik olarak analizin amacına uygun olarak kategorize edilmektedir. CHAID algoritması aşağıda verilmektedir:

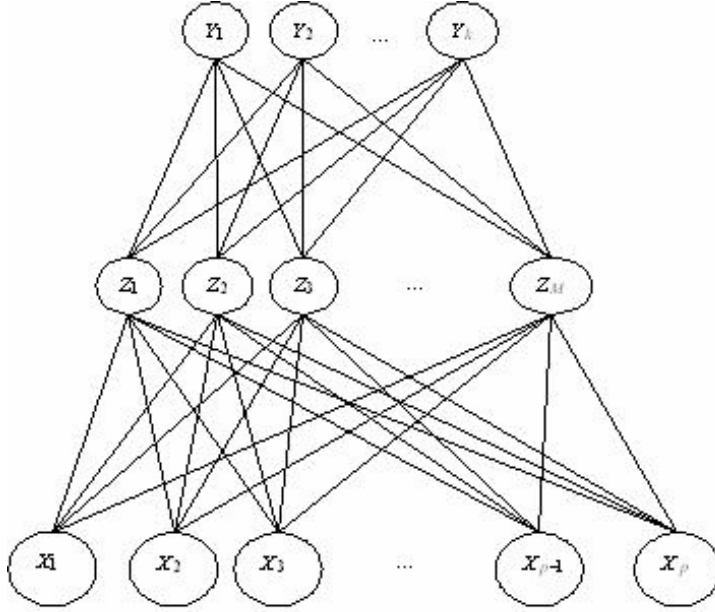
1. Her bir tahmin edici değişken X için, X 'in, Y hedef değişkenini dikkate alan en az öneme sahip kategori çiftini bul (bu, en büyük p değerine sahip olandır). Yöntem, Y 'nin ölçüm düzeyine bağlı olarak p değerlerini hesaplayacaktır.
 - a. Eğer Y sürekli ise F testini kullan.
 - b. Eğer Y isimsel ise X 'in kategorileri satırlarda ve Y 'nin kategorileri sütunlarda olacak biçimde iki yönlü çapraz tablo düzenle. Pearson ki-kare testini veya olabilirlik oranı testini kullan.

- c. Eğer Y sıralı ise bir Y birliktelik modeli uygundur. Olabilirlik oranı testini kullan.
2. En büyük p değerine sahip X'in kategori çifti için, p değerini önceden belirlenmiş alfa düzeyi $\alpha_{birleş}$ ile kıyasla.
 - a. Eğer p değeri $\alpha_{birleş}$ 'den büyük ise bu çifti bir tek kategori altında birleştir. X'in yeni kategori kümesi için süreci Adım 1'den başlat.
 - b. Eğer p değeri $\alpha_{birleş}$ 'den küçük ise Adım 3'e git.
3. X'in ve Y'nin kategori kümesi için uygun Bonferroni düzeltmesini kullanarak, düzeltilmiş p değerini hesapla.
4. En küçük düzeltilmiş p değerine sahip X tahmin edici değişkenini seç (en önemli olan). Bunun p değerini önceden tanımlanmış alfa düzeyi $\alpha_{böl}$ ile kıyasla.
 - a. Eğer p değeri, $\alpha_{böl}$ değerinden küçük veya eşit ise düğümü X'in kategori kümesini temel alarak böl.
 - b. Eğer p değeri, $\alpha_{böl}$ değerinden büyük ise düğümü bölme. Bu düğüm uç düğümdür.
5. Ağaç büyütme sürecini durma kuralları görülene kadar sürdür (SPSS 2002).

3.9.6 Sinir ağları

Karar ağaçları gibi yeni nesil veri madenciliği yöntemlerinden birisi de; (yapay) sinir ağlarıdır. Karar ağaçları gibi kolay uygulanır ve anlaşılır olmasa da, diğer yöntemlerin çözüm sunamadığı pek çok probleme kolaylıkla uygulanıp; başarılı tahminsel modeller sunabilmektedir.

Sinir ağıları, insan beyninin yapısı temel alınarak oluşturulmuş bir hesaplama modelidir. Belirli ağırlıklara sahip birbirine bağlı pek çok basit işlem ünitesinden oluşan bir yapıya sahiptir (Berson *et al.* 2000).



Şekil 3.8 Bir gizli katmanlı, ileri beslemeli sinir ağı şeması

Sinir ağıları ifadesi modellerin ve öğrenme yöntemlerinin büyük bir sınıfını içine alarak geliştirilmiştir. Burada, çok geniş anlamda kullanılan ‘vanilya’ sinir ağı betimlenecektir ki, bazen tek gizli katmanlı geri-beslemeli ağ veya tek katmanlı algılayıcı olarak da anılır. Sinir ağılarını çekici kılan, büyülü ve gizemli kılan yanları vardır. Burada açıklanacağı gibi, bunlar sadece doğrusal olmayan istatistiksel modellerdir.

Sinir ağıları, tipik olarak Şekil 3.8’de gösterilen Ağ Diyagramı ile temsil edilen, iki-aşamalı regresyon veya sınıflandırma modelidir. Bu ağ, regresyon veya sınıflandırmanın ikisine de uygulanabilir. Regresyon için, $K=1$ olur ve tepede olmak üzere bir tek çıktı birimi Y_1 vardır. Bununla beraber, bu ağlar bağlantısız bir üslupla çoklu nicel gözlemler olarak da ele alınabilir, böylece genel durumla karşılaşmış olur.

K-sınıflı sınıflandırma için, k. birim modellemesi k sınıfının olasılığı olmak üzere tepede K birim vardır. Y_k $k=1, \dots, K$ olmak üzere k. sınıf için her biri 0-1 olarak kodlanmış K hedef ölçümü vardır.

Türetilen özellikler Z_m girdilerin doğrusal kombinasyonu olarak yaratılır ve ardından hedef Y_k , Z_m 'nin

$$Z_m = \sigma(\alpha_{0m} + \alpha_m^T X), m=1, \dots, M$$

$$T_k = \beta_{0k} + \beta_k^T Z, k=1, \dots, K \quad (3.14)$$

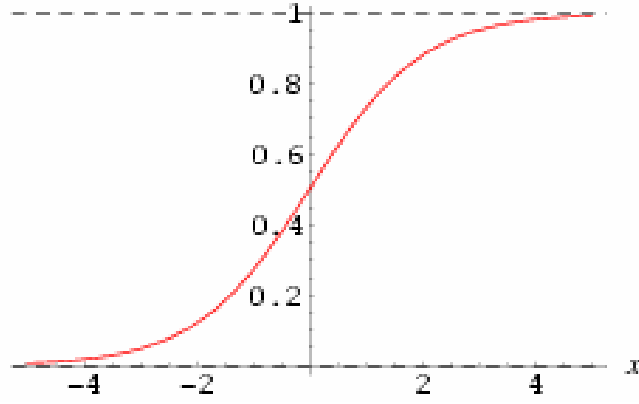
$$f_k(X) = g_k(T), k=1, \dots, K$$

$Z = (Z_1, \dots, Z_M)$ ve $T = (T_1, \dots, T_M)$ olmak üzere, doğrusal kombinasyonlarının fonksiyonu olarak modellenir.

Aktivasyon fonksiyonu $\sigma(v)$ genellikle sigmoid $\sigma(v) = 1/(1 + e^{-v})$ olarak seçilir, Şekil 3.9'da $1/(1 + e^{-v})$ 'nin şekli görülmektedir. Bazen $\sigma(v)$ için Gauss bazlı radyal tabanlı fonksiyonlar kullanılarak, radyal tabanlı fonksiyon ağı üretilir.

Şekil 3.8'deki sinir ağları diyagramları bazen gizli ve çıktı katmanlarındaki her birimde ek yan (bias) birimiyle beslenerek çizilebilir.

Ek bir girdi özelliği olarak sabiti '1' olarak düşünerek, bu yan birimi model (3.14)'te α_{0m} ve β_{0k} katsayılarını yakalar.



Şekil 3.9 Sigmoid fonksiyonunun grafiksel gösterimi

Çıktı fonksiyonu $g_k(T)$, T çıktı vektörünün son dönüşümüne imkan tanır. Regresyon için tipik olarak kimlik fonksiyonu $g_k(T) = T_k$ seçilir. K-sınıf sınıflandırmanın ilk çalışmalarında da aynı zamanda kimlik fonksiyonu kullanılmıştır ama ama daha sonra bu yerini softmax fonksiyonu

$$g_k(T) = \frac{e^{T_k}}{\sum_{l=1}^K e^{T_l}}$$

ifadesine bırakmıştır.

Bu kesinlikle, multilogit modelde kullanılan dönüşümün tam anlamıyla aynısıdır ve toplamı 1 olan pozitif tahminler üretir.

Ağın ortasında yer alan birimler, hesaplanan türetilmiş özellikler Z_m , gizli birimler olarak anılır, çünkü Z_m değerleri doğrudan gözlenmez. Genel olarak birden fazla gizli katman olur. Z_m orijinal girdi X'in temel genişlemesi olarak düşünülebilir, böylece sinir ağları bu dönüşümlerin girdi olarak kullanımıyla, standart doğrusal model veya doğrusal multilogit modeldir.

Belirtilmelidir ki, eğer σ kimlik fonksiyonu ise girdilerde bütün model doğrusal modele daralacaktır. Böylece sinir ağırları hem regresyon hem de sınıflandırmanın her ikisi için de doğrusal modelin doğrusal olmayan genelleştirilmesi olarak düşünülebilir. Doğrusal olmayan σ dönüşümünü dikkate alarak, doğrusal modeller sınıfı büyük ölçüde genişleyecektir. Şekil 3.9'da sigmoid aktivasyonun α_m normuna bağlı değeri görülmektedir ve eğer $\|\alpha_m\|$ çok küçük olursa, birim açıkça aktivasyon fonksiyonunun doğrusal kısmında işlem görecektir (Hastie *et al.* 2001).

3.9.7 Birliktelik (İlişki) kuralları

Birliktelik kuralı analizi ticari veri tabanlarının madenciliğinde gittikçe popüler hale gelen bir araçtır. Hedef, $X = (x_1, x_2, \dots, x_p)$ değişkenlerinin ortak değerleri en sık gerçekleşenleri bulmaktır. Çok sık olarak market sepet analizi denilen iki değerli $X_j \in \{0,1\}$ verilere uygulanmaktadır. Bu kapsamda, gözlemler bir mağazada gerçekleşen satış işlemleridir. Değişkenler mağazada satılan tüm ürünleri temsil etmektedir. i gözlemi için, her bir X_j değişkenine iki değerden biri atanır; öyleki, j . ürün işlemin bir parçası olarak satıldıysa $x_{ij} = 1$, satılmadıysa $x_{ij} = 0$ olur. Bu değişkenler birlikte satılan ürünleri temsil etmektedir. Bu enformasyon raflardaki stoklar, satış promosyonlarında çapraz pazarlama, katalog tasarımı ve satınalma örüntülerini temel alan müşteri segmentasyonunda çok kullanışlıdır.

S_j j . değişkenin bütün mümkün değerlerinin kümesini temsil etsin (destek) ve $S_j \subseteq S_j$ bu değerlerin bir alt kümesi olsun. Hedef, değişken değerleri S_1, \dots, S_p 'nin altkümelerini bulmaya çalışmaktır, öyleki her bir değişkenin görece olarak daha geniş olan kendi alt kümesinde eş anlamlı olarak gerçekleşmesi olasılığını

$$\Pr \left[\bigcap_{j=1}^p (X_j \in S_j) \right] \quad (3.15)$$

bulmaktır. Alt kümelerin kesişimi $\bigcap_{j=1}^p (X_j \in \mathcal{S}_j)$ birleşik kural olarak adlandırılır. Nicel değişkenler için \mathcal{S}_j alt kümeleri bitişik aralıklardır, nitel değişkenler için alt kümeler açıkça tasvir edilmektedir. Belirtmek lazımki, eğer \mathcal{S}_j alt kümesi gerçekten bütün değerler kümesi için $\mathcal{S}_j = \mathcal{S}_j$ ise, X_j değişkeni (3.15) kuralında görünmeyecektir (Hastie *et al.* 2001).

i. Market Sepet Analizi

Market sepet analizinde ifade (3.15) ile verilen genel yaklaşım oldukça başarıyla uygulanmaktadır. Ancak, çok büyük ticari veri tabanları için ($p \approx 10^4, N \approx 10^8$) için market sepet analizini uygulamak mümkün değildir. İfade (3.15) için pek çok basitleştirme yapmak gerekmektedir. Öncelikle, sadece iki çeşit alt küme göz önünde bulundurulmalıdır; $\mathcal{S}_j, \mathcal{S}_j = \mathcal{V}_{0j}$ olacak biçimde X_j 'nin bir tek değeri olarak oluşturulabilir veya $\mathcal{S}_j = \mathcal{S}_j$ olacak biçimde X_j 'nin bütün kümenin değerlerinden oluşturulur. Bu, $J \subset \{1, \dots, p\}$ tamdeğerlerinin alt kümesinin bulunmasında (3.15)'deki problemi basitleştirmekte ve uygun değerler \mathcal{V}_{0j} 'ler, $j \in J$ olacak biçimde

$$\Pr \left[\bigcap_{j \in J} (X_j = \mathcal{V}_{0j}) \right] \quad (3.16)$$

genişletilmektedir.

(3.16)'ı kukla değişkenler tekniğini kullanarak ihtiyaç duyulan sadece iki değerli değişkenlere dönüştürmek mümkündür. Burada, \mathcal{S}_j desteğinin her bir X_j değişkeni için sonlu olduğu varsayılmaktadır. Belirli biçimde, oluşturulan Z_1, \dots, Z_k

değişkenlerinin yeni kümesi, böyle bir değişken her bir v_{ij} değerleri her bir orijinal değişkenler X_1, \dots, X_p ulaşılabilir. Kukla değişkenlerin sayısı K ,

$$K = \sum_{j=1}^p |S_j|$$

olacak biçimde, $|S_j|$ ise X_j ile ulaşılabilen bağımsız değerlerin sayısıdır. Eğer, atanacak Z_k değişkeni ile ilişkili ise her bir kukla değişken $Z_k = 1$ değerini alacaktır; aksi takdirde $Z_k = 0$ olacaktır. Bu (3.16)'ı $K \subset \{1, \dots, K\}$ olmak üzere tamsayıların alt kümesine

$$\Pr \left[\bigcap_{k \in K} (Z_k = 1) \right] = \Pr \left[\prod_{k \in K} Z_k = 1 \right] \quad (3.17)$$

biçimde genişletecektir. Bu market sepet probleminin temel formülasyonudur. K kümesi 'nesne kümesi' olarak adlandırılır. Nesne kümesi içindeki Z_k değişkenlerinin sayısına 'hacim' adı verilir (belirtilmeli ki, hacim p 'den büyük olamaz). (3.17)'nin tahmin edilen değeri veri tabanındaki gözlemlerin fraksiyonundan alınmaktadır; öyleki, (3.17)'deki birlikte gerçekleşme doğru ise

$$\hat{\Pr} \left[\prod_{k \in K} (Z_k = 1) \right] = \frac{1}{N} \sum_{i=1}^N \prod_{k \in K} z_{ik} \quad (3.18)$$

olacaktır. Burada z_{ik} , Z_k 'nin i . gözleminin değeridir. Buna K nesne kümesinin 'destek' veya prevelansı $T(K)$ denir. i gözlemi için $\prod_{k \in K} z_{ik} = 1$ ise nesne kümesi K 'nin 'kapsadığı' söylenir.

Birliktelik kuralı madenciliğinde en küçük destek sınırı t tanımlanır ve Z_1, \dots, Z_k değişkenlerinden biçimlenmiş bütün K_i nesne kümelerinden desteğiyle bu alt sınır t 'den büyük olanlar

$$\{K_i | T(K_i) > t\} \quad (3.19)$$

için veri tabanında aranır.

ii. Önsel (Apriori) Algoritması

Birliktelik kuralı belirlemek için çözüm (3.19) çok büyük veri tabanlarında t eşliğini sağlayan mümkün hesaplamalarla ortaya çıkar, böylece (3.19) bütün mümkün 2^k nesne kümelerinin sadece küçük bir fraksiyonundan oluşur. Önsel algoritma, ifade (3.19)'u çözmek için veriyi geçen küçük bir numarayla boyutun lanetini aşacak bir durumdan yararlanılmıştır. Belirli biçimde, verilmiş destek eşiği t için; $|\{K | T(K) > t\}|$ kardinali görece olarak küçüktür (Agrawal *et al.* 1996, Hastie *et al.* 2001)

K 'daki nesnelerin alt kümelerinden oluşan herhangi bir nesne kümesi L , K 'dan büyük veya eşit desteğe sahip olmalıdır. $L \subseteq K \Rightarrow T(L) \geq T(K)$

İlk geçiş, verilerden her tek nesne kümesinin desteğini hesaplamaktır. Bunlardan, desteği eşikten küçük olanlar çıkarılacaktır. İkinci geçiş, ilk geçişten kalan tek nesnelerin çiftlerinden oluşan iki hacmine sahip bütün nesne kümelerinin desteğini hesaplamaktır. Bir başka deyişle, $|K|=m$ ile sık tekrarlanan bütün nesne kümelerini türetmek, bütün m eski nesne kümelerinden sadece $m-1$ hacmindeki sık tekrarlanan adayları göz önüne almak gerekmektedir. Böyle hacmi 2 olan nesne kümelerinden desteği eşikten küçük olanlar çıkarılacaktır. Veri geçişleri bütün aday kurallardan bir önceki geçişte tanımlanan eşikten küçük desteği olanlara kadar devam edecektir. Önsel

algoritma her bir $|K|$ değeri için tek bir geçişe ihtiyaç duymaktadır. Eğer veri yeterince seyrek (veya t eşiği yeterince büyük) ise büyük veri setleri için anlamlı bir sürede süreç sona erecektir.

Bu stratejinin parçası olarak hızı ve yakınsamayı artırmak için pek çok ek hile vardır (Agrawal *et al.* 1996). Önsel algoritma veri madenciliği teknolojisindeki önemli gelişmelerden birini temsil etmektedir.

Önsel algoritma ile dönüştürülen her yüksek destek (3.19)'daki nesne kümesi K , 'birliktelik kurallarının' kümesine dönüşecektir. $k \in K$ olmak üzere, Z_k nesnelere $A \cup B = K$ olacak biçimde iki ayrık alt kümeye ayrılacaktır ve

$$A \Rightarrow B \quad (3.20)$$

yazılabilir.

İlk nesne kümesi A öncül ve ikinci B ise ardıl olarak adlandırılmaktadır. Birliktelik kuralları veri tabanındaki öncül ve ardıl nesne kümelerinin prevalansını temel alan pek çok özelliğe göre tanımlanır. Kuralın desteği $T(A \Rightarrow B)$, öncül ve ardılın birleşimindeki gözlemlerin fraksiyonudur, öyleki, türetilen K nesne kümesinin desteğidir. (3.18) rasgele seçilen market sepetinde iki nesne kümesinin eş anlı gözlenmesi olasılığı $\Pr(A \vee B)$ 'nin tahmini olarak görülür. Kuralın 'güven' veya 'tahmin edilebilirlik'i $C(A \Rightarrow B)$ desteğinin, öncülünün desteğine bölünmesiyle,

$$C(A \Rightarrow B) = \frac{T(A \Rightarrow B)}{T(A)} \quad (3.21)$$

biçiminde elde edilir. Öyleki, $\Pr(B/A)$ 'nın tahmini olarak görülebilir. $\Pr(A)$ notasyonu sepette gerçekleşen A nesne kümesinin olasılığı, $\Pr(\prod_{k \in A} Z_k = 1)$ için kısaltmasıdır.

‘Beklenen güven’ ardılın desteği $T(B)$ olarak tanımlanır, öyleki koşulsuz olasılık $\Pr(B)$ ’nin tahminidir. Son olarak, ‘kaldıraç’, güvenin, beklenen güvene bölünmesiyle elde edilen kural olarak,

$$L(A \Rightarrow B) = \frac{C(A \Rightarrow B)}{T(B)}$$

biçiminde tanımlanır.

Bu birliktelik ölçüsü $\Pr(A \text{ ve } B)/\Pr(A)\Pr(B)$ ’nin tahminidir.

Örnek olarak, varsayalımki nesne kümesi $K=\{\text{fıstık ezmesi, jöle, ekmek}\}$ olsun ve kuralı $\{\text{fıstık ezmesi, jöle}\} \Rightarrow \{\text{ekmek}\}$ olarak gözönüne alalım. Destek değeri 0,03 için bu kuralın anlamı fıstık ezmesi, jöle ve ekmek market sepetlerinin %3’ünde birlikte yer almaktadır. Bu kuralda güvenin 0,82 olması, fıstık ezmesi ve jöle satın alındığında %82’sinde ekmek alındığını ima etmektedir. Eğer ekmek bütün market sepetlerinin %43’ünde yer alıyorsa, $\{\text{fıstık ezmesi, jöle}\} \Rightarrow \{\text{ekmek}\}$ kuralı 1,95 kaldırıca sahip olacaktır.

Bu analizin hedefi 3.20’deki birliktelik kurallarını 3.21’deki destek ve güvenin yüksek değerleriyle üretmektir. Önsel algoritma bütün veri kümelerini 3.19’daki destek eşiği t ile tanımlanandan yüksek destekte dönüştürür. Güven eşiği bir kümedir ve 3.19’daki nesne kümelerinden biçimlenen bütün kurallar

$$\{A \Rightarrow B | C(A \Rightarrow B) > c\} \quad (3.22)$$

biçiminde bu değerden yüksek güvene sahip olmalıdır. $|K|$ hacmindeki her K nesne kümesi için $A \Rightarrow (K - A)$, $A \subset K$ formunda $2^{|K|-1} - 1$ kurallar vardır. Agrawal *et al.* (1996) önsel algoritmanın varyantını sunmuştur öyleki, (3.19)’daki çözüm nesne

kümelerinden biçimlenen bütün mümkün kurallardan hangi kuralların 3.22'deki güven eşiğinde yaşayabileceği hızla tanımlamaktadır.

Bütün analizin çıktısı,

$$T(A \Rightarrow B) > t \text{ ve } C(A \Rightarrow B) > c$$

kısıtlarını sağlayan (3.20)'deki birliktelik kurallarının koleksiyonudur.

Bunlar genellikle kullanıcı tarafından sorgulanan veri tabanlarında depolanmaktadır. Tipik bir talep, kuralları güven, kaldıraç veya desteğine göre sıralı biçimde göstermektir. Daha belirgin olarak, birisi böyle bir listeyi belirli öncül veya ardıl bir nesneye koşullu olarak isteyebilir. Örneğin, bir istek aşağıdaki gibi olabilir:

Buz pateninin ardıl olduğu, güveni %80'in üzerinde ve desteği %2'den fazla olan bütün işlemleri göster.

Bu talep, bu nesnelere (ardıla) ait enformasyonu sağlayabilir, öyleki buz pateni satışlarının tahminidir. Özel bir ardıl üzerine yoğunlaşmak problemi denetimli öğrenim çerçevesine sokmaktadır.

Birliktelik kuralları market sepeti ve benzerleri için çok büyük veri tabanlarını analiz etmede popüler bir araç olmaktadır. Öyleki, veri çok boyutlu olasılık tablosu biçiminde geldiğinde; Çıktı kolayca anlaşılabilir ve yorumlanabilir (3.17)'deki biçimde ardıl kurallar yapısındadır. Önsel algoritma büyük veri tabanlarına bu analizlerin uygulanmasında, diğer analizi tiplerine göre daha uyumludur. Birliktelik kuralları, veri madenciliği yöntemleri arasında en büyük başarıya sahiptir. Uygulamada verinin kısıtlayıcı yapısının yanısıra, birliktelik kurallarının başka kısıtlamaları da vardır. Hesaplamaların olabilirliğinde kritik olan (3.19)'da verilen destek eşiğidir. Çözüm nesne kümelerinin sayısı, hacmi ve veriden geçiş sağlayanların sayısıbu alt sınırın

düşmesiyle beraber üstel olarak artmaktadır. Böylece, yüksek güven veya kaldıraca sahip ama düşük destekli kurallar keşfedilemeyecektir. Örneğin, yüksek güvenli votka \Rightarrow havyar kuralı düşük düzeyde satışa sahip havyar ardılı nedeniyle açığa kavuşamayacaktır (Hastie *et al.* 2001).

3.9.8 Önemli bileşenler analizi

x 'in gerçek kovaryans matrisi olarak tanımlanan $\Sigma_x = E[xx']$ biliniyor olsun. Burada $E[.]$ beklenen değeri ve $'$ matris veya vektörün transpozunu gösterebilir. Σ_x tanımında x 'in ortalaması içerilmemiştir, çünkü sıfır olacağı varsayılmıştır. Uygulamada, veriden alınan örneklerden ortaya çıkan Σ_x 'in tahminiyle de çalışılabilir. Bu durumda bütün dağılım parametreleri örneklem tahminleriyle değiştirilmelidir. Özel olarak örneklem kovaryans matrisi, \bar{x} , x 'in N gözleminin örneklem ortalaması olmak üzere

$$\hat{E}_x = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})'$$

biçiminde gösterilir.

$\{z_i : i = 1, 2, \dots, n\}$, Σ_x 'in özvektörlerinin bir ortonormal kümesini gösterebilir ve $\{\lambda_i : i = 1, 2, \dots, n\}$ azalan sırada düzenlenmiş uygun özdeğerleri gösterebilir. Bununla beraber, $Z = z_1, z_2, \dots, z_n$ özvektörlerden yapılandırılan $n \times n$ ortogonal matris ve $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ özdeğerlerden yapılandırılan $n \times n$ diagonal matris olsun. $u = Z'x$ $n \times 1$ vektörü tanımlansın. Bir başka deyişle, u 'nun i . ögesi, $u_i = z_i'x$ doğrusal kombinasyondur. Çünkü, $u = Z'x$ ve Z ortogonal matristir ve x

$$\mathbf{x} = \mathbf{Z} \mathbf{u} = \sum_{i=1}^n \mathbf{z}_i u_i$$

biçiminde gösterilebilir.

Bu basit olarak, \mathbf{x} 'i göstermek için özvektörleri farklı (ortonormal) tabanda kullanmak olarak görülebilir. \mathbf{x} 'in \mathbf{z}_i özvektörü yönündeki bileşeni, \mathbf{x} 'in i . önemli bileşeni olarak gösterilen u_i 'dir. \mathbf{x} 'in bu ÖBA gösteriminin önemi için temel sonuçlar, $\{u_i : i = 1, 2, \dots, n\}$ önemli bileşenlerinin hepsi ilişkisiz ve u_i 'nin varyansı λ_i 'dir. Çünkü, özdeğerler azalan sırada düzenlenmiş ve bunun sonucu olarak \mathbf{x} 'in en büyük varyansa sahip bileşeni \mathbf{z}_1 yönündeki u_1 bileşenidir. \mathbf{x} 'in bir sonraki en büyük varyansa sahip bileşeni \mathbf{z}_2 yönündeki u_2 bileşenidir ve böyle devam etmektedir. Gerçekten, gösterilebilir ki ÖBA gösterimi, \mathbf{x} 'in ilk bileşen $u_1 = \mathbf{z}_1' \mathbf{x}$ 'den daha büyük varyansa sahip başka bir birim-norm doğrusal dönüşümü olmaması yönüyle optimal yapıdadır. Dahası, \mathbf{x} 'in bütün birim-norm doğrusal dönüşümleri u_1 ile ilişkisizdir, $u_2 = \mathbf{z}_2' \mathbf{x}$ 'den daha büyük ikinci varyansa sahip bileşende yoktur. Ardışık olarak devam edersek, \mathbf{x} 'in bütün birim-norm dönüşümleri $\{u_1, u_2, \dots, u_{i-1}\}$ ile ilişkisiz i . önemli bileşen 'den daha büyük varyansa sahip başka bir bileşen de yoktur. Böylece, \mathbf{x} 'in ÖBA gösterimi mümkün olduğunca fazla varyansı, en az sayıda (ilişkisiz) bileşene toplar (Ye 2003).

3.9.9 Diskriminant analizi

Sınıflandırma için karar teorisi, optimal sınıflandırma için sonsal (posterior) sınıfını bilmeye ihtiyacımız olduğunu bize söyler. Varsayalım ki $f_k(x)$, X 'in $G=k$ sınıfındaki sınıf-koşullu yoğunluğudur ve π_k , $\sum_{k=1}^K \pi_k = 1$ olmak üzere k sınıfının önsel olasılığı olsun. Bayes teoreminin basit bir uygulaması,

$$\Pr(G = k|X = x) = \frac{f_k(x) \pi_k}{\sum_{l=1}^K f_l(x) \pi_l} \quad (3.23)$$

ifadesini verir.

Görülmektedir ki sınıflandırmaya yetisi anlamında, $f_k(x)$ 'e sahip olmak, hemen hemen $\Pr(G = k|X = x)$ niceliğine sahip olmaya eşittir.

Varsayalım ki her bir sınıf yoğunluğunu çok değişkenli Gauss

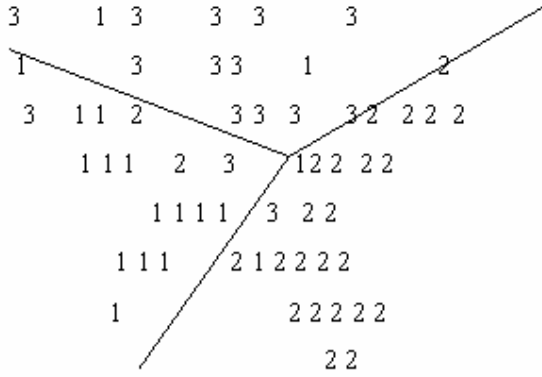
$$f_k(x) = \frac{1}{(2\pi)^{p/2} |\Sigma_k|^{1/2}} e^{-\frac{1}{2}(x-\mu_k)^T \Sigma_k^{-1} (x-\mu_k)} \quad (3.24)$$

olarak modellenmektedir.

Doğrusal Diskriminant Analizi, sınıfların kovaryans matrislerinin $\Sigma_k = \Sigma \forall k$ olduğu özel bir durumda ortaya çıkar. İki sınıf k ve l'nin kıyaslamasında, log-ratio'ya bakmak yeterlidir ve görülür ki,

$$\begin{aligned} \log \frac{\Pr(G = k|X = x)}{\Pr(G = l|X = x)} &= \log \frac{f_k(x)}{f_l(x)} + \log \frac{\pi_k}{\pi_l} \\ &= \log \frac{\pi_k}{\pi_l} - \frac{1}{2} (\mu_k + \mu_l)^T \Sigma^{-1} (\mu_k - \mu_l) + x^T \Sigma^{-1} (\mu_k - \mu_l) \end{aligned} \quad (3.25)$$

x'de lineer bir eşitliktir.



Şekil 3.10 Her biri Gausyen Dağılımdan gelen örnekler için Doğrusal Diskriminant Analizi'ne uydurulmuş modelin karar sınırları

Eşit kovaryans matrisleri, normalizasyon faktörlerini ve üsteki karesel kısmı iptal etmeye neden olur. Bu doğrusal log-odds fonksiyonu ima eder ki, k ve l sınıfları arasındaki karar sınırları – kümenin $\Pr(G = k|X = x) = \Pr(G = l|X = x)$ olduğu yerler- p boyutlu hiper düzlemde x 'de doğrusaldır. Bu, herhangi sınıf çiftleri için de doğrudur, böylece bütün karar sınırları doğrusaldır. Eğer \mathcal{R}^p sınıf 1, sınıf 2,... gibi sınıflandırılacak bölgelere bölünürse, bu bölgeler ayrı hiper düzlemler olacaktır.

Eşitlik (3.25)'den görülür ki, doğrusal diskriminant fonksiyonları

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k \quad (3.26)$$

$G(x) = \arg \max_k \delta_k(x)$ karar kuralı ile denk tanımlamaya sahiptir.

Uygulamada, Gausyen Dağılımların parametreleri bilinmez ve eğitim verisini kullanarak tahmin etmeye ihtiyaç duyulacaktır:

- $\hat{\pi}_k = N_k / N$, N_k k-sınıfındaki gözlemlerin sayısıdır;

- $\hat{\mu}_k = \sum_{g_i=k} x_i / N_k$;
- $\hat{\Sigma} = \sum_{k=1}^K \sum_{g_i=k} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T / (N - K)$.

Şekil 3.10 her biri Gaussyen olan 3 dağılımdan olan tahmin edilmiş karar sınırlarını göstermektedir (Hastie *et al.* 2001).

3.9.10 Kendi kendini düzenleyen haritalar

Bu yöntem K-ortalamalar kümelemenin kısıtlı versiyonu olarak görülebilir; öyleki, prototipler özellik uzayında bir veya iki boyutlu manifoldlar üzerine düşerler. Sonuç manifoldu aynı zamanda, orijinal yüksek-boyutlu gözlemlerin iki boyutlu koordinat sistemine indirgenerek haritalandığı kısıtlanmış topolojik haritaya işaret etmektedir.

Original SOM algoritması çevrimiçidir (online) – gözlemler anında işlenir – ve toplu işlem (batch) versiyonu sonra sonra önerilmiştir.

SOM, $m_j \in R^p$ olmak üzere K prototipin iki boyutlu üçgensel ızgarası(grid) ile göz önüne alınacaktır (Altıgen gibi diğer seçimlerde kullanılabilir). K prototipin her biri tamsayı koordinat çifti $l_j \in Q_1 \times Q_2$ ile parametrize edilir. Burada, $Q_1 = \{1, 2, \dots, q_1\}$, benzer biçimde Q_2 ve $K = q_1 \cdot q_2$ olmaktadır. Prototipler düzlem üzerine düzgün örüntüde dikilmiş düğmeler olarak düşünülebilir. SOM prosedürü düzlemi olabildiğince bükmeye çalışır böylece düğmeler olabildiğince veri noktalarına yaklaşır. gözlemler iki boyutlu ızgaraya indirgenerek haritalanarak model uydurulur.

x_i gözlemleri bir defada işlenir. R^p 'de Öklid mesafesi olarak, x_i 'ye en yakın m_j prototipi bulunur ve ardından m_j 'nin bütün komşuları m_k ,

$$m_k \leftarrow m_k + \alpha(x_i - m_k) \quad (3.31)$$

güncellemesiyle x_i 'ye doğru taşınır.

m_j 'nin komşuları, l_j ile l_k arasındaki mesafenin küçük olduğu bütün m_k 'lar olarak tanımlanır. En basit yaklaşım Öklid mesafesidir ve 'küçük' r eşikliğiyle tanımlanır. Bu komşuluk daima en yakın prototip m_j 'nin kendisini de içerir.

Belirtilmelidir ki, mesafe R^p özellik uzayından ziyade prototiplerin tamsayı topolojik koordinatlarının $Q_1 \times Q_2$ uzayında tanımlanmaktadır. (3.31) güncellemesinin etkisi prototipleri verinin daha yakınına taşımasıdır. Ama aynı zamanda prototipler arasında iki-boyutlu mekansal ilişkiyi kolaylaştırmayı sağlamasıdır.

SOM algoritmasının performansı öğrenme oranı α ve mesafe eşikliği r 'ye bağlıdır. Tipik olarak α 'nın 1,0'den 0,0'a birkaç binin üzerinde yinelemeden (her bir gözlem için) sonra düştüğü söylenebilir. Benzer biçimde r başlangıç değeri R 'den 1'e birkaç binin üzerinde yinelemeden sonra doğrusal olarak düşmektedir (Hastie *et al.* 2001).

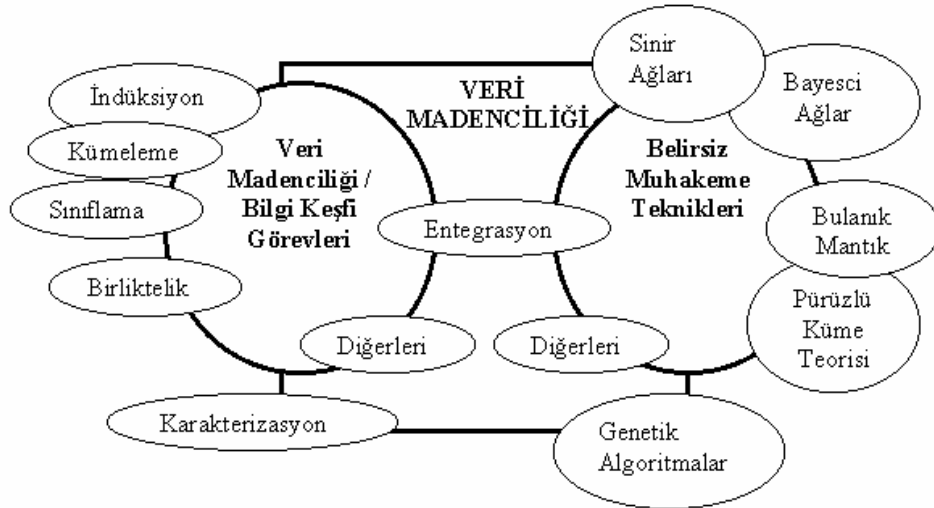
4. BULANIK VERİ MADENCİLİĞİ

Veri madenciliği ifadesi pekçok uygulamayla veriden sonuç çıkarımı üzerine vurgu yaparken, belirsizlik muhakemesi ifadesi belirsizlik sözkonusu olduğunda yeni bilginin çıkarımı anlamı üzerine odaklanır. Buna karşın, belirsizlik muhakemesi, veri madenciliği hedefi üzerinde etkin bir başarıya ulaşır. Aynı zamanda belirtmek lazım ki, bu belirsizlik muhakemesi teknikleri veri tabanı çevresinde analizler yapmak üzere de geliştirilmiştir.

Başlıca belirsiz muhakeme tekniklerini,

- Sinir ağları,
- Bayesci ağlar,
- Bulanık mantık,
- Pürüzlü küme teorisi,
- Genetik algoritmalar

olarak sıralamak mümkündür. Bu yöntemlerin veri madenciliği görevleri ile entegrasyonuna yönelik ilişki Şekil 4.1’de verilmektedir (Chen 2001).



Şekil 4.1 Veri madenciliği görevleri, belirsiz muhakeme teknikleri ve bunların entegrasyonu

Şekil 4.1’de görüldüğü gibi farklı amaçlara hizmet eden belirsiz muhakeme teknikleri vardır. Tezin uygulama alanının Hisse Senetleri Piyasası olduğu dikkate alındığında, belirsiz muhakeme tekniklerinden ihtiyaca en fazla cevap verecek olan yöntemin uygulamada tercih edilmesi gerekmektedir.

Hisse senetleri piyasasının temel varsayımlarından birisi, fiyat oluşumlarını önceden tahmin edilemiyor olmasıdır. Dolayısıyla, hisse senetleri piyasasında fiyatların oluşumu tamamıyla rasgeledir ve Etkin Piyasa Teorisi’ne göre fiyat oluşumlarının önceden tahmin edilemiyor olması, etkin bir piyasanın temel varsayımlarından birisidir (Kovalerchuk and Vityaev 2002). Ayrıca, Kaotik Piyasa Hipotezi’ne göre, rasgeleliğin de ötesinde hisse senetleri piyasası kaotik bir yapıya sahiptir (Kovalerchuk and Vityaev 2002). Bu noktadan hareketle, hisse senetleri piyasası için belirsizliğin de ötesinde bulanıklık esastır ve hisse senetleri piyasası üzerine yapılacak çalışmalarda Kesin Küme Kuramı zaman zaman yetersiz kalmakta; Bulanık Küme Kuramından yararlanmak gerekmektedir. Bu mantıktan hareketle, hisse senetleri piyasasını açıklamakta veya hisse senetleri piyasası üzerine çalışmakta Kesin Küme Kuramı yetersiz kalabilmektedir. Kesin Küme Kuramı yerine Bulanık Küme Kuramı ile çalışmak hisse senetleri piyasasının belirsizlik içeren yapısına daha uygun düşmektedir. Bu nedenle, Hisse Senetleri Piyasası’nın uygulama alanı olarak seçildiği tez çalışmasında Bulanık Küme Teorisi’nden faydalanılması gerekmektedir.

4.1 Bulanık Sonuç Çıkarım Kuralları

Bulanık sonuç çıkarımı, bulanık uzantılı eşitlikleri, bulanık operatörleri ve bulanık kurallarla bağlantılı operatörleri kullanan sonuç çıkarım yöntemini ifade etmektedir. Sonuç çıkarım süreci neticesinde, bulanık kuralları ve sağlanan girdi enformasyonunu temel alan yeni olgular elde edilir.

Bulanık mantık ve klasik iki değerli mantık arasındaki en önemli fark sonuç çıkarım teknikleri olmalıdır. Basit kural biçimini göz önüne alalım.

Eğer A ise C	A kuralın öncülüdür.
A'	A' durum veri tabanında eşleşen durumdur.
C'	C kuralın ardıdır. C' hesaplanan gerçek ardıdır.

İki değerli mantıkta, A öncülü ve A' durumu C ardılı gösteren C' çıkarımı ile tamamen aynı sonucu vermektedir. Diğer taraftan, bulanık mantıkta, kural gerçek ardıl C'nü, A'nün A öncülüne ait olduğu ölçüde verecektir. Çizelge 4.1'de görüldüğü gibi 4 çeşit kural gözönüne alınacaktır.

Çizelge 4.1 Sonuç Çıkarım Kuralı Çeşitleri

Öncül	Ardıl	Kural Türü
Kesin	Kesin	Kesin-Kesin
Kesin	Bulanık	Kesin-Kesin
Bulanık	Kesin	Bulanık-Kesin
Bulanık	Bulanık	Bulanık-Bulanık

Bulanık mantık, iki farklı anlamda görülebilir: (1) dar anlamda, bulanık mantık temel anlamda, yaklaşık muhakemenin mantığıdır ve (2) geniş anlamda, bulanık mantık, bulanık küme teorisi ile eşit genişliğe sahip olup, nesnelere sınıfları üyelikten, üyeye olmamaya tedrici olarak geçer. Geniş anlamıyla bulanık mantık pek çok dalıyla oldukça geniş bir teoridir.

Bulanık teoriye ilginin 1990'lardan itibaren yeniden dirilişinin nedeni birkaç faktöre bağlıdır:

Bulanık mantık, çok karmaşık veya kötü tanımlanmış ve matematiksel olarak ifade etmenin kolay olmadığı sistem davranışlarını yaklaşık ama etkin anlamda tasvir etmektedir. Sistemin karmaşıklığı arttığında, sistemin davranışları hakkında önemli

açıklamalar yapma yetisinde eşik, kesin ve özel karakteristiklerin tamamını kapsayan hale gelene kadar azalmaktadır.

Bir kişisel bilgisayar toplamaktan, bir hastaya teşhis koymaya kadar pek çok olayda süreci otomatikleştirmede insanın nasıl düşündüğü ve bilgisayarın nasıl programlandığı arasındaki fark engel oluşturmaktadır. Böylece, otomasyonda bulanık mantık bir adım ötesidir. Bu durum, oldukça fazla ticari uygulama ortaya çıkarmaktadır. Bu uygulamalar, çok karmaşık sistemleri geliştirmede gereken süreyi indirgemede bulanık mantık düşüncesini pekiştirmede yardımcı olmaktadır.

4.2 Veri Madenciliğinde Bulanık Mantık

Bulanık mantığın veri madenciliğindeki rolü üzerine pek çok açıklama vardır. Bunların bir kısmı aşağıda özetlenmektedir.

1. Bilgi tanelenmesi. Bulanık küme teorisi, bilgi tanelenmesi konusunda veri madenciliğine yardımcı olmaktadır. Bulanık kümeler, sürekli sınırlarda sözel enformasyon tanelerini yakalama kavramlarıdır. Veri madenciliğine katkı sağlayan teknolojiler arasında doğal olarak uyum sağlamaktadır. İnsanlar rakamlar yerine enformasyon taneleri üzerinde işlem yapmaktadır. Enformasyon tanelerinin hacmi verideki bakılması gereken kesin noktaları ima etmektedir ve enformasyonun belirlenmesi (tanelenmesi) düzeyinde ilişki kurulmasına yardımcı olmaktadır (Negoita 2000).

2. Daha iyi tolerans. Bulanık kümeler, sistem karmaşıklığını yönetilebilir yapmada belirsizlikten istifade etmektedir. Bulanık mantık sadece eksik, gürültülü veya kesin olmayan veriyle uğraşmada değil aynı zamanda, veriden, geleneksel sistemlere göre daha kabiliyetli ve düzgün performans sağlayan belirsiz modeller geliştirmekte de yardımcı olmaktadır. Bulanık sistemler belirsizliği tolere edebildiğinden ve verideki sözel belirsizlikleri düzeltmede yardımcı olduğundan kesin girdilerin geçerli olmadığı

veya çok pahalı olduđu durumlarda sađlam, gürültü-tahammüllü model veya tahmin fırsatı sunabilmektedir (Chen 2001).

3. Veri sınıflandırma. Bulanık mantık, bilgi tanelenedirmesiyle ilişkili olarak, yüksek düzeyde özetleme üzerine çalışma avantajı da sunmaktadır. Bu nedenle, bulanık mantık sınıflandırma yapan veri madenciliđi sistemleri için kullanışlıdır (Han and Kamber 2000). Bulanık mantık sınıflandırma için kullanıldığında özellik deđerleri bulanık deđgerlere dönüştürölür örneđin yaş özelliđi hesaplanan bulanık deđerlerle kesikli kategorilere (genç, orta yaşlı, ihtiyar) dönüştürölür.

4. Sinir ađlarıyla ilişkisi nedeniyle veri madenciliđine dolaylı katkısı. Bulanık küme teorisi tek başına ne makine öğrenimi ne de veri madenciliđi tekniđidir, buna karşın, bulanık küme teorisi yapay sinir ađlarında kullanılan bunların öğrenim yetenekleri için iyi bilinen ađırlıklarla yakından ilişkilidir (Chen 2001).

5. Belirsizliđe karşı bilgi keşfinin artan şansı. Bulanık küme teorisi aynı zamanda diđer veri madenciliđi ve belirsizlik muhakemesi yaklaşımlarıyla birlikte kullanılarak; belirsizliđe izin verip, gizli bilginin açığa çıkarılması şansını artırmaktadır. Bu nedenle, bulanık mantık tek başına veri madenciliđi yöntemi olmamasına karşın, veri madenciliđinde hala aktif rol oynayabilmektedir. Örneđin, birliktelik kurallarındaki destek ve güven kavramları belirsizlikle ilişkilidir ki temel anlamda olasılıktır ama bulanık olması gerekli deđildir. Bununla beraber, birliktelik kurallarını pek çok yönden bulanık birliktelik kurallarına genişletmek mümkündür: Kaymaksız süt alan genç kadınların diyet yiyecekler de alması muhtemeldir. Buradaki genç ve muhtemel bulanık kavramları bulanıktır (Chen 2001).

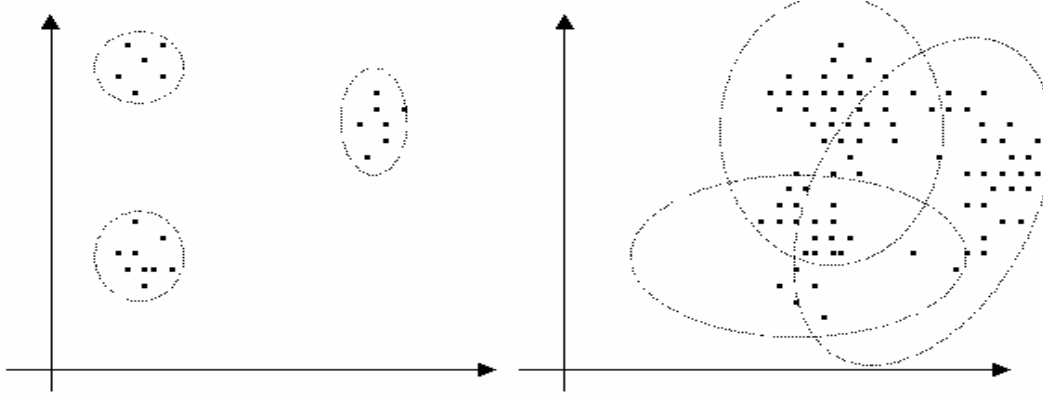
4.3 Bulanık Veri Madenciliđi Yöntemleri

Bulanık yaklaşımın matematiksel programlama ve istatistik yöntemlere uygulanmasına yönelik çalışmalar son on yılda oldukça yaygınlaşmaktadır. Ancak, veri madenciliđi

gibi henüz yeni gelişen bir alanda bulanık yaklaşımın çok fazla yaygın kullanıma sahip olmadığı görülmektedir.

4.3.1 Bulanık kümeleme analizi

Bulanık kümeleme veriyi çok iyi ayrılmış gruplara bölmez ama her noktaya kesin gruba ait olmaya yönelik bir üyelik verir. Şekil 4.2’de sol tarafta verilerin k-ortalamlar kümelemesi, sağdaki şekilde ise klasik bir bulanık kümeleme görülmektedir (Reinhold and Wolfgang 2000).



Şekil 4.2 K-ortalamlar kümeleme ve bulanık kümeleme

Örüntü tanımanın anahtar problemlerinden birisi, bir veri grubundaki kümeleri bulmaktır. Literatürde, pekçok bulanık kümeleme algoritması önerilmektedir. Bunlardan en yaygın kullanıma sahip olanları bulanık c-ortalamlar aşağıda sunulmaktadır (Wang 1997).

4.3.2 Hard ve bulanık c- bölünmeler

Varsayalımki $X = \{x_1, x_2, \dots, x_n\}$ veri kümesi verilmiş olsun. x_k herhangi bir öge olabilir, örneğin, $x_i \in R^p$. $P(X)$ X’in tüm altkümelerini içeren güç kümesi olsun. X’in

hard c-bölünümü $\{A_i \in P(X) : 1 \leq i \leq c\}$ ailesidir, öyleki $\bigcup_{i=1}^c A_i = X$ ve $1 \leq i \neq j \leq c$ için $A_i \cap A_j = \phi$ olur. Her bir A_i bir küme olarak gözlenecektir, böylece X'in $\{A_1, A_2, \dots, A_c\}$ bölünümleri c kümenin içinde yer alacaktır.

Hard c-bölünüm A_i 'deki x_k ögesinin karakteristik (üyelik) fonksiyonuna göre yeniden formüle edilebilir. Tanımlama gerekirse,

$$u_{ik} = \begin{cases} 1 & , x_k \in A_i \\ 0 & , x_k \notin A_i \end{cases} \quad (4.1)$$

$x_k \in X$, $A_i \in P(X)$, $i=1,2,\dots,c$ ve $k=1,2,\dots,n$ olmaktadır. Açıkça, $u_{ik} = 1$ olması x_k 'nin A_i kümesine ait olması demektir. Verilmiş u_{ik} değerine göre, X'in c-bölünümü eşsiz olarak tanımlanabilir ve tersi de doğrudur. u_{ik} 'lar aşağıdaki 3 koşulu sağlamalıdır:

$$u_{ik} \in \{0,1\}, 1 \leq i \leq c, 1 \leq k \leq n \quad (4.2)$$

$$\sum_{i=1}^c u_{ik} = 1, \forall k \in \{1,2,\dots,n\} \quad (4.3)$$

$$0 \leq \sum_{k=1}^n u_{ik} \leq n, \forall i \in \{1,2,\dots,c\} \quad (4.4)$$

ifade (4.2) ve ifade (4.3) birlikte, her bir $x_k \in X$ 'in yalnızca bir kümeye ait olacağı anlamına gelir. İfade (4.4) her bir A_i kümesinin en az bir ve en fazla n-1 veri noktası içermesini gerektirmektedir. u_{ik} 'ları $1 \leq i \leq c$ ve $1 \leq k \leq n$ olmak üzere c x n U matrisi

içinde toplayarak hard c-bölünümün matris gösterimini aşağıda tanımlandığı gibi elde edebiliriz.

Tanım: Hard c-bölünüm. $X = \{x_1, x_2, \dots, x_n\}$ herhangi bir küme, V_{cn} gerçek $c \times n$ $U = [u_{ik}]$ matrislerinin kümesi ve c , $2 \leq c \leq n$ olmak üzere bir sabit olsun. X için Hard c-bölünüm uzayı,

$$M_c = \{U \in V_{cn} \mid (4.2) - (4.4) \text{ geçerlidir} \} \quad (4.5)$$

kümesidir (Wang 1997).

4.3.3 Objektif fonksiyon kümeleme ve hard c-ortalamlar kümeleme

M_c veya M_{fc} uzaylarından optimal bölünümü nasıl seçmeliyiz? Bunun için üç çeşit yöntem bulunmaktadır: Hiyerarşik yöntemler, grafik-teorik yöntemler ve objektif fonksiyon yöntemleri. Hiyerarşik yöntemlerde, yeni kümeleri oluşturmada birleştirme ve bölme teknikleri bazı benzerlik ölçülerini temel alarak kullanılmaktadır; sonuç ise iç içe kümelerin hiyerarşisidir. Grafik-teorik yöntemlerde, X , benzerlik ölçüsüne göre ucundan birbirine bağlı düğümler kümesidir; kümeleme kriteri düğüm grupları arasında tipik olarak bağlantı ölçüsüdür. Objektif fonksiyon yöntemlerinde, bir objektif fonksiyon her bir c 'den oluşmuş küme adaylarının istenebilirliğini ölçmekte ve objektif fonksiyonun local miniması optimal küme olarak tanımlanmaktadır. Objektif fonksiyon yöntemleri kümeleme kriterinin en kesin formülasyonuna izin vermektedir.

En yaygın çalışılan objektif fonksiyon,

$$J_w(U, V) = \sum_{k=1}^n \sum_{i=1}^c u_{ik} \|x_k - v_i\|^2 \quad (4.6)$$

biçiminde tanımlanan toplam grup içi hata kareler toplamıdır. $U = [\mathbf{u}_{ik}] \in M_c$ veya M_{fc} , $V = (\mathbf{v}_1, \dots, \mathbf{v}_c)$ olmak üzere \mathbf{v}_i, A_i kümesinin merkezi olup,

$$\mathbf{v}_i = \frac{\sum_{k=1}^n \mathbf{u}_{ik} \mathbf{x}_k}{\sum_{k=1}^n \mathbf{u}_{ik}} \quad (4.7)$$

biçiminde tanımlanmaktadır.

Açıkçası, \mathbf{v}_i, A_i kümesindeki tüm noktaların ortalama (hard c-bölünüm) veya ağırlıklı ortalamasıdır (bulanık c-bölünüm). $\mathbf{x}_k, \mathbf{v}_i \in R^p$ olduğunu varsayalım. Eğer U hard c-bölünüm ise $J_w(U, V)$

$$J_w(U, V) = \sum_{i=1}^c \left(\sum_{\mathbf{x}_k \in A_i} \|\mathbf{x}_k - \mathbf{v}_i\|^2 \right) \quad (4.8)$$

biçiminde yeniden yazılabilir. Böylece, $J_w(U, V)$ 'ye neden toplam grup içi hata kareler toplamı denildiği açıklanmaktadır. $\mathbf{u}_{ik} \|\mathbf{x}_k - \mathbf{v}_i\|^2$ \mathbf{x}_k ile \mathbf{v}_i temsil eden maruz kalınan hata kareleri olduğundan aynı zamanda yerel yoğunluk ölçüsüdür. Eğer her bir hard küme A_i 'nin içindeki noktalar küme merkezi \mathbf{v}_i 'ye yakın olursa $J_w(U, V)$ küçük olacaktır.

J_w için optimal çift (U, V) 'yi bulmak kolay bir görev değildir. Zorluk sonlu ama büyük olan M_c 'nin boyutundan başlamaktadır. J_w 'nin yaklaşık minimasını bulmak için en popüler yöntemlerden birisi aşağıdaki hard c-ortalamalar algoritmasıdır (aynı zamanda ISODATA algoritması denilmektedir.) (Wang 1997).

Hard c-ortalamlar algoritması

Adım 1. Varsayalımki $X = \{x_1, \dots, x_n\}$, $x_i \in \mathbb{R}^p$ olmak üzere verilmiş n veri noktası olsun. $2 \leq c \leq n$ ve başlangıç $U^0 \in M_c$ olsun.

Adım 2. l iterasyon olmak üzere, $l=0,1,2,\dots$, için c ortalama vektörleri

$$v_i^l = \frac{\sum_{k=1}^n x_k u_{ik}^l}{\sum_{k=1}^n u_{ik}^l} \quad (4.9)$$

$u_{ik}^l = U^l$ ve $i=1,2,\dots,c$ olmak üzere hesapla.

Adım 3. $U^{(l)}$ 'yi $U^{(l+1)} = [u_{ik}^{(l+1)}]$ 'e;

$$U^{(l+1)} = \begin{cases} 1 & , \\ 0 & , \end{cases} \quad \begin{cases} \|x_k - v_i^{(l)}\| = \min_{1 \leq j \leq c} (\|x_k - v_j^{(l)}\|) \\ dy \end{cases}$$

kullanarak güncelle.

Adım 4. $U^{(l)}$ ile $U^{(l+1)}$ 'i kıyasla:

ε küçük bir sabit olmak üzere, eğer $U^{(l+1)} - U^{(l)} < \varepsilon$ dur,

aksi halde $l=l+1$ yap ve Adım 2'ye git.

Hard c-ortalamlar algoritması sezgisel bakış açısından oldukça anlamlıdır:

C hard kümeleri varsay (Adım 1),

Merkezlerini bul (Adım 2),

Veri ve küme merkezleri arasındaki hata karelerini minimize edecek biçimde küme üyeliklerini yeniden konumlandır (Adım 3),

J_w önemli düzeyde düşüp döngüye girince durdur (Adım 4). Hard uzay M_c kesikli ise, local minimumun tasarımı J_w 'ye göre tanımlanamaz. (4.9)'daki $\{\mathbf{v}_i^l\}$ 'yi hesaplama zorunluluğu her bir \mathbf{v}_i 'yi sıfıra eşitleyerek J_w 'nin gradientlerini ayarlamak üzerine kurulabilir.

4.3.4 Bulanık c-ortalamlar

Bulanık c-ortalamlar algoritmasında amaç

$$J_w(U, V) = \sum_{k=1}^n \sum_{i=1}^c (u_{ik})^m \|\mathbf{x}_k - \mathbf{v}_i\|^2 \quad (4.10)$$

ifadesini $m \in (1, \infty)$ ağırlıklandırma sabiti olmak üzere minimize ederek, $U = [u_{ik}] \in M_{fc}$ ve $\mathbf{v}_i \in \mathbb{R}^p$ için $V = (\mathbf{v}_1, \dots, \mathbf{v}_c)$ 'yi bulmaktır. Bu minimizasyon problemi için öncelikle gerekli koşul oluşturulacak ve ardından koşulu temel alan bulanık c-ortalamlar algoritması önerilecektir.

Teorem 3.1. $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ $\mathbf{x}_i \in \mathbb{R}^p$ olmak üzere verilmiş veri seti olsun. $c \in \{2, 3, \dots, n-1\}$ ve $m \in (1, \infty)$ olarak alınsın ve varsayalımki bütün $1 \leq k \leq n$ ve $1 \leq i \leq c$ için $\|\mathbf{x}_k - \mathbf{v}_i\| \neq 0$ olsun. $U = [u_{ik}]$ ve $V = (\mathbf{v}_1, \dots, \mathbf{v}_c)$, eğer sadece;

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_j\|}{\|x_k - v_i\|} \right)^{\frac{2}{m-1}}}, \quad 1 \leq i \leq c, 1 \leq k \leq n \quad (4.11)$$

ve

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m x_k}{\sum_{k=1}^n (u_{ik})^m}, \quad 1 \leq i \leq c \quad (4.12)$$

olmak üzere $J_m(U, V)$ için lokal minimumdur. Bulanık c-ortalama algoritması aşağıda verilmektedir (Wang 1997).

Bulanık c-ortalama algoritması

Adım 1. $X = \{x_1, \dots, x_n\}$ $x_i \in R^p$ olmak üzere verilmiş veri seti olsun.

$c \in \{2, 3, \dots, n-1\}$ ve $m \in (1, \infty)$ olarak düzenlensin ve $U^0 \in M_{fc}$ başlangıç alınsın.

Adım 2. $l = 0, 1, 2, 3, \dots$, için l . iterasyonda c ortalama vektörlerini

$$v_i^{(l)} = \frac{\sum_{k=1}^n (u_{ik}^{(l)})^m x_k}{\sum_{k=1}^n (u_{ik}^{(l)})^m}, \quad 1 \leq i \leq c$$

olarak hesapla

Adım 3. $U^{(l)} = [u_{ik}^{(l)}]$ yi

$$u_{ik}^{(l+1)} = \frac{1}{\sum_{j=1}^c \left(\frac{\|x_k - v_i\|}{\|x_k - v_j\|} \right)^{\frac{2}{m-1}}}, \quad 1 \leq i \leq c, 1 \leq k \leq n$$

ifadesini kullanarak $U^{(l+1)} = [u_{ik}^{(l+1)}]$ 'e güncelle

Adım 4. Eğer $\|U^{(l+1)} - U^{(l)}\| < \varepsilon$ ise dur, aksi takdirde, $l=l+1$ olarak Adım 2'ye git.

4.3.5 FANNY (Fuzzy Analysis) algoritması

Tezin özgün uygulaması olan Erken Uyarı Sistemi'nde kullanılan yöntemlerden birisi de Bulanık c-Ortalamalar Kümeleme algoritmalarından birisi olan FANNY algoritmasıdır.

FANNY algoritması diğer bulanık kümeleme yöntemleriyle kıyaslandığında aşağıdaki özelliklere sahiptir:

- Benzemezlik matrisini kabul eder,
- Küresel nesnelere kümeleme de daha dirençlidir (SPLUS 2001).

$U := (u_{iv})_{i=1, \dots, n; v=1, \dots, k}$ n veri noktası ve k küme sayısı olmak üzere üyelik fonksiyonunu gösterebilir. FANNY (Kaufmann and Rousseeuw 1990) algoritması, $d(x,y)$ mesafe ölçüsü olmak üzere;

$$\sum_{v=1}^k \frac{\sum_{i,j=1}^n u_{iv}^2 u_{jv}^2 d^2(x_i, x_j)}{2 \sum_{i=1}^n u_{iv}^2}$$

fonksiyonunu minimize etmeye çalışır.

Bulanık bölünümü değerlendirebilmek için, gruplara ayrılmanın iyiliği için bir ölçüye ihtiyaç duyulmaktadır. Dunn katsayısı

$$F_k(U) := \sum_{v=1}^k \sum_{i=1}^n \frac{u_{iv}^2}{n}$$

böyle bir ölçüdür.

Hiç bölünüm olmaması (tamamıyla bulanık) için $\frac{1}{k}$ ve kesin bölünüm için 1 arasında gerçekleşir. $[0,1]$ aralığına normalize edilirse,

$$\tilde{F}_k(U) = \frac{k F_k(U) - 1}{k - 1}$$

elde edilir. Sonuçları derecelendirebilmek için bu son eşitlik kullanılmaktadır (Theis and Weihs 1999).

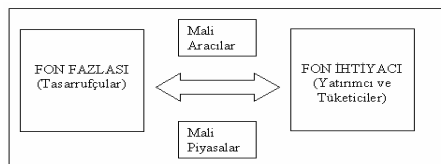
5. SERMAYE PİYASALARI

5.1 Kavram ve Tanımlar

Sermaye piyasası, ülkemizde yarım yüzyıllık bir durgunluktan sonra, 1980'lerin ikinci yarısından itibaren, yeniden gelişmeye başlamıştır. Sermaye piyasalarının işleyişi ve akışının anlaşılabilmesi için öncelikle bazı temel tanım ve kavramların verilmesi gerekmektedir. Bu tanım ve kavramlar aşağıda başlıklar halinde sunulmaktadır.

5.1.1 Mali sistem

Yatırım ve tasarruf kararlarının farklı birimlerce verildiği merkezi olmayan bir ekonomide tasarrufların yatırıma aktarılması işlevi mali sistem aracılığıyla gerçekleşir. Şekil 5.1'de görüldüğü gibi, bu sistem içinde bir yanda gelirinin tamamını tüketmeyen, fon fazlasına sahip kesim, diğer yanda ise gelirinin ötesinde harcama yapması gerektiğinden fon ihtiyacı olan kesim yer alır. İşte mali sistem, mali piyasalar ve bu piyasalarda yer alan mali araçlar ve araçlarla fon fazlasına sahip tasarruf eden kesimden fon ihtiyacı olan yatırımcı ve tüketicilere fonları kanallara eder. Genelde, tasarrufçu kesim hanehalklarından, fon talebi olan kesim ise devlet ve özel işletmelerden oluşur. Mali piyasalardan aktarılan fonların çok büyük bir bölümü devlet ve özel işletmelerin yatırım harcamalarına yönelir. Yatırım için büyük fonlara ihtiyaç duyan bu birimler, gerekli kaynakları tasarrufçulara getiri sağlamak karşılığında mali piyasalardan temin ederler. Dolayısıyla mali piyasalar gelecekteki tüketimle şimdiki tüketim arasındaki tercihin yapıldığı ve faiz oranlarının belirlendiği yerlerdir. Mali sistem iyi çalıştığı ölçüde kısa ve uzun dönemli kaynak dağılımının daha verimli olması mümkündür (Anonim 1989).



Şekil 5.1 Mali Sistem

5.1.2 Mali piyasalar

Bir ülkede fon kullananlar ile fon arz edenler arasında fon akımlarını düzenleyen kurumlar, akımı sağlayan araç ve gereçler ile bunları düzenleyen hukuki ve idari kurallardan oluşan yapıya mali piyasa denir. Tanımdan da anlaşılacağı gibi, mali piyasa para ve sermaye piyasalarından daha geniş ve bu piyasaları da kapsamına alan bir kavramdır.

Mali piyasa beş ana unsurdan oluşmaktadır:

- a) Tasarruf sahipleri (fon arz edenler),
- b) Yatırımcılar (fon talep edenler),
- c) Yatırım ve finansman araçları,
- d) Yardımcı kuruluşlar,
- e) Hukuki ve idari düzen olarak sıralanabilir.

5.1.2.1 Para piyasası

Kısa vadeli fon arz ve talebinin karşılaştığı piyasaya "para piyasası" denir. Para piyasasının tipik özelliği kısa vadeli fonlardan oluşmasıdır. Para piyasalarında vade genellikle bir yılı aşmaz.

Para piyasasından sağlanan fonlar kredi olarak işletmelerin dönen varlıklarının finansmanında kullanılır. Para piyasasının araçlarını ticari senetler; kaynaklarını çeşitli mevduat oluşturmaktadır.

Para piyasasının kendi içinde örgütlenmiş ve örgütlenmemiş para piyasası olarak da bir ayrımı yapılabilir. Örgütlenmiş para piyasası bir bankalar sistemidir. Çünkü işletmelerin nakit ihtiyacı çoğunlukla ticari bankalar tarafından karşılanmaktadır. Örgütlenmemiş

para piyasası banka sisteminin dışında kalan piyasadır. Bankalar dışındaki kişi ve kuruluşlar da bazen işletmelere kısa vadeli fon sağlarlar.

5.1.2.2 Sermaye piyasası

En genel tanımıyla sermaye piyasası orta ve uzun vadeli fon arz ve talebinin karşılaştığı piyasadır. Sermaye piyasası mali piyasa kavramından daha dar ve teknik bir nitelik taşır ve genellikle mali piyasa kavramı içinde yer alır.

Sermaye piyasasının tipik özelliği ve para piyasasından ayrıldığı en belirgin niteliği bu piyasanın orta ve uzun vadeli fonlardan oluşmasıdır. Bu vade bir yıldan fazla olmalıdır. Sermaye piyasasından sağlanan krediler genellikle, işletmelerin bina, makina ve teçhizat gibi duran varlıklarının finansmanında kullanılır.

Para piyasasında olduğu gibi, sermaye piyasasının kaynakları da tasarruf sahiplerinin birikimleridir. Sermaye piyasasının en önemli ve yaygın araçları hisse senetleri ve tahvillerdir (İMKB 2006).

5.1.3 Mali araçlar

Mali sistem içerisinde fonları tasarrufçu kesimden yatırımcı kesime yönlendiren çeşitli mali araçlar yer alır. Bu yardımcı kuruluşlar arasında öncelikle bankalar ve aracı kurumları, onların yanısıra sosyal güvenlik kuruluşları, sigorta şirketleri, menkul kıymet yatırım fonlarını ve ortaklıklarını ve bağımsız dış denetleme kuruluşlarını sayabiliriz (Anonim 1989).

5.1.4 Birincil ve ikincil piyasalar

Sermaye Piyasalarına bakıldığında diğer bir sınıflandırma ise birincil ve ikincil piyasalar şeklinde yapılabilir. Birincil piyasada daha önce ihraç edilmemiş yeni

menkul değerler, yani araçlar satılırken, ikincil piyasada daha önce çıkarılmış araçlar el değiştirir. Dolayısıyla tasarrufların yatırıma kanalize olması birincil piyasada gerçekleşir. İkincil piyasanın fonksiyonu esas olarak mali araçlara likidite ve pazarlanabilme özelliği sağlamaktır. İkincil piyasanın likidite ve pazarlanabilme görevlerini yerine getirememesi durumunda birincil piyasa doğrudan etkilenir ve mali piyasalar işlevlerini yerine getiremez olurlar. Diğer bir deyişle, birincil piyasanın sağlıklı işlemesi etkin ve verimli bir ikincil piyasanın varlığına bağlıdır. Örneğin hisse senedi piyasasında şirket tarafından, bir aracı kuruluş aracılığıyla, yeni hisse senetlerinin çıkarılarak yatırımcılara satılması bir birincil piyasa işlemidir. Öte yandan mevcut hisse senetlerinin borsada bir yatırımcıdan diğerine devri ikincil piyasada gerçekleşir. İkincil piyasadaki işlem şirketi doğrudan etkilemez; ancak bu piyasanın varlığı şirketin birincil piyasada yeni hisse senedi ihracını kolaylaştırır (Anonim 1989, Karan ve Karacabey 2003).

5.1.5 İkincil piyasa türleri

Menkul kıymetlerin alınıp satıldığı ikincil piyasalarda başlıca iki tür piyasa yapısından söz etmek mümkündür. Bunlardan birincisi menkul kıymetler borsalarıdır. Borsalar fiziksel olarak bir mekanda yer alan ve belli kişilerin (borsa üyelerinin) katılımıyla, önceden belirlenmiş kurallar çerçevesinde menkul kıymet alım-satımı yapılan yerlerdir. Borsalarda her menkul kıymet işlem görmez, yalnızca borsaya kote edilen kıymetler alınıp satılır. Bu özellikleriyle borsalar alım satımı yapılan varlığın el değiştirmesinde sürekliliği sağlar ve piyasadaki güven unsurunu pekiştirir. Borsada işlem görmeyen araçlar borsa dışı piyasada el değiştirir. Borsa dışı piyasa, ya da batıda bilinen adıyla, over-the-counter (tezgah üstü) pazar, ülke çapında yer alan aracı ve komisyoncuların oluşturdukları bir iletişim ağıdır. A.B.D. deki uygulamada bu piyasanın işleyişi birtakım kurallara bağlanmış ve gerekli bilgi akımını sağlayacak merkezi bir sistem oluşturulmuştur. Finans literatüründe geleneksel olarak borsa biçimindeki yapıya örgütlü (organize) piyasa denilmekte, böylece dolaylı biçimde borsa dışı piyasaların örgütsüz olduğu ima edilmekteydi. Ancak teknolojik gelişim ve ihtiyaçlar doğrultusunda tezgah üstü piyasanın belli bir

örgütlenmeye yöneldiği düşünülecek olursa, örgütlü piyasa tanımına borsa dışı piyasayı (tezgah üstü) da katmak gerekecektir.

Zaman içinde batıdaki menkul kıymetler piyasalarında gelişen bazı işlem türleri yeni piyasa yapılarının tanımlanmasına yol açmıştır. Bunlardan üçüncü piyasa olarak bilinen yapı, kurumsal yatırımcılar arasında borsaya kote menkul kıymetlerin bloklar halinde borsa dışında alınıp satılmasını içerir. Üçüncü piyasadaki işlemler bir komisyoncunun aracılığıyla gerçekleşir. Alım-satım bloklar halinde yapıldığından ve taraflar komisyoncudan yan hizmetler talep etmediklerinden yüklü işlem masrafları önemli ölçüde azaltılmış olur. Diğer bir piyasa türü ise menkul kıymetlerin alan ve satan taraflar arasında aracısız, doğrudan el değiştirmesidir ve dördüncü piyasa olarak adlandırılmaktadır. Üçüncü ve dördüncü piyasalar örgütlü olmayan piyasa tanımına girmektedir. Bu piyasalar menkul kıymet alım satımında kurumsal (kollektif) yatırımcıların rolünün artmasıyla birlikte gelişmiş ve esas olarak yüksek işlem masraflarını bertaraf etmek amacıyla ortaya çıkmışlardır (Anonim 1989).

5.2 Bankaların ve Bankacılık Sisteminin Türkiye'deki Yeri

Türk mali piyasalarının temel özelliği, bankaların mali piyasaların her alanında ağırlıklı olarak faaliyet göstermektedir.

Bankaların sermaye piyasalarındaki fiili ağırlıklarının temelinde yatan unsurlara bakıldığında ilk göze çarpan, bazı gelişmiş ülkelerde de gözlendiği gibi, bankaların mali piyasalardaki tarihsel üstünlükleridir. Sermaye piyasalarının geliştirilmesi yolundaki kapsamlı çalışmaların başlangıcı 1981 yılıdır. Bu nedenle sermaye piyasaları ve bu piyasalarda aracılık faaliyetini üstlenen banka dışı mali aracı kuruluşlar Türkiye için oldukça yenidirler. Bu durumda bu yeni piyasaların, araçların tasarruf sahibine tanıtılması ve banka dışı mali aracı kuruluşlarla tasarruf sahipleri arasında güven ilişkilerinin oluşturulması zaman almaktadır. Halbuki bankalar uzun bir süredir mali piyasalarda faaliyet göstermektedirler ve faaliyetlerinin niteliği tasarruf sahipleri tarafından bilinmektedir.

Tasarruf sahiplerinin ve fon talep eden kesimlerin alışkanlıklarının bugünden yarına değişmediği dikkate alındığında bunlarla karşılıklı güvene dayalı, denenmiş bir ilişki içinde olan bankaların yeni gelişmekte olan sermaye piyasalarında da avantajlı durumda olmalarının nedeni ortaya çıkmaktadır. Bunun yanısıra bankaların yaygın bir dağıtım şebekesine sahip bulunmaları da onlar için ek bir avantaj olmaktadır.

Bankaların sermaye piyasalarındaki bu tarihsel avantajı yalnızca Türkiye için söz konusu değildir. Mali piyasalar içinde sermaye piyasalarının bankacılık kesiminden sonra gelişmeye başladığı her ülke için aynı durumdan bahsedilebilir. ABD ve Japonya gibi, mali piyasalarının bankacılık ve sermaye piyasası işlemlerinin ayrımı ilkesine göre örgütlendiği ülkelerde bile, başlangıçta aynı olgunun varlığından söz edilebilir. Ancak, anılan ülkelerde, düzenleyici otoriteler bir seçim yaparak bankaları tarihsel avantajlarına karşın, sermaye piyasası faaliyetlerinin dışında tutmayı seçmişlerdir (Anonim 1989).

5.3 Hisse Senetleri Piyasası

Türkiye’de en bilinen ve yaygın kullanılan piyasa olan Hisse Senetleri Piyasası (HSP) hakkındaki detaylar aşağıda başlıklar halinde sunulmaktadır.

5.3.1 Hisse senetleri piyasasının gelişimi

26 Aralık 1985 tarihinde faaliyete geçen İstanbul Menkul Kıymetler Borsası'nda ilk hisse senedi işlemleri, 3 Ocak 1986 tarihinde gerçekleşmiştir. Başlangıçta, 36 Borsa Üyesinin işlem yapma yetkisine sahip olduğu piyasada, 41 anonim şirketin hisse senedi işlem görmekteydi.

İMKB'nin ilk yıllarında fiyatların yazıldığı işlem panolarının kullanımı ile “tek fiyat-çok fiyat” veya “Toplu Açılış Fiyatı Yöntemi” olarak adlandırılan sistemle işlemler yürütülmekteydi. Bu sistem Kasım 1988’e kadar geçerli oldu. Daha sonra, 1993 yılı

Aralık ayına kadar yine panolara emirlerin aktarılmasıyla “Çok Fiyat, Sürekli Müzayede Sistemi” uygulanmasına geçildi. 1993 yılı Aralık ayında ilk olarak düşük işlem hacimli 50 hisse senedi bilgisayar ortamında işlem görmeye başladı. Kasım 1994’e kadar aşamalı olarak tüm hisse senetleri bilgisayarlı sisteme alındı ve pano kullanımı Kasım 1994’de tamamiyle sona erdirildi. Böylece, Kasım 1988 ile Kasım 1994 arasında panolar aracılığıyla uygulanan “Çok Fiyat, Sürekli Müzayede Sistemi”, Kasım 1994’ten sonra bilgisayar aracılığıyla işlemeye devam etmiştir (İMKB 2006).

5.3.2 Borsa üyeleri ve üye temsilcileri

Borsa’ya Sermaye Piyasası Kurulu (SPK) tarafından aracılık faaliyetinde bulunmak üzere yetki belgesi verilmiş olan aracı kuruluşlar (aracı kurumlar ve bankalar) üye olabilir.

Borsa üyeleri;

- a. SPK'dan ‘alım satıma aracılık yetki belgesi’ almış olan **aracı kurumlar**
- b. Bankalar Kanunu'na göre Türkiye'de faaliyette bulunan ve SPK'dan yetki belgesi almış olan **bankalardan** oluşmaktadır.

Borsa’da sadece Borsa üyeleri işlem yapabilirler. Borsada alım satım yapmak isteyen yatırımcılar emirlerini Borsa üyeleri aracılığıyla Borsaya iletirler.

İMKB'nin Hisse Senetleri Piyasası'nda işlem yapma yetkisine sahip olan her üyenin üç harften oluşan bir "Üye Kodu" vardır. Üye temsilcileri, kendi bilgisayarlarından emir gönderirken sistem tarafından üye temsilcilerinin bağlı bulunduğu kurumun üye kodu sisteme emirle birlikte gönderilir.

Gönderilen emirler üzerinde işlem gerçekleştikten sonra, gerçekleşen işlemler dosyasından işlemin tarafları üye kodlarına bakarak görülebilir. Üyelerin, unvan

değişikliğine bağlı olarak, üye kodları yılda bir defa ve sadece yılın ilk gününden itibaren geçerli olmak üzere değiştirilebilir (İMKB 2006).

5.3.3 Hisse senetlerinin borsada işlem görmesi

Hisse Senetleri Piyasası'nda birincil ve ikincil piyasa işlemleri gerçekleştirilir.

i. Birincil piyasa

Hisse senetlerini halka arz eden şirketler ile alıcıların, yani tasarruf sahiplerinin, doğrudan doğruya karşılaştıkları piyasadır. Birincil piyasada uzun vadeli fonların, tasarruf sahibinden ihraççı şirkete akışı söz konusudur ve birincil piyasada yapılan hisse senedi satışları sonucunda şirkete nakit sermaye girer.

ii. İkincil piyasa

Menkul kıymetleri ihraçtan (menkul kıymeti çıkartan şirket tarafından yapılan satıştan) alanlar, bunları tekrar paraya çevirmek istediklerinde, hisse senetlerinde hiçbir zaman, tahvillerde ise vadeden önce bunları ihraç eden kuruluşa müracaat edemezler. İkincil piyasa bu durumdaki menkul kıymetlerin paraya çevrilmesini sağlayan piyasadır ve menkul kıymet borsaları bu tür piyasalara en iyi örnektir.

İkincil piyasada el değiştiren menkul kıymetlerden sağlanan fonların bunları çıkaran şirketle bir ilgisi yoktur. İkincil piyasa, menkul kıymetlerin likiditesini arttırarak birincil piyasaya talep yaratmakta ve gelişmesini sağlamaktadır (İMKB 2006).

5.3.4 Hisse senetleri piyasası'nda fiyatların oluşum yöntemi

İMKB hisse senetlerinin alım satımında 'Çok Fiyat' yöntemi kullanılır. Çok fiyat yöntemi; alım satım emirlerinin Borsa tarafından belirlendiği şekilde her şirket için bilgisayarlı alım satım sistemine kaydedildiği, sürekli müzayedede sistemi ile çalışır.

Sisteme gönderilen alış ve satış emirlerinin seans boyunca farklı fiyat seviyelerinde eşleşebilmesi, dolayısıyla aynı seans içinde farklı fiyatların oluşması nedeniyle bu yönteme "Çok Fiyat Yöntemi" denir. Çok fiyat yönteminin uygulanmasında, ilgili hisse senedinin işlem gördüğü sistemde, bir hisse senedini satın almak veya satmak isteyen üye alış veya satış emrini o seans için tanımlanan marjlar ve fiyat adımlarına uymak koşuluyla istediği fiyattan sisteme girebilir.

Borsada alım satımlar peşin esası üzerinden yapılır. Kıymetin ve bedelinin teslimi Yönetmelikte belirtilen süre içinde yapılır. Yerli ve yabancı menkul kıymetler TL üzerinden alınıp satılır (İMKB 2006).

5.3.5 Hisse senetleri piyasası'nın işleyişi ile ilgili tanımlar

a) İşlem miktarı

Bir piyasada, bir seansta ya da belli bir dönemde alınıp satılan (el değiştiren) menkul kıymet adedidir. Borsa'da, her işlem günü, her iki seansın işlemleri için ayrı ayrı hisse senedi bazında işlem miktarı yayınlanır. Farklı pazarlarda gerçekleşen işlem miktarları toplandığında, İMKB'nin o günkü işlem miktarı ortaya çıkar. Birincil piyasadaki işlemler ayrı olarak gösterilir.

b) İşlem hacmi

Hisse senedinde gerçekleşen her sözleşmedeki hisse senedi sayıları ile işlem fiyatlarının çarpılması sonucu bulunan miktardır. Tüm hisse senetlerinin işlem hacimleri toplamı, Hisse Senetleri Piyasasının toplam işlem hacmini oluşturur.

c) İşlem birimi

Bir sermaye piyasası aracının, kendisi ya da katları ile işlem yapılabilecek asgari sayısını ya da değerini ifade eder. Hisse senetleri piyasası'nda işlem birimi olarak lot kullanılır. 1 lot 1 adet hisse senedi veya 1 YTL nominal değerli hisse senedine denktir. Fon Pazarında da 1 lot 1 adet Borsa Yatırım fonu katılma belgesinden oluşmaktadır.

d) Ağırlıklı ortalama fiyat

Bir sonraki seansta uygulanacak baz fiyatın hesaplanmasına esas teşkil eden, hisse senedinin miktar ağırlıklı fiyatıdır. Her hisse senedi için her seans öncesinde bir önceki seans işlemleri dikkate alınarak ağırlıklı ortalama fiyat hesaplanır.

Hisse senetlerinin ağırlıklı ortalama fiyatları hesaplanırken, değişik fiyattan ve miktarlardan gerçekleşen ve **fiyatı tescil edilen** normal emirler dikkate alınır (Fiyatın tescil edilmesi için işlemin lot olması ve gerçekleşmiş işlemin özel emir şeklinde olmaması gerekir.). İşlem günü sonunda, her hisse senedi için aşağıdaki formül kullanılarak hesaplanır:

$$\text{A.O.F.} = \frac{\sum_{i=1} (M_i \times F_i)}{\sum_{i=1} M_i}$$

M_i = Gerçekleşen emrin içerdiği hisse senedi miktarı
 F_i = Gerçekleşen emrin satış fiyatı

e) Baz fiyat

Bir hisse senedinin seans içinde işlem görebileceği üst ve alt fiyat limitlerinin ve fiyat adımlarının belirlenmesine esas teşkil eden fiyattır. Bu fiyat, menkul kıymetin en son işlem gördüğü seanstaki ağırlıklı ortalama fiyatının en yakın fiyat adımına yuvarlanması ile elde edilir.

f) Referans fiyat

İşlem görülebilecek en üst ve en alt fiyat limitlerinin belirlenmesinde esas teşkil etmeyen, rüçhan hakkı kuponlarının ilk işlem gördüğü seansa işleme baz fiyatla açılmaması nedeniyle, yatırımcılar açısından kuponun yaklaşık değerine ilişkin olarak bir referans olarak kullanılmak üzere, eski hisse senedinin bölünmeden önceki en son işlem gördüğü seanstaki ağırlıklı ortalama fiyatı ve bedelli artırım oranı kullanılarak bulunan bir fiyat türüdür.

g) Fiyat adımları

Fiyat adımı, her hisse senedi fiyatı için bir defada gerçekleşebilecek en küçük fiyat değişimidir. Hisse senetlerinin fiyat adımları, baz fiyatlarına göre belirlenir. Baz Fiyat Aralığı – Fiyat Adımı Tablosu adı verilebilecek aşağıdaki tabloda çeşitli baz fiyat aralıkları ve bu aralıkların her biri için, belirli seviyelerdeki baz fiyat artışına paralel olarak artan bir biçimde, adımlar tanımlanmıştır. Emir girilirken bu fiyat adımlarına uymak zorunludur.

Fiyat adımlarının, girilen emrin fiyatının Baz Fiyat Aralığı-Fiyat Adımı tablosundaki aralığa göre değil, hisse senedinin baz fiyatının bulunduğu fiyat aralığına göre belirlendiği unutulmamalıdır.

h) Fiyat Değişme Sınırları

Borsada işlem gören menkul kıymetler için teşekkül ettirilen pazarlarda arz ve talebin karşılaştırılması, ancak Borsa Yönetimi tarafından belirlenen yöntemle tesbit edilen fiyat aralıkları içinde yapılır.

Her hisse senedi için, seans içinde işlem görebileceği en düşük (taban) ve en yüksek (tavan) fiyatlar, o hisse senedi için "Fiyat Aralığı"nı oluşturur. Bu sınırlar, o seansın baz fiyatı temel alınarak hesaplanır (İMKB 2006).

5.3.6 Hisse senetleri piyasası'nın işleyişi

A) Müşteri ve borsa emirleri

Aa) Müşteri emirleri

Borsa üyelerine müşterileri tarafından verilen alım satım emirleri "Müşteri Emri" olarak adlandırılır. Müşteri emirleri limit veya serbest fiyatlı olabilir. Limitli müşteri emrinde, emri veren alıcı, işlemin gerçekleşmesi için kabul ettiği en yüksek fiyatı; satıcı ise satmaya razı olduğu en düşük fiyatı belli eder. Müşteri, verdiği emirdeki fiyatı belirleme konusunda üyeyi serbest bırakmak istediğinde, "Serbest Fiyatlı" ibaresini taşıyan emir verir.

Ab) Borsa emirleri

Müşteri emirleri, üye temsilcileri tarafından seans içinde Borsa bilgisayar sistemine aktarıldığında "Borsa Emri"ne dönüşür.

B) Alım satım ile ilgili genel kurallar

Verilen bir emrin karşı bir emir tarafından karşılanması sırasında, sistemin uyguladığı bir takım işlem kuralları vardır. Hisse senedi almak veya satmak isteyen üye temsilcisi, bir emrin içermesi gereken tüm bilgileri bilgisayar aracılığıyla sisteme aktarır. İşlemler hisse senetleri piyasası için belirlenen öncelik kurallarına uygun şekilde sistem tarafından otomatik olarak gerçekleştirilir. Bu kurallar, fiyat ve zaman önceliği kurallarıdır.

Ba) Fiyat önceliği

Daha düşük fiyatlı satış emirlerinin, yüksek fiyatlı satış emirlerinden; daha yüksek fiyatlı alım emirlerinin, düşük fiyatlı alım emirlerinden önce işlem görmesini sağlayan öncelik kuralıdır.

Bb) Zaman önceliği

Girilen emirlerde fiyat eşitliği olması halinde, sisteme zaman açısından daha önce kaydedilen emrin daha önce işlem görmesini sağlayan öncelik kuralıdır.

Bc) Müşteri emirlerinin önceliği

Fiyat ve zaman öncelikleri açısından eşitliğin söz konusu olduğu emirler arasında müşteri emirleri, Borsa üyelerinin kendi nam ve hesaplarına verdikleri Borsa emirlerinden önce sisteme girilir (İMKB 2006).

6. BULANIK VERİ MADENCİLİĞİ ve SERMAYE PİYASALARINA UYGULANMASI: HİSSE SENETLERİ PİYASASI İÇİN BİR ERKEN UYARI SİSTEMİ ÖNERİSİ

Hisse senetleri piyasasında işlem manipülasyonu tespitine yönelik erken uyarı sisteminin tasarlanmasında kullanılan tanım, kavram, yöntem ve kullanılan değişkenler ile sistemin işleyişine yönelik açıklamalar aşağıda sunulmaktadır.

6.1 Manipülasyon

Manipülasyon kavramının aşağıda tüm detaylarıyla incelenmesinin yanısıra, birbirleri ile çok sık karıştırılan iki kavram olan spekülasyon ve manipülasyon arasındaki farkların ortaya konabilmesi için spekülasyon kavramının da açıklanmasında yarar görülmektedir. Bu nedenle, manipülasyon kavramının tüm yönleriyle ele alınmasının yanı sıra, spekülasyon kavramı da bu başlık altında incelenmektedir.

6.1.1 Spekülasyonun tanımı ve özellikleri

Spekülasyon, manipülasyon kavramıyla benzerliği nedeniyle çok sık olarak karıştırılan bir kavram olarak karşımıza çıkmasına karşın bir suç teşkil etmemesi nedeniyle, manipülasyondan ayrıştırılması çok büyük önem taşımaktadır.

Hisse senedi fiyatlarının belirlenmesinde, arz-talep unsurları ve finansal-ekonomik değişkenlerin etkisi yanında psikolojik davranışın etkisinin varlığı da bilinmektedir. Bu çerçevede; bireysel ve kurumsal yatırımcılar sürü psikolojisi ile hareket ederek spekülatif şişkinliklere (speculative bubbles) yol açabilirler. Hisse senedine olan talebin artması, arzın bu artışa ayak uydurmaması fiyatları daha da yükseltirken, kendi kendini besleyen bir süreç ortaya çıkacak, gerçekte hiçbir finansal, ekonomik ve politik gerekçesi olmayan bir artış trendi yaşanacaktır. Tasarruf sahipleri her geçen gün fiyatların daha da yükseldiğini görerek, daha fazla kazanmak için pozisyonlarını korurlar ve sürekli canlı bir talebin varlığı ile arz noksanlığı fiyat yükselişlerini

hızlandırır. Daha sonra fiyatların yeterli derecede arttığına inanan kurumsal tasarrufçuların hisse senedi portföylerini ellerinden çıkarmaları bir anda fiyatların düşmesine, yoğun satış emirlerinin borsaya kanalize olmasına neden olur. Fiyatların düştüğü bu dönem refah dönemini aksine bir çöküş (crash) dönemi olmakta, piyasayı yükseltip kar realizasyonunu gerçekleştiren kurumsal tasarrufçular yüksek karlar elde etmektedir. İşte, söz konusu kurumsal tasarrufçu spekülative faaliyette bulunan bir spekülâtördür. Şu halde spekülasyon, güçlü portföy yapısına sahip yatırımcıların sürü psikolojisinden yararlanarak sermaye kazancı (capital gain) ağırlıklı olmak üzere yüksek kazançlar elde etme amaçlı faaliyetleridir şeklinde tanımlanabilir (Özby 1990).

Özby (1990)'ın tanımından da anlaşıldığı üzere, spekülasyonda, yatırımcı psikolojisinden faydalanarak kazanç sağlanmaktadır. Spekülâtör, yatırımcının kazanç beklentisinden faydalanmakta ve onları istediği biçimde yönlendirebilmektedir.

Konuyu daha iyi anlamak bakımından Özby'nin (1990) belirttiği spekülative faaliyetlerin unsurlarını incelemekte fayda vardır:

1. Spekülasyon için güçlü bir portföy yapısına sahip olunmalıdır. Hisse senedi fiyatlarını belirli düzeylere getirebilmek için piyasa hacminin belirli bir yüzdesi spekülâtörün elinde olmalıdır. Aksi takdirde spekülative faaliyet amacına ulaşamayacaktır.

2. Spekülative faaliyet sürü psikolojisinden hareket eder. Tasarruf sahiplerinin toplu yönde hareket etmeleri ve kolay motive olmaları spekülative faaliyetlerin uygulanabilirliğini mümkün kılar. Aksi takdirde, fiyatların gereksiz arttığını gören tasarruf sahiplerinin spekülative amaçlı girişime destek vermemeleri bu girişimi başarısız kılacaktır.

3. Spekülasyonda amaç sermaye kazancı (capital gain) dır. Spekülative faaliyetler, özünde kısa süreli alım-satım işlemlerine dayalı olup, uzun vadede temettü (dividend) getirisi beklentisi içinde değildir. Özellikle günlük işlemciler (daily traders) aynı gün

içinde mümkün olduğunca fiyatı yükseltip o gün içinde kar realizasyonunu amaçlar. Öte yandan sürü psikolojisinin uzun vadeli reel pozitif etkisinden yararlanmayı amaçlayan uzun vadeli spekülâtörler (long speculators) 15-20 gün hatta bir aylık süre içinde fiyatları belirli seviyelere getirme amacını güderler. Ancak, her türlü spekülâtör sermaye kazancı amaçlı girişimde bulunur.

4. Yüksek kazanç sağlama temel amaçtır. Spekülâtif faaliyetleri mümkün kılacak finansal-ekonomik ve politik ortamları her zaman yakalamak mümkün değildir. Dolayısıyla spekülâtör bu imkanı bulduğunda azami kazancı sağlamak için vur-kaç yöntemini kullanmak isteyecektir.

Spekülasyonun unsurlarından da anlaşılacağı gibi sermaye piyasalarının işleyiş mekanizmasına zarar verecek bir faaliyet olmadığı gibi canlılık getirmesi nedeniyle zaman zaman faydalı olduğu bile düşünülmektedir.

6.1.2 Manipülasyonun tanımı ve özellikleri

En basit şekilde, piyasaların ve dolayısıyla fiyatların yapay olarak düşürülmesi, yükseltilmesi ya da aynı seviyede tutulması olarak tanımlanabilecek manipülasyon; piyasaların, genişlik ve derinlikten yoksun olduğu dönemlerde sık karşılaşılan bir olgudur. Diğer bir deyişle manipülasyon; bilerek ve isteyerek, menkul kıymet fiyatlarını kontrol etmek veya yapay şekilde etkilemek suretiyle yatırımcıları aldatmayı veya dolandırmayı amaçlayan davranışları ifade eder. Yatırımcıları aldatmayı amaçlayan manipülasyon faaliyeti, klasik dolandırıcılık eyleminin piyasalarla ilgili bir şeklidir. Bu özelliği nedeniyle piyasadaki işlemler ve bilgi akışı ile ilgili çok değişik versiyonda manipülasyonla karşılaşılabilir (Özbay 1990, Tezcanlı 1993, Ataman 1999).

Alım-satım işlemleri piyasaların varlık nedenleridir. Alım-satım işlemleri ile fiyatlar oluşmakta ve buradan yansıyan sinyallerle piyasadaki gerek aktif gerek potansiyel yatırımcılar yeni kararlar almakta veya mevcut kararlarını gözden geçirmektedirler. Bu sinyallerin ekonomik kaynak tahsisi açısından sağlıklı sonuçlar doğurması, ancak

dayandıkları arz ve talebin, serbest piyasa koşulları altında bilgisel anlamda etkin bir şekilde karşılaşması halinde mümkündür. Manipülasyon ile sermaye piyasası araçlarının fiyatının veya piyasada önemsenen diğer önemli ölçütlerin etkilenmesi amaçlandığından, piyasaların bilgisel anlamda ve kaynak tahsisi açısından etkin sonuçlar doğurması ilkesi zedelenmekte, yerine göre bu ilke boşa çıkarılmaktadır. Serbest arz ve talebin birbirinden bağımsız kombine yargısını yansıtması beklenen piyasa, önceden yazılmış senaryoya göre oynanan bir sahne performansına çevrilmektedir (Özbay 1990).

Bu bağlamda meşru alım satım işlemleri ile manipülasyon amaçlı alım satımların ayırddilmesi piyasalar açısından hayati önem arz etmektedir. Bu sınır iyi çizilemediği zaman piyasanın masum katılımcılarının kriminal kovuşturmalara konu edilerek piyasanın likiditesinin ve adalet duygusunun zayıflatılmasına neden olunurken, aksi durumda piyasa fiyatlarının güvenilirliğinin kalmaması ve böylece ilginin ve likiditenin azalması sözkonusu olabilmektedir (Özbay 1990).

Manipülasyonun tanımından piyasalara verdiği zarar açıktır. Ancak, yanlış tespitin verdiği zarar da azımsanmayacak ölçüdedir. Dolayısıyla, manipülasyon tespitine yönelik bir mekanizmanın veya sistemin çok hassas bir dengeye sahip olması gerekmektedir. Sistem, manipülasyonu gözden kaçırmayacak kadar duyarlı ama manipülasyonu diğer işlemlerden ayırt edebilecek kadar da kontrollü olmalıdır. Bu nedenle, manipülasyonun olabildiğince iyi anlaşılması gerekmektedir ki, tasarlanan sistem doğru bir tanımdan hareket edebilsin ve hassas dengeyi koruyabilsin.

Manipülasyon sonuçları bakımından piyasaya zarar vermektedir ve şüphesiz ki bir çıkar için yapılmaktadır ama manipülatif faaliyetlerin amacının iyi tanımlanması gerekmektedir. Özbay (1990) ve Tezcanlı'ya (1993) göre manipülatif faaliyetlerin başlıca üç amaçla yapıldığı görülmektedir:

1. Aktif piyasanın varolduđu imajını vererek diđer yatırımcıları piyasaya çekip, fiyatları yükseltmek ve elde bulunan hisse senetlerini daha yüksek fiyattan elden çıkarmak amaçlardan birisidir.

2. Fiyatları aynı düzeyde tutmak diđer bir yöntemdir. Deđeri düşük kalmıř hisse senetlerini olması gereken fiyat düzeyine yaklařtırmak amacıyla da manipülatif faaliyetlerde bulunulabilmektedir. Bu uygulamaya genellikle halka arza aracılıhta başvurulmaktadır. ABD’de kamunun aydınlatılması kořuluyla bu tür faaliyetlere izin verilmektedir.

3. Fiyatların ařađıya çekilerek, açađa satılan hisse senetlerinin daha düşük fiyatlardan satın alınması amacıyla da manipülasyon yapılmaktadır.

Manipülatif faaliyetlerin amaçları incelendiđinde,

1. Fiyatları yapay olarak yükseltmek,
 2. Fiyatları sabit tutmak,
 3. Fiyatları yapay olarak düşürmek,
- olmak üzere üç yol izlendiđi görülmektedir.

6.1.3 Manipülasyon çeřitleri

Manipülasyon kavram olarak aynı cümlelerle ifade edilebilse de farklı manipülasyon türlerinden bahsetmek mümkündür. Tezcanlı’ya (1993) göre temel anlamda, harekete dayalı manipülasyon, bilgiye dayalı manipülasyon ve işleme dayalı manipülasyon olmak üzere üç tip manipülasyon vardır:

i. Harekete Dayalı Manipülasyon (Action Based Manipulation): Hisse senetlerinin cari fiyatlarını ve dolayısıyla firma deđerini deđiřtirebilen işlemler olarak tanımlanır.

ii. Bilgiye Dayalı Manipülasyon (Information Based Manipulation): Kamuoyuna bilgi veren kaynakların fiyatları etkileyebilecek yalan, yanlış yanıltıcı ve mesnetsiz bilgi vermesi ya da kamuya bilgilendirmekle yükümlü olanların bu bilgiyi açıklamaması veya eksik açıklama yoluyla yapılır. Burada suç işleyen kişinin bu fiili kendi çıkarı doğrultusunda kullanması şart değildir. Yükümlünün bu kusurundan dolayı bazı yatırımcıların zarara uğraması yeterlidir.

iii. İşleme Dayalı Manipülasyon (Operation Based Manipulation): Hisse senetleri üzerinde işlem yapılan firmanın değerini değiştirmeye çalışmadan ya da piyasaya yanlış bilgiler yaymadan, sadece alım satımlar piyasayı manipüle etmeye gayret edenlerin kullandıkları bir stratejidir.

Görüldüğü gibi temel anlamda üç farklı tip manipülasyon vardır. Ancak, bu manipülasyon tiplerinden sadece birisi olan ‘İşleme Dayalı Manipülasyon’ bu tezin konusudur. İlerleyen bölümlerde sistem tasarımında ve uygulamasında manipülasyon ifadesinden sadece manipülasyonun bu tipinin anlaşılması gerekmektedir.

6.1.4 İşlem bazlı manipülasyonda yaygın olarak kullanılan manipülasyon yöntemleri ve davranış kalıpları

İşlem bazlı manipülasyonda kullanılan yöntemlerin her somut olaya göre değerlendirilmesi ve bu çerçevede suç oluşup oluşmadığının tespit edilmesi gerekmektedir. Bununla beraber, menkul kıymet borsalarında yaygın olarak kullanılan bir takım manipülasyon yöntemleri bulunmaktadır. Söz konusu yöntemlerde yapılan işlemler genellikle tek başına yasal bir görüntüye sahip olmakla birlikte, belirli bir amaç için koordineli olarak gerçekleştirildiklerinde manipülasyona neden olmaktadır (SPK 2006).

SPK’nın (2006) tanımından da anlaşıldığı gibi manipülasyon organize ve koordineli çalışılan bir suç olarak karşımıza çıkmaktadır. Her manipülasyon olayı kendine özgü bir niteliğe sahip olmasına karşın sıklıkla tekrarlanan bazı kalıplar da söz konusudur.

I. Manipülasyon yöntemleri

Tezcanlı (1993) ve SPK'nın (2006) manipülasyonla ilgili belirttiği gibi '**Her somut olayda farklı yöntemlerin kullanılması mümkündür. Dolayısıyla, manipülasyon kriterlerini sınırlı biçimde saymak mümkün değildir.**' Zaman içerisinde tespit edilebilen ve tedbirler alınan yöntemlerin yerine yenileri geliştirilmekte ve manipülasyonu belli kalıplar içerisinde tanımlamak mümkün olmamaktadır. Buna karşın, konunun daha iyi anlaşılması bakımından Tezcanlı (1993) ve SPK'da (2006) tarafından aktarılan, SPK denetimlerinde tespit edilen ve uluslararası uygulamalarda da kabul edilen manipülasyon yöntem ve kriterleri ile davranış kalıplarına aşağıda yer verilmektedir.

A- Mülkiyette Değişiklik Yaratmayan İşlemler

a) Kurgulu Emirler (Matched Orders): Kurgulu emirlerde manipülasyon grubunu oluşturanların, aynı miktar ve fiyattan aynı zamanda alım ve satım emirlerini vermesi ve söz konusu emirlerin eşleşerek işlemin gerçekleşmesi söz konusudur. Bu emirlerin amacı, hisse senedinde aktif bir piyasa izlenimi yaratmak suretiyle diğer yatırımcıları söz konusu hisse senedini almaya teşvik etmektir. Böylece, hisse senedinin piyasasındaki aktifliği görendiğer yatırımcılar piyasaya alıcı olarak girerek fiyatın yükselmesini sağlamaktadırlar. Manipülatör ise elindeki hisse senetlerini satarak haksız bir kazanç elde etmektedir. Öte yandan, söz konusu emirler ile fiyat artış veya azalışı da amaçlanabilmektedir.

b) Kendinden Kendine İşlemler (Wash Sales): Kendinden kendine işlemlerde hisse senetleri, alıcının ve satıcının aynı kişi olduğu, dolayısıyla mülkiyette bir değişikliğin gerçekleşmediği bir işleme konu olmaktadır. Manipülatör oldukça sık bir şekilde işlem gerçekleştirerek, söz konusu işlemlerle hisse senedinin piyasasında işlemlerin arttığı izlenimini yaratmak veya fiyatı etkilemek suretiyle diğer yatırımcıları yanıltmaya çalışmaktadır.

B- Fiyat Adımlarını Yükselterek Seri Halde Yapılan Tek Lotluk İşlemler: Maliyeti düşük olan ve fakat fiyatı etkileme imkanı bulunan bu tür 1 lotluk işlemlerin gerçekleştirilmesindeki amaç, genellikle fiyatın son gerçekleşen seviyesinin üstünde veya altında oluşmasını, kapanış fiyatının aynı şekilde yüksek görünmesini sağlamaktır. Bu işlemlerin genellikle hisse senedinin teminat değerinin daha yüksek görünmesini sağlamak ve böylece daha fazla kredi kullanmak veya ilave özkaynak koymaktan kaçınmak için yapıldığı görülmektedir.

C- Menkul Kıymette Yoğunlaşma (Runs): Söz konusu uygulama bir yatırımcının hisse senedinde yoğun alım (veya satım) yapmak suretiyle hisse senedinin piyasasında bir canlılık sağlamaya çalışmasını kapsamaktadır. Yoğun alımlardaki amaç diğer yatırımcıları etkileyerek onları alıma yöneltmek suretiyle hisse senedinin fiyatını yükseltmek, bir noktadan sonra da hisse senetlerini yüksek fiyattan elden çıkartarak menfaat sağlamaktır.

D- Arzı Kısıtlama (Corners): Bu tür uygulamalarda manipülatör, önemli miktardaki menkul kıymeti, diğer piyasa katılımcılarının daha sonra elindeki stoku daha yüksek fiyattan almak zorunda olduklarını bilerek almaktadır.

E- Menkul Kıymetin Piyasasına İstikrar Kazandırma (Market Stabilisation): Bu uygulama, yeni bir ihracın yapıldığı menkul kıymetin fiyatının hızlı düşmesini önlemek amacıyla menkul kıymette işlem yapılmasını kapsamaktadır. Söz konusu uygulamada yetkili aracı kuruluşlar ihracın başarısız olmaması için işlem yapmaktadır. Fiyat istikrar kuralları, sadece düşüşü önleme yönünde kullanıldığı, fiyatı arttırıcı rolü olmadığı ve gerekli kamuyu aydınlatma yükümlülükleri yerine getirildiği takdirde piyasa için yararlı bir mekanizmadır. Ancak, düzenlenmediği ve kuralsız yapıldığı takdirde manipülatif etkisi olabilmektedir.

F- Seans Açılışını Belirleme: Seans başlarında verilen emirler ve gerçekleşen işlemler, piyasa katılımcılarının seans aralarında geçen zamanda meydana gelen olaylara ilişkin yargılarını yansıtmaları itibariyle önemlidir. Dolayısıyla, seans başlarındaki işlem

hacmini ve/veya fiyatları etkileyerek seansın gidişini belirlemek bir manipülasyon stratejisi olarak kullanılabilir.

G- Seans Kapanışını Belirleme (Marking Close): Seans kapanışındaki işlemler günün özetini sunmaktadır. Bu nedenle, son durum itibarıyla hisse senedinin hangi fiyat seviyesinde olduğunu görmek isteyen katılımcılar kapanış fiyatına bakacaklardır. Medya ve borsa bültenleri de seansın seyrini ayrıntılı olarak sunmadığı için, piyasayı seans boyunca takip etmemiş olanlar, piyasanın yönünü seansın kapanışındaki fiyat seviyesinde izlemektedir. Yine yatırım fonları ve kredili menkul kıymet hesaplarının özkaynaklarının değerlendirilmesinde genellikle kapanış fiyatı bir değerlendirme ölçüsü olarak dikkate alınmaktadır.

Bu nedenlerle seansın sonuna doğru arz, talep ve/veya fiyatı etkileyerek değerlendirme sonuçlarını, sonraki seans için piyasayla ilgili diğer kişilerin beklentilerini ve alım satım kararlarını etkilemek, seans açılış fiyatını belirlemede olduğu gibi manipülasyon için elverişli bir yol olabilmektedir. Genellikle seans sonunda yapılan küçük miktarda işlemlerle kapanış fiyatı istenen düzeye getirilmektedir.

H- Fiktif Hesaplar Kullanma (Parking and Warehousing): Söz konusu uygulamada asıl olarak kimliği açıklanmamış bir kişi tarafından, başkası adına açılmış hesaplar üzerinden işlemler gerçekleştirilmektedir. Farklı kişilere ait fiktif hesaplar, kurgulu emirler ve kendinden kendine işlemlerin diğer piyasa katılımcıları tarafından fark edilmesini önlemek amacıyla kullanılmaktadır.

I- Açığa Satış: Açığa satış işlemi de bazen manipülatif amaçlı olabilmektedir. Açığa satış, satışın yapıldığı anda yatırımcının menkul kıymete sahip olmaksızın satış yapması olarak tanımlanabilir. Açığa satış işlemlerinin anlamlı olabilmesi için, ilgili menkul kıymetin fiyatının düşeceği beklentisinin olması gerekmektedir. Bu nedenle de, açığa satış işlemlerini manipülatif amaçlı kullanmak isteyenler, fiyat düşürücü emirleri hedefleyerek işlem gerçekleştirmektedirler.

II. Manipülatif davranış kalıpları

Manipülasyon yöntemlerinde olduğu gibi, manipülatif davranışlarda da sınırlı sayıda kalıptan bahsedememek doğaldır. Fakat, manipülasyon yöntemlerinde olduğu gibi manipülatif davranışlarda da sıklıkla rastlanan bazı kalıplar sözkonusudur. Manipülatif davranışlarının kalıplarla sınırlı olmadığı gibi, kalıplarla örtüşen davranışların da her zaman manipülasyona işaret etmediğine de dikkat edilmesi gerekmektedir. Yatırımcıyı bilgilendirmek ve manipülatif davranış ile manipülatif olmayan davranış arasındaki farkın ayrımını sağlayabilmek üzere SPK (2006) tarafından aktarılan bir dizi manipülatif davranış kalıbı aşağıda verilmektedir:

a) Verilen emirlerin ve gerçekleştirilen işlemlerin menkul kıymetin günlük işlem hacmi içinde önemli bir oran teşkil etmesi ve özellikle bu işlemlerin menkul kıymetin fiyatında önemli değişimlere neden olması,

b) Menkul kıymette önemli oranda pozisyona sahip kişiler tarafından verilen emirlerin ve gerçekleştirilen işlemlerin menkul kıymet fiyatında önemli bir değişikliğe yol açması,

c) Menkul kıymette yükselişten menfaat elde edecek olanlarla işlem gerçekleştirenlerin birlikte hareket etmeleri,

d) Seans içinde sürekli olarak aynı anda veya yakın zaman aralıklarıyla alım ve satım emirleri verilmesi ve kendi vermiş olduğu emirlerin eşleşmesi ile hisse senedinin mülkiyetinde herhangi bir değişikliğe yol açmayan işlemler gerçekleştirilmesi, bu emirlerin ve işlemlerin kişinin toplam işlemleri içinde ve toplam işlem hacmi içinde önemli bir yer tutması,

e) Seans içinde kısa zaman aralıklarıyla alıcı pozisyonundan satıcı pozisyonuna veya satıcı pozisyonundan alıcı pozisyonuna dönülerek gerçekleştirilen işlemlerin yatırım

mantığıyla yapılmaması, bu işlemlerin hisse senedinin günlük işlem miktarının önemli bir oranını teşkil etmesi ve bu işlemlerin menkul kıymetin fiyatında önemli değişikliklere yol açması,

f) Verilen emirlerin ve gerçekleştirilen işlemlerin seansın çok kısa bir bölümüne denk gelmesi ve hisse senedinin fiyatında ters yönde bir değişime neden olması,

g) Hisse senedinin sırasında bekleyen en iyi alış ve en iyi satış fiyatını değiştirmeye yönelik emirler verilmesi veya hisse senedinin alış ve satış emirlerinin kompozisyonunu değiştirmeye yönelik alış ve satış emirleri verilmesi ve bu emirlerin gerçekleşmeden önce iptal edilmesi,

h) İşlemlerin yoğun olarak seans başında veya seans sonunda gerçekleştirilmesi,

i) Çok sayıda aracı kurum üzerinden kendi adına veya vekaleten yönetilen veya kontrol edilen hesaplar vasıtasıyla emirler verilmesi ve işlem gerçekleştirilmesi,

j) Değişik aracı kurumlardaki farklı hesaplar üzerinden emirler verilmesi ve işlem gerçekleştirilmesi, işlemlerin gerçekleştirildiği dönemde menkul kıymetin fiyatında önemli değişikliklerin meydana gelmesi ve hesap sahiplerinin hesaplarında gerçekleştirilen işlemleri açıklayabilecek mali güçlerinin ve yapılan işlemleri izah edebilecek menkul kıymet işlemleriyle ilgili bilgi ve tecrübelerinin bulunmaması,

k) Birlikte hareket eden kişiler tarafından farklı aracı kurumlar ve farklı hesaplar üzerinden seans içinde birbirleriyle eşleşmesini sağlayacak şekilde devamlı olarak üst veya alt fiyat basamaklarından karşılıklı alım ve satım emirleri verilmesi ve karşılıklı olarak verilen emirler sonucunda işlemler gerçekleştirilmesi, bu emir ve işlemlerin hisse senedinde gerçekleşen toplam işlemlerin önemli bir oranını oluşturması,

1) Satın alınan hisse senedi bedellerinin, t+2 gününde aynı hisse senedinden yapılan satışları teminat olarak kabul eden başka bir aracı kurum tarafından satış tutarında kullanılan kredi ile ödenmesi, yapılan satışlara ilişkin takas yükümlülüğünün de t+2 gününde satın alınan hisse senetleriyle karşılanması, söz konusu işlem ve kredi sisteminin işlem gerçekleştirilen günlerde devamlı olarak uygulanması, olarak sıralanabilir.

6.1.5 Manipülasyon suçunu önlemeye yönelik olarak alınabilecek tedbirler

Manipülasyonun tamamıyla önlenmesi mümkün olmasa da, alınacak bazı tedbirlerle azaltılması mümkündür. Güngör (2001) tarafından önerilen tedbirler,

- Manipülasyon olaylarının önüne geçilebilmesi için şirketlerin halka açıklık oranlarının yükseltilmesi ve böylelikle piyasadaki hisselerin derinliğinin artırılması ilk akla gelen çözüm olarak ortaya çıkmaktadır,
- Bir diğer konu ise yatırımcıların eğitim düzeyinin yükseltilmesidir. Böylelikle yatırımcılar bilinçlenecek, manipülatörlerin hareket alanı daralacaktır. Ayrıca, piyasada kurumsal yatırımcıların ağırlığının artırılması da yatırım kararlarını profesyonelce veren kesimin borsadaki ağırlığını arttıracak ve piyasayı daha sağlıklı bir yapıya kavuşturacaktır. Ancak sayılan konuların kısa vadede gerçekleştirilmesi oldukça zor görünmektedir,
- Öte yandan manipülasyonla ilgili yasal düzenleme konusunda herhangi bir eksiklik bulunmadığı ve bu çerçevede anlık müdahaleler, kamuyu aydınlatma uygulamalarının daha sıkı takibi, Kurul kaydından çıkarma ve işlem yasağı uygulamalarının etkili yöntemler olarak ortaya çıktığı düşünülmektedir,
- Bunun ötesinde, elektronik bir gözetim sisteminin oluşturulması da yapılacak tespitleri kolaylaştırabilecek ve süreci hızlandırabilecektir,

olarak sunulmaktadır.

Yukarıda sayılan tedbirler dikkate alındığında, yasal düzenlemelerin yeterli yetkiyi tanıdığı görülmektedir. Ancak, tedbirler arasında sayılan halka açıklık oranının artırılması, piyasanın derinlik kazanması, yatırımcıların konuyla ilgili eğitilip bilinçlendirilmesi ve kurumsal yatırımcıların katılımının yükseltilmesi inisiyatif kullanılarak gerçekleştirilecek olgular değildir. Bu nedenle, yasal düzenlemelerle tanınan yetkinin daha etkin kullanımı için elektronik bir gözetim sisteminin oluşturulması daha anlamlı ve yapılabilir. Sözkonusu gözetim sisteminin içerik olarak bir erken uyarı sistemi yapısına sahip olması veya erken uyarı sistemi işlevini de yerine getirmesi caydırıcılığı artıracığından manipülasyonu tamamıyla ortadan kaldırırsa da teşebbüslerin sayısını azaltacağı kesindir.

6.2 Erken Uyarı Sistemleri

Manipülasyonun azaltılmasına yönelik alınacak tedbirler arasında en uygulanabilir olanının bir erken uyarı sistemi olduğu Bölüm 6.1’de vurgulanmıştır. Erken uyarı sisteminin yerine getireceği işlevin belirginleşmesi için manipülasyon tespitinde izlenen yolun tarif edilmesi gerekmektedir.

Manipülasyon, herhangi bir menkul kıymetin arz ve talebini etkilemeye ve böylelikle bir menfaat elde etmeye yönelik bir faaliyet olduğu için, manipülasyon incelemeleri esas olarak fiyat ve miktar hareketlerinin incelenmesine dayanmaktadır. Menkul kıymetlerin fiyat ve miktar hareketlerinde ortaya çıkan ve herhangi bir rasyonel beklenti ya da bilgiye dayanmayan olağan dışı değişiklikler, büyük ölçüde bir manipülasyon olayı ile karşı karşıya olunduğu sonucunu doğurduğundan, bu fiyat ve miktar hareketlerinin ayrıştırılması ve daha sonra bu işlemleri gerçekleştiren yatırımcıların, aracı kurumların ve inceleme konusunu oluşturan menkul kıymetlerin ihraççısı şirket ile şirket çalışanları arasındaki ilişkilerin analizi gerekmektedir (Ataman 1999).

Ataman'ın (1999) belirttiği gibi manipülasyon incelemeleri özünde fiyat ve miktar hareketlerinin incelenmesine dayanmaktadır. Olağandışı fiyat ve miktar hareketleri manipülasyona işaret etmekte ve sözkonusu işlemde paydaş olanların bir hiyerarşi içerisinde incelenmesi gerekmektedir. Görüldüğü gibi, manipülasyon tespiti aslında hesaplamaya dayalı bir dizi işlem gerektiren analitik bir süreçtir ve hisse senedi piyasasının doğal akışı içerisinde izlediği yoldaki paydaşların arasındaki ilişkiyi ortaya koymaktır. Bu tez kapsamında tasarlanan erken uyarı sisteminin de temel mantığı, insangücüyle yapılan hesaplamalar zincirini bilgi teknolojilerinden yararlanarak gerçekleştirmek ve şüpheli bir durumda olası ilişkileri ortaya koyacak analizleri yapmaktır.

6.3 Dünyada ve Türkiye'de Erken Uyarı Sistemi Örnekleri

Konuya ilişkin olarak, New York Borsası (NYSE 2005) tarafından uygulanmakta olan sistem iyi bir örnek teşkil ettiğinden kısaca değinilmesi faydalı olacaktır. New York Borsası'nda 'Stock Watch' olarak adlandırılan ve manipülasyon ve içeriden öğrenenlerin ticaretini önceden tespit edebilmeye yönelik bilgisayarlı bir erken uyarı sistemi kullanılmaktadır. Stock Watch NYSE (2005) tarafından 'NYSE'de kayıtlı menkul kıymetler için yasal olmayabilecek anormal fiyat ve miktar hareketlerini gözlemleyen, NYSE'nin son teknoloji bilgisayar destekli gözetim birimidir.' biçiminde tanımlanmaktadır.

NYSE'de kullanılan bir başka bilgisayarlı gözetim sistemi de ASAM (Automated Search and Match) olarak adlandırılan bir veri tabanı uygulamasıdır. ASAM veri tabanı yaklaşık 800.000 üst düzey yöneticinin, avukatın, bankacının ve muhasebecinin, yaklaşık 80.000 şirket yöneticisinin ve 30.000 iştirak veya bağlı kuruluş çalışanlarının bilgilerinin mevcut olduğu bir veri tabanıdır. ASAM sayesinde, işlem gerçekleştirenler, ihraççılar ve yöneticileri, aracı kurumlar ve yöneticileri ve ilgili olabilecek diğer piyasa katılımcılarının listesi çıkarılmakta ve bilgisayar tarafından birbirleriyle ilişkileri bulunabilecek kişi ve kurumlar otomatik olarak eşleştirilmektedir. Yine anılan sistem, bu eşleştirme sırasında ilgili kişi ve kurumların ikamet ettikleri ve çalışmakta oldukları

bölgeleri de değerlendirerek, geçmişte yaşanan olaylarla da ilişki kurmak suretiyle bu işlemi gerçekleştirmektedir. Bu işlem sonucunda şüpheli görünen işlemler bulunduğu tespit edildiğinde borsa üyelerine yaptırım uygulanmakta, kendi yetkisini aşan aykırılıkların tespiti halinde ise SEC'ye iletmektedir (Ataman 1999).

Erken uyarı sistemlerine bir başka örnek de Tayland Borsası (SET) tarafından kullanılan ATOMS (Automated Tools for Market Surveillance) sistemidir. ATOMS sistemi içerisinde yer alan Otomatik Uyarı Sistemi, gerçek zamanlı olarak fiyat ve/veya miktar hareketlerini, geçmiş değerlerle kıyaslayarak; olağandışı durumlarda sesli uyarı vermektedir. Uyarı neticesinde, olağandışı örüntü, soruşturulmaktadır.

ATOMS sisteminin,

1. Olağandışı fiyat değişimleri için uyarı veren 'Fiyat Uyarı Modülü'
2. Olağandışı miktar değişimleri için uyarı veren 'Miktar Uyarı Modülü'
3. Fiyat ve miktar kombinasyonunda olağandışı değişimler için uyarı veren 'Fiyat ve Miktar Uyarı Modülü' olmak üzere

3 modülü bulunmaktadır. Ayrıca, ATOMS bünyesinde ön soruşturma ve dokümantasyon imkanı sağlayan alt sistemler de mevcuttur (SET 2006).

Stock Watch ve ASAM gibi sistemler, manipülasyon gibi sermaye piyasası suçlarının tespitinde bilgi teknolojilerinin kullanımına iyi birer örnek teşkil etmektedir ama bunların dışında da finansal erken uyarı sistemi önerileri sözkonusudur. Tezcanlı (1993) sermaye piyasaları için gözetim sistemlerinin önemine dikkat çekerek, istatistik temelli, bilgisayar destekli bir gözetim sistemi modeli önermiştir. Önerilen model, bir 'Gözetim Analisti'nin fiyatları takip ederek, resmi ve resmi-olmayan bilgi kaynaklarından haber akışını dikkate almasını ve olağandışı bir durumla karşılaştığında; hisse senedi alım satımı yapan kişiler, piyasa yapıcılar gibi aktörlerle telefonla görüşüp, manipülasyon veya içerden öğrenme durumu olduğu kanaati oluşması halinde gözetim birimi yetkililerine raporlanmasını öngörmektedir. Ayrıca, Borsa'da işlem gören hisse senetlerinin bilançolarını kullanarak Lojistik Regresyon yöntemine dayalı işlem

manipülasyonuna yönelik bir erken uyarı sistemi önerilmiştir. Önerilen sistemde manipülasyon tespitini sağlamak üzere bir Lojistik Regresyon Modeli geliştirilmiştir.

Genel anlamda sermaye piyasası suçları özelde de manipülasyon tespitine yönelik, Sermaye Piyasaları için geliştirilmiş gözetim ve erken uyarı sistemlerinin yanısıra Bankacılık için geliştirilmiş Erken Uyarı Sistemleri de sözkonusudur. Bankacılığa yönelik olarak geliştirilen erken uyarı sistemlerine Dünya ve Türkiye'den örnekler aşağıda verilmektedir.

Çilli ve Temel (1988) Diskriminant Analizi ve Faktör Analizi'ne dayalı bir erken uyarı sistemi önermektedirler. Çilli ve Temel (1988) tarafından önerilen sistem bankacılık krizlerini tespit etmeye yöneliktir.

TCMB, Bankacılık Genel Müdürlüğü bünyesinde yer alan Bankalar Veri Kontrol Müdürlüğü ve Bankalar Gözetim Müdürlüğü tarafından yürütülen, bir bankacılık gözetim sistemi de söz konusudur. Gözetim sistemi, Bankalardan elde edilen verilerin içsel tutarlılıkları kontrol edildikten sonra 75 adet rasyo ile özkaynak yeterliliği, aktiflerin niteliği, karlılık, likidite ve pasif yapısı incelenerek, uyarı raporları hazırlanmaktadır. Bunun yanısıra, performans raporları, bilanço ve kar-zarar tablolarında yer alan kalemlerin oransal-yapısal gelişmeleri, yabancı para pozisyonu, likidite, krediler portföyü, nakit akımları analizi gibi raporlar da TCMB Bankacılık Genel Müdürlüğü tarafından hazırlanmaktadır (Gönül ve Eroğlu 1999).

Gaytan and Johnson'un (2002) bankacılık krizleri için erken uyarı sistemleri literatür taramasında ise önerilen modellerin Probit ve Logit modellere dayandığı görülmektedir.

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

Erken Uyarı Sistemi tanımı altında yapılan çalışmalar incelendiğinde aynı tanım altında bazı gösterge setlerinin incelenmesinin de yer aldığı görülmektedir. Özellikle, Bankacılık ve Ekonomik Krizleri öngörme amacıyla takip edilmesi gereken ekonomik göstergeleri tespit etmeye yönelik bu çalışmalar bir sistemden çok gösterge setini ortaya koymakta ve bu çalışmaların ‘Erken Uyarı Göstergeleri’ ifadesiyle anılmasının daha doğru olduğu düşünülmektedir. Bu noktadan hareketle, Stock Watch, ASAM, ATOMS ve Tezcanlı’nın (1993) sistemi Erken Uyarı Sistemi olarak tanımlanabilirken; Çilli ve Temel (1988), Gönül ve Eroğlu (1999), Gaytan ve Johnson (2002) ile Gerni vd.’nin (2005) çalışmaları Erken Uyarı Göstergeleri olarak tanımlanmalıdır.

Bu tez kapsamında Erken Uyarı Sistemleri değerlendirildiği için, Stock Watch, ASAM, ATOMS ve Tezcanlı’nın (1993) sistemlerinin etüd edilmesi esastır. Stock Watch, ASAM ve ATOMS sistemleri, sermaye piyasası suçlarını tesbite esas sistemler olması ve halihazırda kullanılıyor olmaları nedeniyle gizliliğe sahip olmakta; sistem içeriği konusunda, teknik ve kullanılan yöntemlere yönelik detaylara ulaşamamıştır. Sistemlerle ilgili daha çok tanım ve işleve yönelik detaylara erişmek mümkün olmuştur ki, ulaşılan bilgilerde sistemlerin aktarımı sırasında verilmiştir.

Sözkonusu sistemlerden VM’den yararlanıldığı ifade edilen sadece ‘Stock Watch’ olmasına karşın kullanılan yöntemlere ulaşmak mümkün olmamıştır. ASAM’da ise tamamıyla bir veritabanı sorgusu sözkonusudur. ATOMS bünyesinde izlenen yaklaşım ise fiyat ve/veya miktar değişimlerini geçmiş değerlerle kıyaslayarak olağandışı değişimleri tespite yöneliktir. Tezcanlı (1993) ise ASAM’a benzer bir telefon destekli gözetim sistemi önerirken, sistemin erken uyarı bileşeni Lojistik Regresyon Yöntemi ile bir istatistiksel model kurarak, geçmiş manipülasyon örneklerine dayalı olarak oluşturduğu bir şablonla manipülasyon şüphesine sahip örnekleri kıyaslamaya dayanmaktadır.

Bu tez kapsamında geliştirilen Erken Uyarı Sistemi;

- Sadece Stock Watch’da bahsedilen VM’yi temel almanın ötesinde Bulanık Küme Teorisi’nden de istifade etmektedir,

- ATOMS’da fiyat ve/veya miktar deęişimleri gemiş ile kıyaslanırken, gemiş ile kıyaslama yerine işlem esnasındaki farklılaşmaları esas almaktadır,
- Tezcanlı’nın (1993) önerdiği sistem Lojistik Regresyon Modeli’ne dayalı bir şablon oluşturup, gemiş tecrübelerle dayalı olarak manipölasyon tespitini önerirken; veriden öęrenmeye dayalı olarak verideki anomalilerin tespitine dayanmaktadır,
- Her ülkenin HSP için işleyiş farklılığı dikkate alındığında, geliştirilen sistem Türkiye koşullarına özgüdür.

6.4 Bulanık Veri Madenciliğine Dayalı İşlem Manipölasyonu Tespitine Yönelik Erken Uyarı Sisteminin Tanımı ve İş Akışı

Bu tez kapsamında incelenen ve uygulamaya esas olan problem Hisse Senetleri Piyasası’nda (HSP) işlem manipölasyonu tespitidir. Uygulama olarak işlem manipölasyonu tespitini esas alan bir Erken Uyarı Sistemi tasarlanmaktadır.

Sistem tasarlanırken Bilişim ve İstatistik olmak üzere izlenebilecek iki temel yaklaşım sözkonusudur. Bu yaklaşımlardan istatistięi temel alan İstatistiksel Öęrenme esas alınmaktadır.

Sistem modellenirken, Bölüm 6.3’de aktarılan Erken Uyarı Sistemleri’den farklı olarak HSP’nin kaotik yapısına uygun biçimde işleyişi sağlayan Bulanık Küme Teorisi’nden faydalanılmaktadır. Her ülkenin HSP işleyişindeki farklılıklar dikkate alındığında, Türkiye’ye özgü bir Erken Uyarı Sistemi için modelleme yapmak gerekmektedir. Bu açıdan bakıldığında, Türkiye özelinde, HSP İşlem Manipölasyonu tespitine yönelik Veri Madenciliğine dayalı olarak gerçekleştirilen ilk Erken Uyarı Sistemi’dir.

HSP’de işlem manipölasyonu incelemesinde, HSP’de işlem gören şirketler, işlemlere aracılık eden HSP üyeleri (aracı kuruluşlar) ve işlemleri yapan yatırımcılardır. HSP

işlem verileri, İMKB endeksleri, işlem gören şirketlerin mali tabloları değerlendirilebilecek temel veri kaynakları olarak görülmektedir.

Erken uyarı sistemi modellenirken izlenebilecek yaklaşımlardan birisi, geçmişe dayalı örneklerden faydalanarak bir veya birden fazla İstatistiksel Öğrenme tekniğine dayalı istatistiksel bir tahmin modeli kurmak olabilirdi. Şablon görevi görecektir bir istatistiksel tahmin modeli kurmanın, doğası gereği belli bir hata payını içereceği ve geçmişten farklı bir davranışı algılamakta başarılı olamayacağı açıktır. Ayrıca, Tezcanlı (1993) ve SPK'ya (2006) göre her somut olayda farklı manipülasyon yöntemlerinin kullanılması mümkündür. Dolayısıyla, manipülasyon kriterlerini sınırlı biçimde saymak mümkün değildir. Bu nedenle, her manipülatif olayı içerebilecek bir model kurmak mümkün olmadığına göre şablon bir model kurmak yerine, manipülasyonun tanımından giderek verideki (fiyat ve miktar değişimindeki) anomalileri tespit edecek bir yaklaşım izlemek gerekmektedir. Bu noktadan hareketle, verideki işlemde kaynaklanan farklılıkları açığa çıkarmak en uygun yaklaşım olarak düşünülmüştür.

ATOMS sisteminde olduğu gibi hisse senedinin fiyat ve miktar değişimindeki anomalileri, geçmişteki değişim miktarıyla kıyaslayarak tespit etmenin de, geçmişteki değişim sınırları içerisinde manipülasyona imkan tanıdığı düşünülmektedir. Bu nedenle, manipülasyon arayışı incelenen işlem dönemi içerisindeki anomalileri dikkate almaktadır. Anomaliden kasıt, işlem dönemindeki fiyat ve miktar değişimi bakımından diğer hisse senetlerinden ayrışacak biçimde farklılık gösterenlerdir.

Manipülasyonu tespit etmenin, piyasaya vereceği zarar açısından önemi açıktır. Buna karşın, yanlış tespitinde yatırımcıyı piyasadan uzaklaştırarak diğer bir zarara yol açacağı da dikkate alınması gerekmektedir. Dolayısıyla manipülasyon tespit edilirken, manipülatif olmayan davranışın da ayrıştırılması önem taşımaktadır. Bu nedenle, manipülasyon tespitinde, tanımından gelen ve manipülasyona işaret eden tüm unsurların varlığının araştırılması esas olarak alınmalıdır. Erken uyarı sistemi tasarlanırken bu hassasiyet dikkate alınmaktadır.

Erken uyarı sistemi tasarlanırken, HSP işlem verileri esas alınmıştır. İMKB endeksleri, işlem gören şirketlerin mali tablolarına ait veriler dikkate alınmamıştır. Şüphesiz ki, mümkün olduğunca fazla değişkeni sisteme katarak daha kapsamlı bir sistem tasarlamak mümkündür. Ancak, Sistemin amacı ‘Erken Uyarı’ olduğu için mümkün olduğunca hızlı bir sistem tasarlamak esas alınmış ve sadece HSP işlem verileri üzerinden sistem geliştirilmiştir.

Hisse senetleri piyasası;

- Hisse senedi işlemlerinin gerçekleştiği İstanbul Menkul Kıymetler Borsası (İMKB),
 - Müşteriler adına hisse senedi işlemlerini gerçekleştiren üyeler (aracı kuruluşlar),
 - Emirleri veren müşteriler (yatırımcılar),
- olmak üzere üç temel bileşenden oluşmaktadır.

Erken uyarı sistemi tasarlanırken, manipülasyon tespitinde HSP’nin bütün bileşenleri dikkate alınmaktadır. HSP’nin işleyişine uygun olarak, işlem manipülasyonuna yönelik erken uyarı sisteminin tasarımı üç basamaklı bir süzgeç olarak gerçekleştirilmiştir:

1. basamakta, hisse senetleri etüd edilmektedir. İMKB HSP’de gerçekleşen işlemler arasında diğerlerinden ayrışan davranış örüntüsüne (pattern) sahip hisse senedi olup olmadığı incelenmektedir. İşlem davranışında örüntüler;

- Fiyat değişimindeki farklılıklar

veya

- İşlem miktarlarındaki farklılıklar,

olmak üzere iki başlık altında incelenmektedir. Hisse senedi işlem davranışları açısından diğerlerinden farklılık yakalandığında sarı alarm durumu gerçekleşecek ve

2. basamak işlem etüdüne geçilecektir. Eğer farklı örüntüde bir davranış söz konusu değilse alarm durumu söz konusu olmayacaktır.

2. basamakta, Hisse Senetleri (HS) üzerinde gerçekleşen işlemler arasında diğerlerinden ayrışan örüntüler tespit edildiğinde; gerçekleşen işlemlerin, belli aracı

kuruluşlar (üyeler) üzerinde yoğunlaşp yoğunlaşmadığı sorgulanacaktır. Eğer işlemlerin belli aracı kuruluşlar (üyeler) üzerinde yoğunlaştığı görülürse turuncu alarm durumuna geçilecek ve 3. basamak işlem etüdüne başlanacaktır. Aksi taktirde, alarm hali sona erecektir.

3. basamakta, HSP’de davranış örüntüleri açısından diğerlerinden ayrışan hisse senedine rastlanması ve söz konusu işlemlerin aynı aracı kuruluşlar tarafından işlem talimatlarının verilmesi halinde, aynı müşteri ya da müşteriler adına işlem yapılıp yapılmadığı sorgulanacaktır. İşlemlerin aynı müşteri ya da müşteriler adına gerçekleştiği belirlenirse kırmızı alarm durumuna geçilerek; ‘İşlem Manipülasyonu Gerçekleştiği’ uyarısı yapılacaktır.

HSP işlem verileri sistemin girdisi olmak üzere, sistemin çıktıları ardışık olarak manipülatif işleme maruz kalmış olması bakımından şüpheli hisse senedi/senetleri, manipülasyon şüphesi taşıyan işlemler, manipülatif işlemin gerçekleşmesine aracılık eden aracı kuruluş(lar) ve işlem emrini veren yatırımcı(lar)dır.

6.5 Bulanık Veri Madenciliğine Dayalı İşlem Manipülasyonu Tespitine Yönelik Erken Uyarı Sisteminde Kullanılan Değişkenler ve Tanımları

İşlem manipülasyonu tespitine yönelik erken uyarı sisteminin tasarımında, sistemin erken uyarı özelliğini yerine getirebilmesi için hızlı işlemesi gerekliliğinden dolayı, sadece HSP işlem verileri dikkate alınmıştır. Fiyatlar üzerinde işlem manipülasyonu yapıldığı kesinleşmiş bir örnek üzerinden inceleme gerçekleştirilecek olup, sistemin işleyişinde dikkate alınacak HSP işlem verilerine ait değişkenler aşağıda tanımlarıyla yer almaktadır.

Zaman (T): İşlemin gerçekleşme zamanını göstermektedir. Saat/Dakika/Saniye biçimindedir.

Menkul: İşlem gerçekleştirilen hisse senedinin kodunu göstermektedir.

Fiyat: İşlem gerçekleştirilen hisse senedinin fiyatını göstermektedir.

Alan üye (AU): Hisse senedini satın alan üyenin (aracı kuruluşun) kodunu göstermektedir.

Alan hesap no (AH): Hisse senedini satın alan yatırımcının hesap numarasını göstermektedir.

Satan üye (SU): Hisse senedini satan üyenin (aracı kuruluşun) kodunu göstermektedir.

Satan hesap no (SH): Hisse senedini satan yatırımcının hesap numarasını göstermektedir.

Tasarlanan Erken Uyarı Sistemi'nin basamakları ile değişkenlerin eşleştirmesi aşağıdaki biçimde gerçekleşmektedir:

1. Basamak: Hisse senedi işlem örüntüsü ayrışması,

- Zaman,
- Menkul,
- Fiyat,
- Alan üye,
- Satan üye,
- Alan hesap no,
- Satan hesap no,

değişkenleri etüd edilerek değerlendirilecektir.

2. Basamak: Aracı kuruluş (üye) davranışları benzeşmesi;

- Fiyat,
- Alan üye,
- Satan üye,
- Alan hesap no,
- Satan hesap no,

değişkenleri kullanılarak etüd edilecektir.

3. Basamak: Müşteri davranışları benzeşmesi;

- Alan hesap no,
- Satan hesap no,

değişkenleri kullanılarak sorgulanacaktır.

6.6 Bulanık Veri Madenciliğine Dayalı İşlem Manipülasyonu Tespitine Yönelik Erken Uyarı Sisteminin Modellenmesi

Tasarlanan erken uyarı sistemine yönelik modelleme, Bölüm 6.4'te verilen iş akışı doğrultusunda bulanık mantığa dayalı olarak model uydurmaya ihtiyaç duyulmayan İstatistiksel Öğrenme tekniklerinin kullanımıyla adımsal olarak hazırlanmıştır. Her bir adımda, eleme yapılarak sürece devam edilmektedir. İşlem gören bütün hisse senetlerini fiyat değişimi, aracı kuruluş ve yatırımcı ilişkileri açısından manipülasyon şüphesiyle sorgulamak hem iş yükü, hem de zaman maliyeti açısından anlamlı değildir. Dolayısıyla, öncelikle fiyat değişiminde diğer hisse senetlerinden ayrışanlar tespit edilmektedir. Ardından şüphe taşıyan hisse senetleri için aracı kuruluş ve yatırımcı ilişkileri etüd edilmektedir. Bu noktadan hareketle, erken uyarı sistemi üç basamaklı bir süzgeç olarak çalışmakta ve sırasıyla, hisse senedi, aracı kuruluş ve yatırımcılardan şüphe taşımayanları eleme esasına dayanmaktadır.

1. basamakta, İMKB Hisse Senetleri Piyasası'nda (HSP) gerçekleşen işlemler arasında diğerlerinden ayrışan davranış örüntüsü (pattern) gösterenler tespit edilmektedir. Hisse senetlerinin fiyat hareketlerinde olağandışı bir değişim olup olmadığına bakılmaktadır.

Özbay (1990)'a göre manipülasyon tanımına bakılırsa; 'En basit şekilde, piyasaların ve dolayısıyla fiyatların yapay olarak;

- Düşürülmesi,
- Yükseltilmesi,
- Aynı seviyede tutulması,

olarak tanımlanabilecek manipülasyon...' ifadesi yer almaktadır. Bu ifade dikkate alınrsa yapılan diğer işlemlerden ayrışacak biçimde;

- Düşük,
- Yüksek,
- Durağan (orta),

fiyat veya miktar değişimi manipülasyona işaret etmektedir. Bu noktadan hareketle, manipülasyon kriterlerini sınırlı biçimde saymanın da mümkün olmadığı dikkate alınarak; fiyat değişim oranının 3 parçaya ayrılması ve diğerlerinden ayrışacak biçimde farklılaşanların manipülasyon şüphesi ile incelenmesi hedeflenmiştir. Bu hedefe ulaşmak amacıyla, nesnelere aralarındaki benzerlik ölçütüne dayalı olarak gruplamayı esas alan denetimli istatistiksel öğrenme tekniklerinden,

- K-ortalamar Kümeleme Analizi K=3 alınarak,

3 grup elde etmek için uygulanmaktadır.

Gruplama yapılırken alternatif olabilecek İstatistiksel Öğrenme tekniklerinden,

- Denetimli bir yöntem olan K-en yakın komşuluk yöntemi, komşuluğu sorgulanabilecek bir hedef nokta belirlemek mümkün olmadığından,
- Denetimsiz bir yöntem olan Hiyerarşik Kümeleme Analizinde, çıktı bölgesini, bir başka deyişle küme sayısını belirlemek mümkün olmadığından ve karar verirken insan yargısı kullanımı gerektiğinden,
- Denetimsiz bir yöntem olan Kendi Kendini Düzenleyen Haritalar ise sınıflandırmasında zorluk çekilen küresel vb. yüzeye sahip bir veri söz konusu olmadığı için,

kullanılmamıştır. Bu yöntemlerin yerine amaca daha fazla hizmet edecek olan denetimli İstatistiksel Öğrenme tekniklerinden K-ortalamar Kümeleme Analizi kullanılmıştır. K-ortalamar Kümeleme Analizi'nde, hem bir model uydurmak gerekmemekte, hem de

küme sayısını en başından belirlemek mümkün olmaktadır. K-ortalamlar Kümeleme Analizinin K=3 için uygulanmasıyla, fiyat (ve/veya miktar) değişim oranlarına göre,

- Düşük,
- Yüksek,
- Durağan (orta),

değişime sahip gruplar (kümeler) elde edilmektedir.

Elde edilen kümeler incelenerek, diğerlerinden ayrışan küme ya da kümeler (ki muhtemelen diğerlerine göre daha az işleme sahip olan kümeler olacaktır), şüpheli işlem olarak incelemeye tabi tutulmaktadır. Şüpheli işlemlerin söz konusu olması halinde Sarı Alarm haline geçilecektir. Böylece 1. Basamak işlem etüdünün 1. etabı tamamlanarak; 2. etap olan değişimleri bulanıklaştırma aşaması gerçekleştirilecektir.

Elde edilen kümelerin, küme merkezlerinin birbirlerine yakın olması halinde diğerlerinden ayrışan örüntü olmadığı veya bir başka deyişle şüpheli işlem olmadığı kararına varılarak alarm verilmeden süreç sona erecektir.

İkinci etapta, 1. basamak işlem etüdünün birinci etabında şüpheli olarak tespit edilen işlemler buldukları kümelere aidiyet derecelerini tespit etmek üzere,

• FANNY Bulanık 2-ortalamlar Kümeleme algoritmasının uygulamasına tabi tutulacaktır. Bu algoritmanın 2 gruba ayırıştırmak için uygulanmasının nedeni,

- Düşük düzeyde fiyat değişimine sahip işlemler,
- Yüksek fiyat değişimine sahip işlemler,

olmak üzere şüpheli işlemlerin iki gruptan oluşmasıdır. Böylece, manipülasyon amaçlı alım ve satım işlemlerinin karşılıklı gerçekleşip gerçekleşmediği sorgulanmış olacaktır. Kesin K-Ortalamlar Kümeleme Analizi'nin yerine Bulanık C-Ortalamlar Kümeleme Analizi'nin uygulanmasının nedeni ise gözlemlerin iki kümeye ayrılmasına karşın, her bir gözlemin kümelere aidiyet derecesinin farklı olması ve bir anlamda bulunduğu kümenin özelliklerini aidiyet derecesi kadar temsil etmesidir. Ayrıca, her bir gözlem için iki kümeye de aidiyet dereceleri ayrı ayrı verilmektedir. Her bir gözlem, toplamları

1 olan farklı üyelik fonksiyonu değerleriyle veya aidiyet dereceleriyle her iki kümede de temsil edilmektedir.

Şüpheli işlemler toplamları 1 olan iki adet üyelik değeriyle iki kümeye de aidiyet göstermektedir. İşlemlerin, aidiyet derecesinin daha büyük olduğu kümeleri daha iyi temsil ettiği noktasından hareketle $\mu_{1,i}$ ve $\mu_{2,i}$ sırasıyla, i . gözlem için birinci ve ikinci kümeye aidiyet dereceleri olmak üzere,

$$w_i = \text{Max}(\mu_{1,i}, \mu_{2,i}) \quad i = 1, 2, \dots, n$$

her bir fiyat değişim oranının bulunduğu kümeyi en büyük temsil etme düzeyi elde edilmiştir. Dolayısıyla fiyat değişim oranları (fdo) kesin değerleriyle değil, üyelik dereceleriyle kümeye aidiyetlere düzeyinde ağırlıklandırılarak;

$$bfdo_i = w_i fdo_i \quad i = 1, 2, \dots, n$$

biçiminde bulanık fiyat değişim oranı haline getirilecektir.

FANNY algoritmasının uygulanması sonucu elde edilen Normalize edilmiş bulanıklık katsayısı olan,

- Dunn Katsayısı

sistemin bulanıklık düzeyini ortaya koyacaktır.

Bulanık fiyat değişim oranları her bir şüpheli işlem için hesaplandıktan sonra Sarı Alarm halinde 2. basamak işlem etüdüne geçilecektir.

2. basamakta, 1. basamak işlem etüdü sonrasında şüpheli olarak tespit edilen Hisse Senetleri (HS) için gerçekleşen işlemlerin; belli aracı kuruluşlar (üyeler) üzerinde

yoğunlaşp yoğunlaşmadığı sorgulanacaktır. İşlemlerin aynı üye ya da üyeler üzerinde yoğunlaşp yoğunlaşmadığının tespit edilmesi için,

- Bulanık girdilere sahip CHAID Karar Ağacı Algoritması

kullanılacaktır.

CHAID karar ağaçları algoritmasının yerine kural çıkarımında kullanılabilir alternatifler olan,

- CART, ikili ağaç türettiği için tüm olası alt grupları yansıtmadığından,
- Birliktelik kuralları, karar ağaçlarında olduğu gibi karar sürecini kolaylaştıracak görsellik sunmadığı ve elde edilen birliktelik kurallarının içinden seçim yapılacaklar için belirli eleme kriterleri gerektiği için

kullanılmamıştır.

CHAID karar ağacı algoritması ile şüpheli olarak tespit edilen her bir hisse senedinin bulanık fiyat değişim oranları hedef olmak üzere, alan üye, satan üye, alan hesap no ve satan hesap no değişkenlerine göre haritalanacaktır. Karar ağaçları algoritmaları ile olasılıksal olarak her bir hisse senedinin alan ve satan üyenin hangisi ya da hangilerinden ağırlıklı olarak geldiği belirlenecektir. İncelenen hisse senetleri için işlem yapan üyelerden bir veya birkaçında anlamlı bir yoğunlaşma görüldüğünde, ilgili hisse senedi 3. Basamak işlem etüdüne sevk edilecek ve ‘Turuncu Alarm’ durumuna geçilecektir. Herhangi bir yoğunlaşma söz konusu değilse alarm durumu sona erecektir.

Ayrıca, bu basamakta, alan ve satan üyelerden şüpheli olanların altında hangi müşterilerin de işlem yaptığı belirlenip 3. basamakta aralarındaki ilişkiler etüd edilecektir.

3. Basamakta, manipülasyon tespitinde son aşama olan müşteriler arası ilişki sorgulanacaktır. Müşteriler arası ilişkilerin sorgulanmasında;

- Önsel (Apriori) Birliktelik Kuralları Algoritması

yöntemi kullanılacaktır.

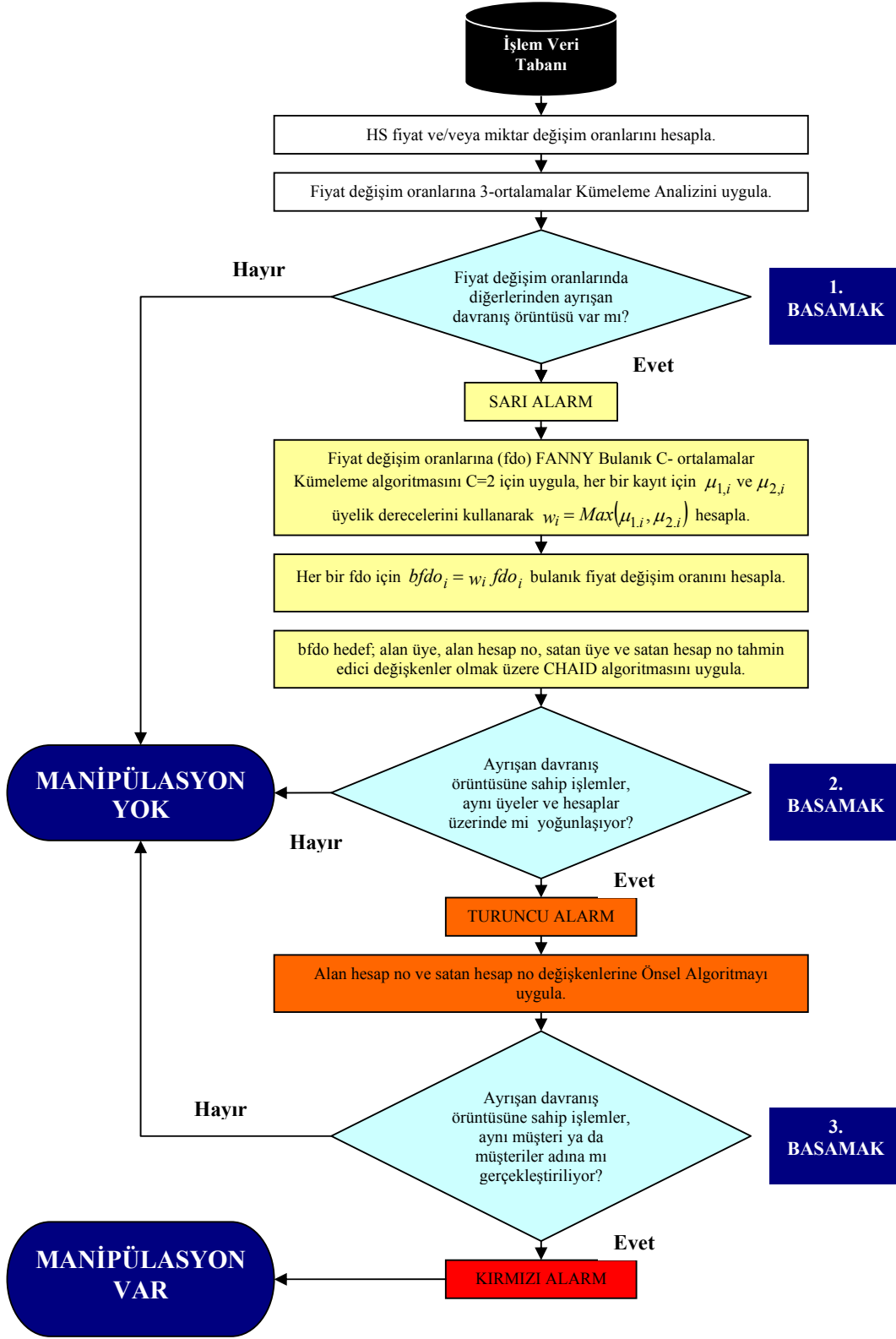
Birliktelik kuralları yerine kural çıkarımında kullanılabilecek alternatif olan,

- Karar ağaçları, birliktelik kurallarında olduğu gibi hızlı sonuç vermemekte ve süreç alternatifler arasından seçim yapmayı gerektirmediği için görsel destek gerekmemekte; sadece mevcut durum sorgulanıp, elde edilen birliktelik kurallarının tespiti gerekmekte olduğu için,

kullanılmamıştır.

Müşteri bilgileri arasında, şüpheli işlem esnasında (veya öncesinde) beraber davranış sergileyen bir grubun varlığı araştırılacaktır. Beraber davranış sergileyen bir kişi veya grubun varlığının tespiti halinde ‘Kırmızı Alarm’ durumuna geçilecek ve manipülasyon yapıldığı uyarısı verilecektir. Beraber davranış sergileyen yatırımcılara rastlanmaması halinde alarm hali sona erecektir. Bu basamağın başka bir özelliği de; manipülasyon ve spekülasyon ayrımını ortaya koyuyor olmasıdır. Belli aracı kuruluşlarda yoğunlaşma olmasına karşın karşılıklı alım-satım işlemleri bakımından yatırımcılar arası ilişki gözlenemiyorsa, manipülatif değil spekülatif bir davranışın varlığından söz edilebilir ki, bu da suç kapsamına giren bir eylem değildir. Çünkü, manipülasyonda fiyat mekanizmasını hedef alan bir eylem organize edilmişken, spekülasyonda büyük sermayeye sahip yatırımcının sürü psikolojisinden faydalanarak kazanç sağlama esastır. Dolayısıyla, spekülasyonda alım satım ilişkileri bakımından 3. basamakta belli hesaplar arasında yoğunlaşma beklenmemektedir. Hesaplar arasında alım-satım ilişkileri bakımından bir yoğunlaşmanın varlığı ise manipülasyona işaret etmektedir.

Tasarlanan erken uyarı sisteminin iş-akış diagramı Şekil 6.1’de verilmektedir. Ayrıca, erken uyarı sisteminin iş akışında kullanılabilecek yöntem alternatifleri, uygulanan yöntemin tercih nedeni ve diğer yöntemlerin tercih edilmeme nedenleri Çizelge 6.1’de; uygulanan yöntemler, kullanılan değişkenler ve yöntemin gerçekleştirildiği yazılımlar da Çizelge 6.2’de verilmektedir.



Şekil 6.1 Erken uyarı sistemi iş akış diagramı

Çizelge 6.1 Erken uyarı sistemi iş akışının işlem ve yöntem değerlendirmesi

1. BASAMAK	
1. ETAP	
İŞLEM	Fiyat ve/veya miktar hareketlerinde, yüksek düzeyde değişim, orta düzeyde değişim ve düşük düzeyde değişim gösteren işlemleri içeren hisse senedini/senetlerini tespit ederek; diğerlerinden ayrışanların tespitini sağlayacak gruplama yapma.
Yöntem Alternatifleri	K-ortalamlar Kümeleme Analizi K-en Yakın komşuluk yöntemi Hiyerarşik Kümeleme Analizi Kendi Kendini Düzenleyen Haritalar
Tercih Edilen Yöntem ve Tercih Nedeni	K-ortalamlar Kümeleme Analizi, N gözlemi istenen belirli sayıda K gruba, belirli bir benzemezlik (benzerlik) ölçütüne göre ayırmakta kullanıldığı için tercih edilmiştir.
Diğer Yöntemlerin Tercih Edilmeme Nedeni	Denetimli bir yöntem olan K-en yakın komşuluk yöntemi, komşuluğu sorgulanabilecek bir hedef nokta belirlemek mümkün olmadığından, Denetimsiz bir yöntem olan Hiyerarşik Kümeleme Analizinde, çıktı bölgesini, bir başka deyişle küme sayısını belirlemek mümkün olmadığından ve karar verirken insan yargısı kullanımı gerektiğinden, Denetimsiz bir yöntem olan Kendi Kendini Düzenleyen Haritalar ise sınıflandırmasında zorluk çekilen küresel vb. yüzeye sahip bir veri söz konusu olmadığı için tercih edilmemiştir.
2. ETAP	
İŞLEM	Şüpheli işlemlerin, fiyat değişim oranına göre yüksek değişim ve düşük değişim olmak üzere iki gruba ayrılarak, her bir işlemin gruplara aidiyet derecelerinin belirlenmesi.
Yöntem Alternatifleri	K-ortalamlar Kümeleme Analizi Hard c-ortalamlar Kümeleme Analizi FANNY Bulanık c-Ortalamlar Kümeleme Algoritması Diğer Bulanık c-ortalamlar Kümeleme Algoritmaları
Tercih Edilen Yöntem ve Tercih Nedeni	FANNY Bulanık c-Ortalamlar Kümeleme Algoritması, N gözlemi eşanlı olarak c farklı kümeye farklı aidiyet (üyelik) dereceleriyle atamaktadır. Böylece her bir gözlemin her bir gruptaki temsil kabiliyetini gözlemek mümkün olmaktadır.
Diğer Yöntemlerin Tercih Edilmeme Nedeni	K-ortalamlar Kümeleme Analizi, üyelik derecelerini vermediği için, Hard c-Ortalamlar Kümeleme Algoritması sadece bir tek kümeye aidiyete izin verdiği için, Diğer Bulanık Kümeleme Algoritmaları ise FANNY kadar dirençli olmadığı için tercih edilmemiştir.
2. BASAMAK	
İŞLEM	Hisse Senetleri (HS) için gerçekleşen manipülatif şüphesi taşıyan işlemlerin; belli aracı kuruluşlar (üyeler) üzerinde yoğunlaşp yoğunlaşmadığı ve yoğunlaşma varsa hangi yatırımcıların bu işlemleri gerçekleştirdiği sorgulanacaktır.
Yöntem Alternatifleri	CHAID Karar Ağacı Algoritması, CART Karar Ağacı Algoritması Birliktelik kuralı algoritmaları
Tercih Edilen Yöntem ve Tercih Nedeni	CHAID karar ağacı algoritması ağaç görünümüyle sunduğu görsel destek ile karar verme sürecinde kolaylık sağlamakta ve diğer karar ağaçları algoritmalarından farklı olarak ikiden fazla dallanarak ilerlediği için tüm alt gruplarla ilişkileri ortaya koymaktadır.

Çizelge 6.1 Erken uyarı sistemi iş akışının işlem ve yöntem değerlendirilmesi (Devam)

2. BASAMAK (Devam)	
Diğer Yöntemlerin Tercih Edilmeme Nedeni	CART karar ağacı algoritması ikili (binary) olarak dallandığı için tüm alt grupları sunmadığından, Birliktelik kuralları algoritmaları ise görsel desteğe sahip olmadığından, karar verme sürecinde tüm ilişki kuralları arasından seçim yapmak gerektiğinden
3. BASAMAK	
İŞLEM	Manipülatif işlem şüphesi taşıyan ayrışan davranış örüntüsüne sahip işlemlerin belli müşteri ya da müşteriler adına gerçekleştirilip gerçekleştirilmediği sorgulanacaktır.
Yöntem Alternatifleri	Önsel birliktelik kuralları algoritması Karar ağaçları algoritmaları
Tercih Edilen Yöntem ve Tercih Nedeni	Önsel birliktelik kuralları algoritması, tüm birliktelik kurallarını taramak yerine belirli düzeyin üzerinde birlikteliğe sahip kuralları ortaya çıkarmakta ve çok hızlı sonuç vermektedir.
Diğer Yöntemlerin Tercih Edilmeme Nedeni	Karar ağaçları algoritmaları görsel destek sağlamasına karşın, birliktelik kuralları kadar hızlı çalışmamakta ve karar vermeyi gerektirecek bir seçim süreci gerektirmediği için kullanılmamıştır.

Çizelge 6.2 Erken uyarı sisteminde uygulanan analizler ve kullanılan yazılımlar

SÜREÇ		DEĞİŞKENLER	İŞLENEN DEĞİŞKENLER	ANALİZ YÖNTEMİ	KULLANILAN YAZILIM
1. BASAMAK	1. ETAP	Zaman Menkul Fiyat Alan üye Satan üye Alan hesap no Satan hesap no	Fiyat değişim oranı (fdo)	K-ortalamalar Kümeleme Analizi	SPSS 12.0
	2. ETAP	Fiyat değişim oranı	Fiyat değişim oranı	FANNY Bulanık c-Ortalamalar Kümeleme Algoritması	S-Plus 6.2
2. BASAMAK		Bulanık fiyat değişim oranı Alan üye Satan üye Alan hesap no Satan hesap no	Bulanık fiyat değişim oranı Alan üye Satan üye Alan hesap no Satan hesap no	Bulanık Hedefli CHAID Karar Ağacı Algoritması	SPSS Answer Tree 3.1
3. BASAMAK		Alan hesap no Satan hesap no	Alan hesap no Satan hesap no	Önsel Birliktelik Kuralları Algoritması	TANAGRA

7. BULANIK VERİ MADENCİLİĞİNE DAYALI İŞLEM MANİPÜLASYONU TESPİTİNE YÖNELİK ERKEN UYARI SİSTEMİNİN UYGULAMASI

Tasarlanan erken uyarı sisteminin başarısının ölçümü için İMKB HSP'de fiyat manipülasyonu yapıldığı belirlenmiş bir hisse senedinin sistemce tespit edilme başarısı sorgulanacaktır.

Bu sorgulama için söz konusu hisse senedinde (HSX) manipülasyon yapıldığı güne (T) ait bütün işlem verileri değerlendirmeye alınmıştır. Sistemin amacı manipülasyon yapılan hisse senedi HSX'i, manipülasyonda dahil bulunan alan ve satan üyelerle, alan ve satan yatırımcıları doğru olarak tespit etmektir.

T tarihinde İMKB HSP'de;

- 291 hisse senedi için,
- 114 üye (aracı kuruluş) tarafından,
- 99.287 adet hesap numarası (yatırımcı) için alım işlemi,
- 99.553 adet hesap numarası (yatırımcı) için satım işlemi,
- 100.614 işlem yapılarak,

gerçekleştirilmiştir.

Gerçekleşen 100.614 işlem için her bir hisse senedinin bir önceki işleme göre fiyat değişim oranı hesaplanmıştır. Bu hesaplama neticesinde her bir hisse senedinin ilk işlemi ve sadece bir tek işlem gerçekleşen hisse senedi işlemleri göz ardı edilmiş ve işlemler 100.319 kayıt ile gerçekleştirilmiştir.

Erken uyarı sistemi için hız faktörünün önemi göz önüne alındığından, verileri mümkün olduğunca eleyerek gitmek biçiminde bir yol izlenmiştir. Bu nedenle öncelikle veri içerisinde normal kabul edilebilecek değişim miktarına sahip verileri ayıklamak hedeflenmiştir.

1. basamak işlem etüdü iki aşamadan meydana gelmektedir. 1. aşamada hisse senetlerinin fiyat değişim oranına göre 3 parçaya bölünmesi ve bu bölünüm sonucunda;

- Düşük düzeyde fiyat değişimine sahip işlemler,
- Orta düzeyde fiyat değişimine sahip işlemler,
- Yüksek fiyat değişimine sahip işlemler,

gruplarına ulaşmak için hareket edilmiştir. Bu noktadan hareketle, fiyat değişim oranı

- 3-Ortalamalar Kümeleme Analizi

uygulanarak 3 kümeye ayrılmış olup; orta düzeyde fiyat değişimine sahip işlemler kümesinde bir yığılma olması beklenmiş ve bu kümenin normal veya manipülatif olmayan işlemlere sahip küme olduğu düşünülmüştür. Diğer iki kümenin ise normalden düşük veya normalden yüksek fiyat değişim oranına sahip işlemler olması nedeniyle şüpheli işlemler olduğu düşünülecektir. Bu noktadan hareketle, fiyat değişim oranına 3-ortalamalar Kümeleme Analizi SPSS 12.0 paket programının kullanımıyla uygulanmış ve elde edilen kümeleme analizi sonuçları Çizelge 7.1 ve Çizelge 7.2’de verilmiştir.

Çizelge 7.1 3-ortalamalar kümeleme analizi sonucunda elde edilen son küme merkezleri

	Kümelere		
	1	2	3
Fiyat Değişim Oranı	32,9536	0,0013	-24,7523

Çizelge 7.2 3-ortalamalar kümeleme analizi sonucunda elde edilen her bir kümedeki gözlem sayıları

Küme	1	109
	2	100.101
	3	109
Geçerli gözlem sayısı		100.319
Kayıp gözlem		0

3-ortalamalar kümeleme analizi sonucunda elde edilen Çizelge 7.1 ve Çizelge 7.2'ye göre 100.319 kayıt;

- Düşük düzeyde fiyat değişimine sahip işlem sayısı = 109,
- Orta düzeyde fiyat değişimine sahip işlem sayısı = 100.101,
- Yüksek fiyat değişimine sahip işlem sayısı = 109,

olarak kümelendi. Dolayısıyla, 100.101 adet normal işlem gerçekleşirken, toplam 218 adet şüpheli işlem gerçekleşmiştir.

218 adet şüpheli işlemin ait olduğu hisse senetlerinin dağılımına bakılarak, belli hisse senetlerinde yoğunlaşma olup olmadığı sorgulanmıştır. Bu sorgulama sonucunda 218 adet şüpheli işlemin de aynı hisse senedi için gerçekleştiği gözlenmiştir. Bir başka deyişle, fiyat değişim oranı diğer hisse senetlerinden ayrılan 218 işlem sadece bir hisse senedinde (HSX) gerçekleşmiştir. Bu durum, söz konusu hisse senedinde manipülatif işlem şüphesini kuvvetlendirmiş olup, 1. basamak işlem etüdünde ilgili hisse senedi için sarı alarm durumuna geçilmiştir. Ayrıca, HSX hisse senedi için gerçekleşen tüm işlemler incelendiğinde tüm gün boyunca toplam 7.623 adet işlem gerçekleştiği ve 7.622 fiyat değişiminden sadece 218 adetinin şüpheli olduğu belirlenmiştir.

1. basamak işlem etüdünün, 2. etabı HSX hisse senedi için şüphe taşıyan 218 adet işlem dikkate alınarak gerçekleştirilmiştir. Bir başka deyişle gerçekleşen 100.614 işlem için hesaplanan 100.319 fiyat değişim oranından, şüphe taşıyan toplam 218 adet işlem dikkate alınmıştır.

İkinci etapta, 1. basamak işlem etüdünün birinci etabında şüpheli olarak tespit edilen 218 adet işlemin buldukları kümelere aidiyet dereceleri, S-PLUS 6.2 yazılımında yer alan

- FANNY Bulanık 2-ortalamalar Kümeleme algoritması

ile incelenmiştir. Bu algoritmanın 2 gruba ayırtmak için uygulanmasının nedeni,

- Düşük düzeyde fiyat değişimine sahip işlemler,
- Yüksek fiyat değişimine sahip işlemler,

olmak üzere şüpheli işlemlerin iki gruptan oluşmasıdır. Kesin K-Ortalamlar Kümeleme Analizi'nin yerine Bulanık C-Ortalamlar Kümeleme Analizi'nin uygulanmasının nedeni ise gözlemlerin iki kümeye ayrılmasına karşın, her bir gözlemin kümelere aidiyet derecesinin farklı olması ve bir anlamda bulunduğu kümenin özelliklerini aidiyet derecesi kadar temsil etmesidir. Ayrıca, her bir gözlem için iki kümeye de aidiyet dereceleri ayrı ayrı verilmektedir. Her bir gözlem, toplamı 1 olan farklı üyelik fonksiyonu değerleriyle veya aidiyet dereceleriyle her iki kümede de temsil edilmektedir. Aynı biçimde, 218 adet işlemde toplamı 1 olan iki adet üyelik değeriyle iki kümeye de aidiyet göstermektedir. İşlemlerin, aidiyet derecesinin daha büyük olduğu kümeleri daha iyi temsil ettiği noktasından hareketle $\mu_{1,i}$ ve $\mu_{2,i}$ sırasıyla, i . gözlem için birinci ve ikinci kümeye aidiyet dereceleri olmak üzere,

$$w_i = \text{Max}(\mu_{1,i}, \mu_{2,i}) \quad i = 1, 2, \dots, 218$$

her bir fiyat değişim oranının bulunduğu kümeyi en büyük temsil etme düzeyi elde edilmiştir. Dolayısıyla fiyat değişim oranları (fdo) kesin değerleriyle değil, üyelik dereceleriyle kümeye aidiyetlere düzeyinde ağırlıklandırılarak;

$$bfdo_i = w_i fdo_i \quad i = 1, 2, \dots, 218$$

biçiminde bulanık fiyat değişim oranı haline getirilmiştir.

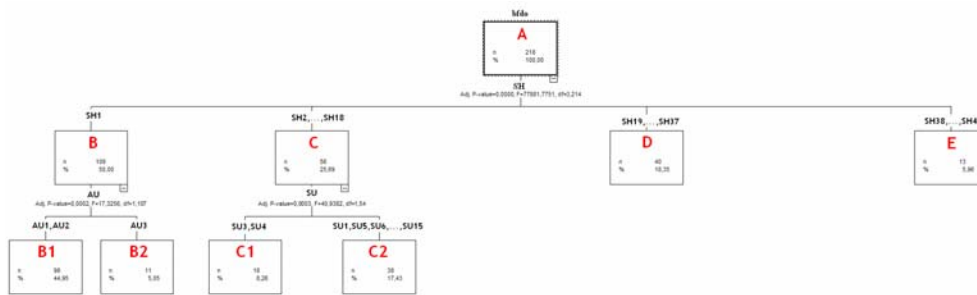
Şüpheli 218 adet işlem için fiyat değişim oranının kesin değeri fdo, iki küme için üyelik fonksiyonu değerleri $\mu_{1,i}$ ve $\mu_{2,i}$, kümelere en büyük aidiyet derecesi w_i ve üyelik fonksiyonlarıyla ağırlıklandırılarak elde edilen bulanık fiyat değişim oranı bfdo Ek 1'de verilmiştir. FANNY algoritmasının uygulanması sonucu elde edilen Normalize edilmiş bulanıklık katsayısı olan,

- Dunn Katsayısı $k = 0,966189$

elde edilmiştir. Normalize edilmiş Dunn katsayısı k 'nın $[0,1]$ aralığında gerçekleştiği ve sifıra yaklaştıkça bölünümün bulanıklaştığı, bire yaklaştıkça da kesinleştiği bilindiğine göre bölünümün kesine yakın gerçekleştiği görülmektedir.

Şüphesiz ki, FANNY algoritmasından elde edilen üyelik fonksiyonlarının dışında da üyelik fonksiyonu atamak mümkündür. Ancak, FANNY Bulanık C-ortalamalar Kümeleme Analizi sonucu elde edilen üyelik fonksiyonlarının kullanımıyla; analiz ve analizde kullanılan verileri bütünsel olarak sonraki işlem etüdüne sevk etmek mümkün olmaktadır. Böylece, uygulanan FANNY Bulanık C-ortalamalar Kümeleme Analizi ile elde edilen üyelik katsayılarıyla bulanıklaştırılmış fiyat değişim oranlarının CHAID Karar Ağacı algoritmasında uygulanması entegre biçimde gerçekleşebilmiştir.

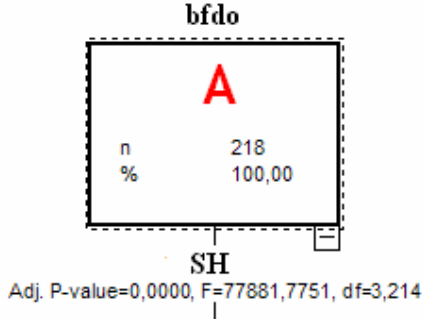
2. basamak işlem etüdünde bulanık fiyat değişim oranı $bfdo$ hedef değer, alan üye, alan hesap no, satan üye ve satan hesap no tahmin edici değişkenler olmak üzere, hedef değeri bulanıklaştırılmış CHAID karar ağacı algoritması kullanılarak alan üye ve satan üye de bir yoğunlaşma olup olmadığına bakılmış ve üyelerde bir yoğunlaşma varsa bu yoğunlaşmanın hangi hesap numarasının işlemlerinde yoğunlaştığı tespit edilmesi hedeflenmiştir. Hedef değeri bulanık CHAID algoritmasının SPSS Answer Tree yazılımında uygulanması sonucu elde edilen karar ağacı aşağıda Şekil 7.1'de verilmiştir.



Şekil 7.1 Hedef değeri bulanık CHAID karar ağacı

Şekil 7.1’de yer alan karar ağacından aşağıdaki çıkarımlar yapılabilmektedir:

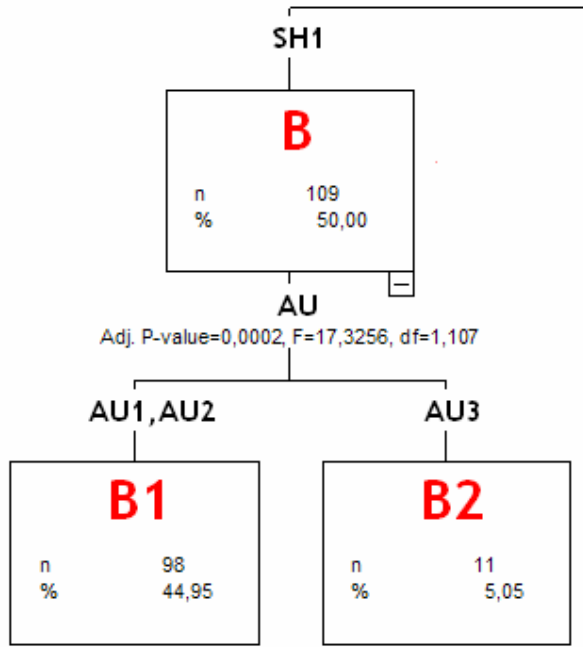
A. Bulanık fiyat değişim oranı (bfdo) hedef değeri ile en güçlü ilişki ($p < 0,00001$), satan hesap numarası (SH) değişkeni arasında gerçekleşmektedir. SH değişkeni için veri dört farklı gruba ayrılmaktadır.



B. SH değişkenininin birinci grubunu oluşturan SH1 hesap nolu yatırımcının şüpheli işlemlerin yarısını oluşturan pozitif yöndeki işlemlerin tamamını (109 işlem) gerçekleştirdiği görülmektedir. SH1 ile alan üye (AU) değişkeni olmak üzere sadece bir değişkenin çok güçlü ilişkide ($p = 0,0002$) olduğu ve değişkenin iki gruba ayrıldığı görülmektedir.

B1. Birinci grupta AU1 ve AU2 olmak üzere iki üyenin toplam 109 işlemin 98’ini gerçekleştirdiği görülmektedir.

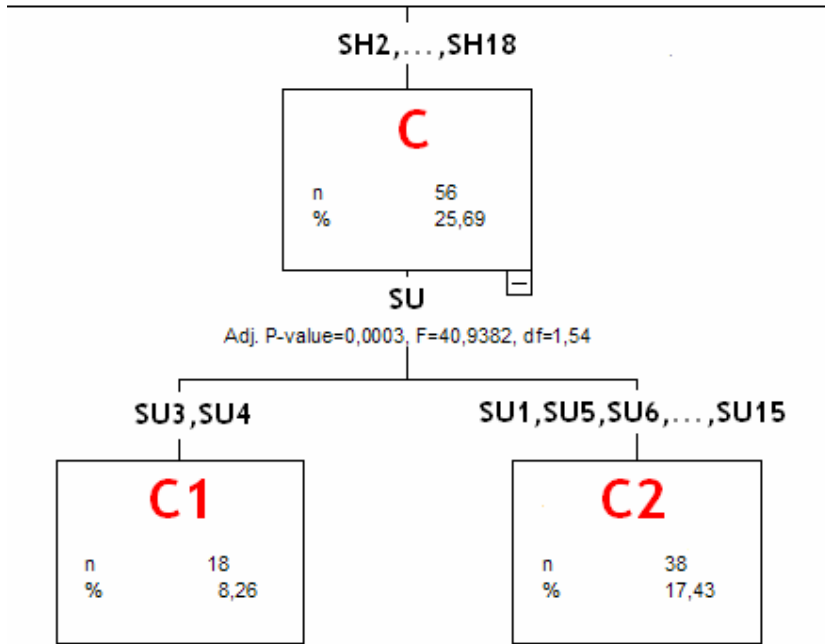
B2. İkinci grupta ise sadece AU3’ün yer aldığı ve 109 işlemde geriye kalan 11’ini gerçekleştirdiği görülmektedir.



C. SH deęişkeninin ikinci grubunu, SH2'den SH18'e kadar toplam 17 adet yatırımcının oluşturduęu ve Őüpheli işlemlerin negatif yarısını oluřturan 109 işlemden 56'sının bu gruptaki yatırımcılar tarafından gerekleřtirildięi grlmektedir. SH deęişkeninin ikinci grubunun satan üye (SU) deęişkeninden SU3, SU4 ve SU1, SU5, SU6, ... , SU15 yelerini ieren iki grpla ok gl iliřkiye sahip olduęu ($p=0,0003$) grlmektedir.

C1. SU3 ve SU4 olmak zere 2 satın alan ye tarafından toplam 56 işlemin 18'inin gerekleřtirildięi grlmektedir.

C2. SU1, SU5, SU6, ... , SU15 olmak zere 12 satın alan ye tarafından 56 işlemden geriye kalan 38'inin gerekleřtirildięi grlmektedir.



D. SH deęişkeninin üçüncü grubunu, SH19'dan SH37'e kadar toplam 19 adet yatırımcının oluşturduğu ve şüpheli işlemlerin negatif yarısını oluşturan 109 işlemden 40'nın bu gruptaki yatırımcılar tarafından gerçekleştirildięi görülmektedir. Bu grupta başka bir deęişkenle önemli düzeyde kuvvetli ilişkiye sahip başka bir deęişken görülmemektedir.

E. SH deęişkeninin üçüncü grubunu, SH38'den SH48'e kadar toplam 19 adet yatırımcının oluşturduğu ve şüpheli işlemlerin negatif yarısını oluşturan 109 işlemden 13'ünün bu gruptaki yatırımcılar tarafından gerçekleştirildięi görülmektedir. Bu grupta başka bir deęişkenle önemli düzeyde kuvvetli ilişkiye sahip başka bir deęişken görülmemektedir.



Şekil 7.1'deki karar ağacı yukarıdaki yorumların ışığında incelendiğinde söz konusu hisse senedi HSX için karar ağacının B dalında belirgin bir yoğunlaşma görülmektedir. Üyeler açısından değerlendirildiğinde, AU1, AU2 ve AU3 üyelerinde belirgin bir yoğunlaşma dikkat çekmektedir. Ağacın diğer dallarında ise böyle bir yoğunlaşma gözlenmemektedir. AU1, AU2 ve AU3 üyelerindeki belirgin yoğunlaşmanın yanı sıra alan üyelerden sadece satan hesaplardan SH1 nolu yatırımcının şüpheli 109 işlemin tamamını gerçekleştirdiği de görülmektedir. Dolayısıyla, belli üyelerde belirgin bir yoğunlaşma gözlenmesi nedeniyle erken uyarı sistemi turuncu alarm durumuna geçirilerek 3 basamak işlem etüdünü işletmek gerekmektedir. 3. basamak işlem etüdünde satan yatırımcı SH1 nolu yatırımcının gerçekleştirdiği işlemler ile alan yatırımcılar gerçekleştirdiği işlemler arasındaki ilişki sorgulanacaktır.

3. Basamak işlem etüdünde, 2 basamak işlem etüdünde üzerinde yoğunlaşma görülen alan üyeler AU1, AU2 ve AU3 üyelerinin işaret ettiği satan hesap numarası SH1 olan yatırımcının satan hesap no ve alan hesap no olmak üzere diğer yatırımcılarla ilişkileri sorgulanacak ve başta SH1 nolu yatırımcı olmak üzere, belirli yatırımcı ya da arasında bir ilişki yoğunlaşmasının varlığı aranacaktır. Alan ve satan hesap numaraları arasındaki birliktelik arayışı 3 üyenin yoğunlaştığı ve bir hesaba işaret ettiği 109 işlem üzerinde gerçekleştirilmiştir.

Bu sorgulama,

- Alan hesap no
- Satan hesap no

değişkenleri arasındaki mevcut ilişkilerin varlığını tespit etmek üzere,

- Önsel (Apriori)

birliktelik kuralı algoritması kullanılarak sorgulanmıştır. Bu sorgulama sonucunda elde edilen sonuçlar aşağıda sunulmuştur.

109 işlem üzerinde Tanagra Yazılımı ile gerçekleştirilen Önsel (Apriori) birliktelik kuralı analizi sonucunda 2 birliktelik kuralı elde edilmiştir. Elde edilen birliktelik kuralları Çizelge 7.3'de verilmektedir.

Çizelge 7.3 Birliktelik kuralları

No	Öncül	Ardıl	Destek	Güven	Kaldıraç
1	AH=AH3	SH=SH1	85,3	100,0	100,0
2	SH=SH1	AH=AH3	85,3	85,3	100,0

Çizelge 7.3’de görülen karar kurallarının yorumları aşağıda sunulmaktadır.

1. birliktelik kuralında öncül, alan hesaplardan AH3 nolu hesaba sahip yatırımcı; ardıl ise satan hesaplardan SH1 nolu hesaba sahip yatırımcıdır. Destek % 85,3 olarak görülmektedir ki, 109 işlemin %85,3’ünde alan hesap olarak AH3 hesap nolu yatırımcı ve satan hesap olarak SH1 hesap nolu yatırımcının birlikte yer aldığını göstermektedir. Güven %100,0 olarak görülmektedir ki, alan hesap olarak AH3 hesap nolu yatırımcının öncül olarak bulunduğu işlemlerin %100’ünde veya bir başka deyişle tamamında satan hesap olarak SH1 nolu yatırımcı ardıl olarak yer almaktadır. Kaldıraçın %100 olduğu görülmektedir ki, SH1’in ardıl olduğu işlemlerin %100’ünde AH3’ün öncül olduğunu göstermektedir.

2. birliktelik kuralında öncül, satan hesaplardan SH1 nolu hesaba sahip yatırımcı; ardıl ise alan hesaplardan AH3 nolu hesaba sahip yatırımcıdır. Destek % 85,3 olarak görülmektedir ki, 109 işlemin %85,3’ünde satan hesap olarak SH1 hesap nolu yatırımcı ve alan hesap olarak AH3 hesap nolu yatırımcının birlikte yer aldığını göstermektedir. Güven %85,3 olarak görülmektedir ki, satan hesap olarak SH1 hesap nolu yatırımcının öncül olarak bulunduğu işlemlerin %85,3’ünde alan hesap olarak AH3 nolu yatırımcı ardıl olarak yer almaktadır. Kaldıraçın %100 olduğu görülmektedir ki, AH3’ün ardıl olduğu işlemlerin %100’ünde SH1’in öncül olduğunu göstermektedir.

Yukarıdaki birliktelik kuralları ve yorumları incelendiğinde, 109 işlemin %85,3’ünde AH = AH3 ise SH = SH1 ve aynı şekilde 109 işlemin %85,3’ünde SH = SH1 ise AH = AH3 olduğu görülmektedir ki; 109 işlemin %85,3’ünde,

$$SH = SH1 \Leftrightarrow AH = AH3$$

olduđu açıkça gör÷lmektedir. Dolayısıyla satan hesap no SH1 ile alan hesap no AH3 arasında oldukça yüksek bir alım-satım ilişkisi gör÷lmektedir ki, bu ilişki şüpheli ilişkilerde açığa çıkmaktadır. Yatırımcılar arasında şüpheli işlemler açısından bu denli güçlü bir ilişkiyle, manipülasyonun son halkası olan yatırımcılar arasındaki ilişkinin varlığı da kanıtlanmıştır.

Erken uyarı sisteminin tüm süreçleri gözden geçirilirse:

1. basamak işlem etüdünün birinci etabında şüpheli hisse senedi HSX ve şüpheli 218 işlem tespit edilmiş olup sistem sarı alarm durumuna geçirilmiştir. İkinci etapta şüpheli işlem kümesine aidiyet derecesini gösteren her bir işleme ait değişim miktarına bağlı üyelik dereceleri hesaplanmış ve şüpheli işlemler 2. basamak işlem etüdüne bulanık fiyat değişim oranlarıyla sevk edilmiştir.

2. basamak işlem etüdünde şüpheli üyeler AU1, AU2 ve AU3 ve bu üyeler vasıtasıyla işlem yapan yatırımcı SH1 tespit edilmiş olup; sistem turuncu alarm durumuna geçirilmiştir. Yoğunlaşma tespit edilen şüpheli 109 işlemde tespit edilen üyeler kanalıyla işlem yapan yatırımcıya ait SH1 hesabı dikkate alınarak 3. basamak işlem etüdüne geçilmiştir.

3. Basamak işlem etüdünde, 2. basamak işlem etüdünde şüpheli işlemleri gerçekleştiren SH1 hesap nolu yatırımcı ile şüpheli işlemleri gerçekleştiren diğer yatırımcılar arasında birliktelikler sorgulanmış ve bu sorgulama sonucunda AH3 hesap nolu yatırımcının birlikteliği tespit edilmiştir. Bu tespit sonucunda sistem kırmızı alarma durumuna geçirilmiş ve işlem manipülasyonu gerçekleştirildiğine dair uyarı ilgili birimlere iletilmiştir.

Erken uyarı sisteminin işlevini yerine getirmesi açısından sistemin işlem zamanı da oldukça büyük önem taşımaktadır. Bu nedenle erken uyarı sisteminin toplam işlem zamanı açısından da değerlendirilmesi gerektiği düşünülmektedir.

Erken uyarı sisteminin işlem zamanı, uygulanan yöntemler açısından değerlendirildiğinde,

- K-ortalamlar kümeleme analizinin SPSS 12.0 paket programında uygulanmasına yönelik işlem zamanı 1,56 saniye,
- Bulanık C-ortalamlar kümeleme algoritmasının S-Plus 6.2 yazılımındaki işlem zamanı 0,047 saniye,
- Bulanık hedefli CHAID karar ağacı algoritmasının SPSS Answer Tree yazılımındaki işlem zamanı 3,53 saniye,
- Önsel (A priori) birliktelik kuralı algoritmasının Tanagra Yazılımındaki işlem zamanı 0,01 saniye,

olarak gerçekleşmiştir.

Görüldüğü gibi, erken uyarı sisteminin toplam işlem zamanı 5,147 saniye olarak gerçekleşmekte ve sistemin işlem hızı olarak da işlevini yerine getirdiği görülmektedir.

Sistemin işleyişindeki başarıyı test etmek için aynı veri seti ikiye bölünerek;

- 50.119 kayıt için

aynı sürecin işletilmesi düşünülmüştür. SPSS 12.0 yazılımı ile K-ortalamlar Kümeleme Analizi K=3 olarak sözkonusu kayıtlar için işletilmiş ve elde edilen sonuçlar Çizelge 7.4 ve Çizelge 7.5'te verilmektedir.

Çizelge 7.4 3-ortalamlar kümeleme analizi sonucunda elde edilen son küme merkezleri

	Kümeler		
	1	2	3
Fiyat Değişim Oranı	32,9536	0,0013	-24,7523

Çizelge 7.5 3-ortalamalar kümeleme analizi sonucunda elde edilen her bir kümedeki gözlem sayıları

Küme	1	109
	2	49.901
	3	109
Geçerli gözlem sayısı		50.119
Kayıp gözlem		0

Çizelge 7.4 ve Çizelge 7.5’den elde edilen sonuçlar incelendiğinde sistemin 100.319 kayıt içeren veri setinin tamamında yakaladığı şüpheli 218 işlemi 50.119 işlem içerisinde de yakaladığı gözlenmiştir. Şüpheli olarak tespit edilen 218 işlemin her iki veri seti içinde aynı işlemler olduğu belirlenmiştir. Bu nedenle 2. ve 3. adımları işletmeye gerek kalmamıştır. Dolayısıyla, sistemin başarıyla çalıştığı teyit edilmiştir.

Tezcanlı’nın (1993) sermaye piyasaları; Çilli ve Temel (1988) ile Gaytan and Johnson’un (2002) bankacılık; Gerni vd. (2005) ise ekonomik krizler için önerdikleri gözetim veya erken uyarı sistemleri incelendiğinde, finansal erken uyarı sistemlerinin geçmiş tecrübelerle elde edilen kalıplar üzerine sabit bir model kurmaya dayalı olduğu ve sistemlerin hiçbir esnekliğe sahip olmadığı gibi karar verme konusunda insan faktörünün çok etkin olduğu da görülmektedir. Esnekliğin olmaması ve insan faktörünün etkinliği subjektiviteyi artıran faktörler olarak sistemlerin temel zaafalarını oluşturmaktadır.

Sözkonusu sistemler geçmiş tecrübelerle dayalı ekonometrik modeller kurarak, anormal davranışı tespiti çalışmaktadır ve gözlenmemiş yeni bir durumla karşılaşıldığında sistemler bunu algılayamamaktadır. Buna karşın, bu tez kapsamında tasarlanan erken uyarı sistemi, bir kalıp model kurmak yerine veriden öğrenmeyi esas alarak, veri içerisindeki anomalileri belirleyerek manipülasyon tespitini gerçekleştirmektedir. Sağlanan esneklik, geçmişte rastlanmayan yeni bir durumla karşılaşıldığında İstatistiksel Öğrenme tekniklerinin kullanımıyla tespit edilmekte, kalıplara ihtiyaç duyulmamaktadır. Diğer sistemlerde olduğu gibi geçmişteki anomali örneklerinden

model uydurarak deęil, manipölasyonun kendi tanımından gidilerek sistem tasarlanmıřtır.

Tezcanlı'nın (1993) sistemi bilançolara dayalı olarak her bir hisse senedi için detaylı birtakım hesaplamalardan sonra, manipölasyon var yok gibi tanımlamalar yapmaktadır. Tasarlanan sistem ise manipölasyon içeren hisse senedini tespit edebildięi gibi, eşanlı olarak manipölatif işlemleri belirleyip, manipölasyonu gerçekleřtiren aracı kuruluş(lar) ve manipölasyonu gerçekleřtiren yatırımcı(ları)yı da belirleyebilmektedir.

New York Borsası'nda kullanılan ASAM veritabanı bünyesinde, işlem gerçekleřtirenler, ihraççılar ile yöneticileri, aracı kurumlar ile yöneticileri ve ilgili olabilecek dięer piyasa katılımcılarının tamamına yönelik bilgi bulunması, ihtiyaç halinde birbirleriyle ilişkileri bulunabilecek kiři ve kurumların otomatik olarak belirlenebilmesine imkan tanımaktadır. Ancak, halen İMKB bünyesinde HSP paydařlarına ait böyle bir veritabanı bulunmaması nedeniyle tasarlanan sistemde de böyle bir bileřen yer almamaktadır. Tasarlanan sisteme, ASAM benzeri bir veritabanının da bir bileřen olarak eklenmesi; řüphesiz ki sistemin etkinlięini daha da artıracaktır.

8. SONUÇ ve ÖNERİLER

Tez çalışmasının sonuç ve öneriler kısmı, tez çalışmasının genel değerlendirmesinin yapıldığı Ardıl Değerlendirme ve Sonraki Çalışmalara Öneriler olmak üzere iki başlık altında değerlendirilmektedir.

8.1 Ardıl Değerlendirme

Veri madenciliğinin temel malzemesi olan büyük hacimde veri ön koşulu, işlem verileri nedeniyle finans kesimi için oldukça uygun bir yapı sunmaktadır. Ancak, veri madenciliğinin finansal uygulamaları incelendiğinde bankacılık üzerine bir yoğunlaşma olduğu gözlenirken; sermaye piyasaları üzerine çok fazla çalışma yapılmadığı gözlenmektedir. Veri madenciliği uygulamalarının bankacılık sektörü üzerine yoğunlaşmasındaki öncelikli nedenlerden birisinin bankacılık sisteminin dünyadaki ağırlığı olduğu düşünülmektedir. Her ne kadar, ekonomik krizlerden çıkarılan en önemli derslerden birisi, güçlü bir ekonomik sistem için sermaye piyasalarının etkinliğinin artırılması yönünde olsa da, bankacılık sistemine göre görece olarak yeni bir piyasa olması ile daha karmaşık ve riskli bir yapıya sahip olması yatırımcıyı bankalara yönlendirmektedir. Bununla beraber, son dönemlerde global anlamda sermaye piyasalarına yöneliş olduğu da açıktır.

Sermaye piyasalarına yönelimde yatırımcı açısından cesaret kırıcı bir başka faktör de, fiyat mekanizmasını bozan manipülasyondur. Sermaye piyasaları açısından hayati öneme haiz olan fiyat mekanizmasının çıkar amaçlı ve kasıtlı olarak bozulması, yatırımcılar açısından oldukça caydırıcı olmakta ve sermaye piyasalarının temel hedeflerinden olan sermayenin tabana yayılmasının önündeki en büyük engellerden birisi olarak ortaya çıkmaktadır.

Zaman içerisinde manipülasyonu önleyici bir dizi tedbir alınsa da, işlem manipülasyonu denilen manipülasyon tipini önleyici bir mekanizma geliştirmek mümkün olmamaktadır. Bu durumda, işlem manipülasyonu gerçekleşmesinin ivedilikle

belirlenmesi yoluna gidilmekte ve fiyat mekanizmasının bozulmasının, bir başka deyişle piyasa işleyişinin normale döndürülmesinin hız kazandırılması çalışmalarına ağırlık verilmektedir. Bu çalışmaların neticesinde teknolojinin gelişiminden de faydalanarak manipülasyon tespitini mümkün olduğunca hızlı olarak gerçekleştirecek erken uyarı sistemleri üzerinde çalışılmaktadır. Erken uyarı sistemleri mantık olarak, insanların gözlem ve elle yaptıkları hesaplamaları otomasyona bağlamak olarak şekillenmektedir. Bir anlamda, veri içerisinden gizli bilgilerin otomatik olarak açığa çıkarılması olarak da yorumlanabilir ki, veri madenciliği tanımının büyük hacimde veri ifadesinin tamamlayıcısı olmaktadır.

Sermaye piyasalarının en gelişmiş olduğu örneklerden birisi olan ABD’de erken uyarı sistemlerinde veri madenciliğinden faydalandığı bilinmektedir. Bu çalışmada da hisse senetleri piyasası için veri madenciliğine dayalı olarak işlem manipülasyonu tespiti sağlayan bir erken uyarı sistemi tasarlanması hedef alınmıştır. Hisse senetleri piyasasının kaotik yapısına uygun olarak veri madenciliği uygulaması için bulanık yaklaşım izlenmiştir. Dolayısıyla, bulanık veri madenciliğinin kullanımıyla olabildiğince hızlı bir erken uyarı sistemi tasarlanmıştır.

En basit şekilde, piyasaların ve dolayısıyla fiyatların yapay olarak düşürülmesi, yükseltilmesi ya da aynı seviyede tutulması olarak tanımlanabilecek manipülasyon; bilerek ve isteyerek, menkul kıymet fiyatlarını kontrol etmek veya yapay şekilde etkilemek suretiyle yatırımcıları aldatmayı veya dolandırmayı amaçlayan davranışları ifade eder. Uluslararası kabule sahip manipülatif davranış kalıpları olmasına karşın; her somut olayda farklı yöntemlerin kullanılması mümkün olup; manipülasyon kriterlerini sınırlı biçimde saymak mümkün değildir. Dolayısıyla, manipülasyonu tespit etmek üzere tasarlanan bir sistemde;

1. Geçmiş manipülasyon örneklerini veya daha geniş anlamda manipülasyon yöntemlerini ve manipülatif davranış kalıplarını modellemek,
2. Manipülasyonun tanımından giderek, fiyatlardaki yapay düşürülme, yükseltilme veya sabit tutmayı fiyat değişimleri içerisinden yakalamak,

esas olmak üzere, iki temel yaklaşım izlemek mümkündür. Geçmiş manipülasyon örneklerine veya manipülasyon kalıplarına dayalı olarak hazırlanan bir veya birden fazla

İstatistiksel model uydurarak sistemin tasarlanması halinde, geçmişte rastlanmamış bir manipülatif davranış karşısında, manipülasyon tespitinde sistem başarısız olabileceği gibi işlem verileri üzerinde birden fazla istatistiksel modelin taranması süreci de zaman bakımında oldukça maliyetli olacaktır. Bu noktadan hareketle, model uydurma yerine manipülasyonun tanımından gidilerek, fiyat ve/veya miktar değişimindeki anomalileri belirlemek sistem tasarımında esas alınmıştır.

Hisse Senetleri Piyasası (HSP) işlem verileri, borsa endeksleri, mali tablo verileri ve makroekonomik göstergeler potansiyel girdilerdir. Buna karşın, manipülasyon sürecinin temel hareket noktası yani bir anlamda özü HSP işlemleri olduğu için ve Sistem’de hız önemli bir faktör olduğundan, Sistem HSP işlem verileri üzerine yapılandırılmıştır. Dolayısıyla, bir anlamda fayda-maliyet ödünleşimi sonucunda yavaş çalışan detaylı bir sistem yerine, hızlı çalışan ve amaca hizmet eden temel bir sistem esas alınmıştır.

Erken Uyarı Sistemi’nin tasarımında İstatistiksel Öğrenme bakış açısı esas alınarak, fiyat değişim oranlarındaki anomalileri tespit için önceki basamağın çıktısı, sonraki basamağın girdisi olacak biçimde; işlem yapılan hisse senedi, işleme aracılık eden aracı kuruluş, talimatı veren yatırımcı sırasını izleyen üç basamaklı ardışık bir süreç tasarlanmıştır. Sürecin başlangıcında tüm işlemler analize tabi tutularak, Sistemin her basamağında, şüphe taşımayan işlemler ayıklanacak biçimde bir işleyiş esas alınmıştır. Sistemin her basamağında elemeye gidilmesi, Sistem’e hız kazandırmanın yanı sıra şüphe taşımayan işlemleri gerçekleştiren yatırımcıları haksız yere zan altında bırakabilecek bir sonuca ulaşmamaktır. Şüphesiz ki manipülasyon, HSP’ye zarar veren ve yatırımcıları piyasadan uzaklaştırıp, piyasanın gelişimine engel olmakta ve sığ piyasaya yol açmaktadır. Ancak, manipülasyonun yanlış tespiti sonucu haksız yere zan altında kalan kişilerin sözkonusu olması halinde, bu da piyasaya katılım konusunda potansiyel yatırımcıları ürkütürerek gene piyasanın gelişimine darbe vurulmuş olacaktır. Bu nedenle, Sistem’in işleyişinde manipülasyon kararı alınıp, derinlemesine bir araştırmaya geçilmezden önce manipülatif davranışın, manipülatif olmayandan ayrıştırılabilmesi için azami gayret sarfedilmiştir. Bu nedenle, manipülatif işlem tespiti, hisse senedi, aracı kuruluş, yatırımcı akışıyla takip edilerek; işlem tüm bileşenleriyle belirlendikten sonra manipülasyon yapıldığı uyarısında bulunmaktadır.

Kavramsal anlamda manipülasyonla karıştırılan bir olgu olan spekülasyonu, manipülasyonla ayırtmak da oldukça büyük bir önem taşımaktadır. Büyük miktarda sermayeye sahip yatırımcıların sürü psikolojisinden faydalanarak kazanç sağlaması olarak tanımlanabilecek spekülasyon, manipülasyon gibi fiyat mekanizmasını hedef alan bir davranış değildir. Dolayısıyla, manipülasyon bir suç unsuru iken spekülasyon bir suç unsuru değildir. Bu nedenle, manipülasyon ile spekülasyonun ayırtılması da oldukça büyük bir önem taşımaktadır ki, analitik anlamda spekülasyonu manipülasyondan ayırtıracak temel bulgu fiyatlarla oynamak üzerine bir kurgu ve organizasyonun bulunmamasıdır. Bu nedenle, hisse senedi fiyat değişiminde diğerlerinden ayrışacak biçimde bir örüntünün bulunması ve aracı kuruluş(lar)da ve belli yatırımcılarda da belli yoğunlaşma spekülasyon içerisinde değerlendirilebilirken, belli hesaplar arasında işlem yoğunluğu spekülasyon için kabul edilebilirlik sınırlarını aşarak; manipülasyon tanımına girmektedir. Bu nedenle, sistemin ilk iki basamağında manipülatif işlem şüphesi taşıyan bazı davranışlar gözlense de, üçüncü basamakta belirli hesaplar da yoğunlaşma olmaması halinde sözkonusu işlemler spekülatif işlemler olarak değerlendirilmektedir. Çünkü, sürü psikolojisinden yararlanması beklenen spekülatif işlem gerçekleştiren yatırımcının büyük miktarda alım ve satımlarla, çok sayıda farklı hesabı kuyruğuna takarak kazanç sağlaması beklenir. Dolayısıyla, çok sayıda yatırımcı yerine belirli yatırımcılarda yoğunlaşma spekülasyonun doğasına aykırı olmakta ve Sistemde üçüncü basamak işlem etüdü spekülatif ve manipülatif işlemi birbirinden ayırmakta ve manipülatif işlem gerçekleştiğini kesinleştirmektedir.

Erken uyarı sistemi tasarlanırken veri madenciliğine istatistikî bakış açısı olan İstatistiksel Öğrenme baz alınmıştır. Sistem üç basamaklı bir süzgeç olarak ardışık işlemlerle manipülasyon tespit etmeye yönelik olarak tasarlanmış ve bir günlük zaman diliminde fiyat manipülasyonu yapıldığı bilinen bir hisse senedini, tüm gerçekleşen işlemler bakımından, manipülatif işlemleri ve manipülasyon gerçekleşen hisse senedini, işlemlere aracılık eden aracı kuruluşlar ve manipülatif işlemi gerçekleştiren yatırımcılar bakımından başarıyla tespit edip etmediği sorgulanmıştır. Sistemin 1. basamağı manipülatif işleme sahip hisse senedi olup olmadığını fiyat değişim oranından tespite yönelmiş olup; 2 etaptan oluşmaktadır. Birinci basamak 1 etap işlemi olan fiyat değişiminde diğerlerinden ayrışan örüntü tespiti için, değişim miktarının görelî olarak;

- Yüksek miktarda deęişim
- Orta düzeyde deęişim
- Düşük miktarda deęişim

olmak üzere üç grupta toplanması esas alınmıştır. Bu gruplar oluşturulurken manipülasyonun tanımından gidilerek, kazanç sağlamak amacıyla manipülasyonda;

- Fiyatların kasıtlı olarak yükseltilmesini temsilen yüksek miktarda deęişim,
- Fiyatların sabit tutulmasını, orta düzeyde deęişim,
- Fiyatların düşürülmesini düşük miktarda deęişim,

temsili esas alınmıştır. Mantık olarak, diğerlerinden ayrışacak kadar az gözleme sahip grup veya grupların varlığı manipülasyon açısından şüpheli işlemlere işaret etmektedir. Sistemin tasarımında izlenecek yaklaşım, model uydurma olmayacağı için model uydurmayı gerektirmeyecek İstatistiksel Öğrenme tekniklerinden faydalanılması esas alınmaktadır. Bu noktadan hareketle, K-ortalamlar Kümeleme Analizi, K-en yakın Komşuluk, Hiyerarşik Kümeleme Analizi ve Kendi Kendini Düzenleyen Haritalar alternatifleri oluşturmuştur. Alternatifler arasından denetimli bir İstatistiksel Öğrenme yöntemi olan K-ortalamlar Kümeleme Analizi, küme içinde homojen, kümeler arası heterojen gruplar yaratması ve küme sayısının önceden belirlenebilmesine imkan tanınması nedeniyle tercih edilmiştir. K-ortalamlar Kümeleme Analizi, 100.319 kayıttan oluşan tüm fiyat deęişimlerini 3 gruba ayırıştırmak üzere fiyat deęişim oranlarına uygulanmış, yüksek ve düşük düzeyde deęişim oranına sahip toplam 218 gözlem şüpheli olarak dikkate alınmıştır. Bu deęerlendirmede küme merkezlerinin birbirine yakın deęerler alması halinde manipülatif açıdan şüpheli işlem olmadığı düşünülüp, Sistem 1. Basamak 1. Etap işlem etüdünde süreci sona erdirecekti. Tespit edilen 218 şüpheli işlem sarı alarm halinde 2. etap işlem etüdüne alınarak; yüksek ve düşük düzeyde fiyat deęişimi kümelerine ait oldukları düzeyde manipülatif işlemi temsil kabiliyetine sahip oldukları düşüncesinden hareketle Bulanık Kümeleme algoritmalarıyla iki gruba ayrılmaları esas alınmıştır. Kesin c – Ortalamalar, Bulanık Kümeleme Algoritması ve FANNY Bulanık Kümeleme alternatiflerini oluşturmaktadır. Alternatifler arasında en dirençli olan ve bir tek kümeye aidiyeti deęil bütün kümelere ayrı ayrı üyelik deęerleriyle ifade eden FANNY Bulanık Kümeleme Algoritması tercih edilmiştir. Tespit edilen 218 şüpheli işlem FANNY Bulanık Kümeleme Algoritması ile iki gruba ayırmak üzere işleme tabi tutulmuştur. İşlem sonunda her bir gözlem için elde

edilen iki üyelik derecesinden büyük olanı, ait olduğu kümeyi daha iyi tespit edeceği düşüncesiyle tercih edilmiş ve bu değer fiyat değişim oranını ağırlıklandırmak üzere kullanılarak; fiyat değişim oranları bulanıklaştırılmıştır. Bulanıklaştırılmış fiyat değişim oranlarının tespit edilmesiyle 1 Basamak işlem etüdü sonlandırılmış ve 2. Basamak işlem etüdüne geçilmiştir.

İkinci basamak işlem etüdünde amaç manipülatif işlem şüphesini taşıyan işlemlerin gerçekleşmesine aracılık eden aracı kuruluşların varlığını sorgulamaktır. Sorgulama, 218 işlemi gerçekleştiren aracı kuruluşlar arasında yoğunlaşma olup olmadığını temel almaktadır. Dolayısıyla, şüpheli işlemlerle eşleşen aracı kuruluşları tespit etmeyi sağlayacak birliktelikler belirlenecektir. Birliktelik kuralları algoritmaları ve karar ağaçları algoritmaları yöntem alternatiflerini oluşturmaktadır. Hızlı ve kolay karar verebilmek için yoğunlaşma olup olmadığını görsel destekli olarak gözlemleme imkanı tanıyacak karar ağaçları algoritmaları uygun yöntem olarak belirlenmiş ve tüm alt grupları açığa çıkarmak için ikiden fazla gruba ayırıştırarak dallanma özelliğiyle diğer karar ağaçları algoritmalarından ayrılan CHAID algoritması tercih edilmiştir. Bulanıklaştırılmış fiyat değişim oranları hedef değer ve alım ve satım işlemine aracılık eden aracı kuruluşlar ile alım ve satım işlemi gerçekleştiren yatırımcılar tahmin edici değişken olmak üzere CHAID yöntemi uygulanmıştır. Dolayısıyla, CHAID algoritması bulanıklaştırılmış hedef değer için uygulanmış olup; CHAID algoritması değiştirilerek Bulanık Hedefli CHAID algoritmasına dönüştürülmüştür. Bulanık Hedefli CHAID algoritmasının uygulanmasıyla, 3 aracı kuruluşta yoğunlaşma tespit edilmiş ve 218 şüpheli işlemin bir yatırımcı tarafından sözkonusu 3 aracı kuruluş vasıtasıyla bu işlemleri gerçekleştirdiği gözlenmiştir. Böylece, 2. basamak işlem etüdü de tamamlanmış olup; turuncu alarm halinde manipülasyonun son ayağı olan yatırımcılar arası ilişkilerin sorgulandığı 3. basamağa geçilmiştir. Ayrıca, 3. basamak manipülasyon ve spekülasyon ayrımının da tespit edileceği basamaktır. 3. basamakta, yatırımcılar arası ilişkiler bakımından bir yoğunlaşma gözlenmemesi halinde şüpheli işlemlerin manipülatif değil spekülatif olduğu kararına varılacaktır.

3. basamak işlem etüdünde yatırımcılar arası ilişkiler etüd edilecektir. Önsel birliktelik kuralları algoritması ve karar ağaçları algoritmaları İstatistiksel Öğrenme yöntem

alternatiflerini oluşturmaktadır. Yatırımcılar arası ilişkilerin tespitinde hız faktörü öncelik taşıdığı ve karar vermeyi gerektirecek bir seçim süreci sözkonusu olmadığından görsel desteğe ihtiyaç duyulmadığından önsel birliktelik kuralları algoritması tercih edilmiştir. Önsel birliktelik kurallarının alım ve satım yapan hesaplardaki ilişkiyi sorgulamak üzere uygulanması sonucu şüpheli 109 satım işlemi gerçekleştiren hesap numarası ile %85,3 oranında birliktelik sergileyen alım yapan bir yatırımcı tespit edilmiştir. Dolayısıyla, manipülasyonun son ayağını oluşturan yatırımcılar arası ilişkiler de tespit edilmiş olup; kırmızı alarm halinde manipülasyon gerçekleştiği uyarısıyla işlemler ileri incelemeleri gerçekleştirecek yetkilere sevk edilmiştir.

Tasarlanan Erken Uyarı Sistemi yeniden gözden geçirildiğinde; Erken Uyarı Sistemi tasarlandıktan sonra, İMKB HSP’de işlem manipülasyonu yapıldığı tespit edilmiş gerçek bir örneğin işlemleri üzerinde işletilmesi sonucunda; hisse senedi, üyeler ve yatırımcılar olmak üzere işlem manipülasyonunun tüm unsurlarıyla başarıyla tespit edilebildiği gözlenmiştir. Ayrıca, CHAID karar ağaçları algoritması tez kapsamında değiştirilerek; Bulanık Hedefli CHAID karar ağaçları algoritması geliştirilmiştir. Tespitin yapılması kadar, tespit süresinin de önemi dikkate alınarak tüm algoritmaların işlem zamanları toplamının 5,147 saniye olarak gerçekleşmesi sistemin başarıyla işlediğini göstermektedir.

8.2 Sonraki Çalışmalara Öneriler

Bu tez çalışmasında, Erken Uyarı Sistemi tasarlanırken istatistiksel bakış açısı kullanılarak İstatistiksel Öğrenme yöntemlerinden faydalanılmıştır. Üç basamaklı bir süzgeç yapısında tasarlanan erken uyarı sisteminin, her bir basamağındaki işlemler İstatistiksel Öğrenme yöntemlerinin kullanımıyla otomatize edilmiş olmasına karşın basamaklar arasındaki geçişlerde karar verme insan yargısına bırakılmıştır. Sonraki çalışmalarda, basamaklar arası geçişlerde karar mekanizması insan yargısından arındırılarak, tamamıyla otomatize edilebilir.

Erken Uyarı Sistemi'nin tasarımı HSP işlem verilerinin içerdiği değişkenler esas alınmıştır. Sistemin mümkün olduğunca hızlı sonuç vermesi esas alındığı için sadece HSP işlem verilerinin içerdiği değişkenlere yer verilmiştir. Şüphesiz ki, İMKB Borsa Endeksi, işlem gören şirketlerin mali tabloları ve HSP'yi etkileyen makro ekonomik göstergelerin de sistemde kullanılacak değişken kolleksiyonuna eklenmesi mümkündür. Ancak, bu tercih Erken Uyarı Sistemi'nin süreçlerini uzatacağı ve zaman maliyetini oldukça artıracacağı için sadece HSP işlem verilerine dayalı bir sistem tasarlanmıştır. Hız faktörünün önem taşımadığı bir erken uyarı sisteminin tasarlanmasının düşünülmesi halinde İMKB Borsa Endeksi, işlem gören şirketlerin mali tabloları ve HSP'yi etkileyen makro ekonomik göstergelerin de yer aldığı bir sistem tasarlanması mümkündür.

Sistemin uygulamasında manipülasyon yapıldığı bilinen bir hisse senedinin, manipülasyonun gerçekleştiği dönemdeki işlem verileri kullanılarak sistemin başarısı test edilmiş ve manipülatif davranış kalıplarının sorgulanması gerçek veriye erişimdeki zorluklardan dolayı gerçekleştirilememiştir. Ancak, bir başka çalışmada, manipülatif davranış kalıplarının sistem üzerinde sorgulanması benzetim yöntemlerinin kullanımıyla gerçekleştirilebilir.

Sistemin tasarımı geçmiş manipülasyon örneklerine veya manipülasyon kalıplarına dayanan İstatistiksel Tahmin Modelleri'nin kullanımı yerine, verideki (fiyat değişimindeki) anomalilerden manipülasyon tespiti esas alınmıştır. Manipülasyon kalıplarına dayalı olarak yapılandırılmış bir sistemin temel eksikliği olabilecek yeni manipülasyon yöntemlerinin sistemce tespit edilememesi sorunu da Makine Öğrenimi ve Yapay Zeka yöntemlerinden faydalanılarak yeni manipülasyon kalıplarını da bünyesine katabilecek Akıllı Erken Uyarı Sistemi'nin tasarımıyla bir başka çalışma da inceleme konusu yapılarak aşılabılır.

Sistem, mevcut koşullar temel alınarak tasarlanmış ve Sermaye Piyasası paydaşlarına ait bilgilerinin yer aldığı bir veritabanının halihazırda mevcut olmaması nedeniyle, Sisteme böyle bir veritabanı entegre edilmemiştir. Mevcut altyapı koşulları dikkate alınarak, örüntü tespiti sadece HSP işlem verilerine dayandırılarak gerçekleştirilmiştir.

Paydaşların yer aldığı kapsamlı bir veritabanının mevcudiyeti halinde, veritabanı sisteme entegre edilerek ilişki ağını daha detaylı ve sağlıklı olarak tespit edebilecek bir sistem tasarlanabilir ve ASAM benzeri bir özellik erken uyarı sistemine bir başka çalışma kapsamında eklenebilir.

Erken Uyarı Sistemi istatistiksel bakış açısından değerlendirildiği için İstatistiksel Öğrenme tekniklerinin uygulaması esas olmuş ve sistemde kullanılan algoritmaları sunacak bir tek yazılım bulmak mümkün olmadığı için sistem SPSS 12.0, S-Plus 6.2, SPSS Answer Tree ve TANAGRA yazılımlarının kullanımıyla işletilmiştir. Sisteme bilişim bakış açısıyla bakılması halinde sözkonusu algoritmaları bir tek platformda kullanıma sunabilecek görsel destekli bir bağlantı yazılımı veya kullanıcı arayüzü geliştirilmesi bir başka çalışma kapsamında değerlendirilebilir.

Ayrıca, Erken Uyarı Sistemi'nin bilişim bakış açısından değerlendirilerek bir yazılım olarak hazırlanması da bir başka çalışmaya konu edilebilir.

KAYNAKLAR

- Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. and Verkamo, A. I. 1996. Fast discovery of association rules. Advances in knowledge discovery and data mining, AAAI/MIT Press, Cambridge, MA, 307-328.
- Anonim. 1989. Sermaye Piyasası Özel İhtisas Komisyonu Raporu. Devlet Planlama Teşkilatı, 111, Ankara.
- Ataman, T. 1999. Gelişmiş piyasalarda uygulanan manipülasyon inceleme yöntemleri. Sermaye Piyasası Kurulu, 46, Ankara.
- Berson, A., Smith, S. and Thearling, K. 1999. Building data mining applications for CRM. McGraw Hill, 510, USA.
- Brachman, R. and Anand, T. 1996. The process of knowledge discovery in databases: a human centered approach. Advances in knowledge discovery and data mining, AAAI/MIT Press, Cambridge, MA, 37-58.
- Cabena, P., Hadjinian, P., Stadler, R., Verhees, J. and Zanasi, A. 1997. Discovering data mining: from concept to implementation. Prentice Hall PTR, Upper Saddle River, New Jersey, 195, USA.
- Chen, Z. 2001. Data mining and uncertain reasoning: an integrated approach. John Wiley & Sons, Inc., 370, Canada.
- CRISP-DM Special Interest Group. 2004. CRISP- DM (Cross Industry Standart Process for Data Mining) 1.0, Step-by-step data mining guide. [<http://www.crisp-dm.org/CRISPWP-0800.pdf>]
- Çilli, H. ve Temel, T. 1988. Türk bankacılık sistemi için bir erken uyarı modeli. T.C. Merkez Bankası, Araştırma, Planlama ve Eğitim Genel Müdürlüğü, APE tartışma tebliği no: 8814.
- Dombi, J. 1990. Membership function as an evaluation. Fuzzy Sets And Systems, 35; 1-21.
- Fayyad ,U., Piatetsky-Shapiro G. and Symth, P. 1996. From data mining to knowledge discovery in databases. AI Magazine, 17(3); 37-54 .
- Gaytan, A. and Johnson, A.J. 2002. A review of the literature on early warning systems for banking crises. Central Bank of Chile, working papers no: 183.

- Gerni, C., Emsen, S. ve Değer, M.K. 2005. Erken uyarı sistemleri yoluyla Türkiye'deki ekonomik krizlerin analizi. *Ekonometri ve İstatistik*, 2; 39-61.
- Giles, R. 1988. The concept of grade of membership. *Fuzzy Sets And Systems*, 25; 297-323.
- Gönül, A. ve Eroğlu Z. A. 1999. Etkin bankacılık denetiminde temel prensipler: Türkiye ve diğer ülke uygulamaları. Devlet Planlama Teşkilatı. 93. Ankara
- Güngör, A.İ. 2001. Sermaye piyasası kanununun 6. ve 7. bölümleri, içerden öğrenenlerin ticareti, manipülasyon. Sermaye Piyasası Kurulu, 91, Ankara.
- Han, J and Kamber, M. 2000. Data mining: concepts and techniques. Morgan Kaufmann Series in Data Management Systems, 550, USA.
- Hartigan, J.A. and Wong, M.A. 1979. A k-means clustering algorithm. *Applied Statistics*, 28; 100-108.
- Hastie, T., Tibshirani, R. and Friedman, J. 2001. The elements of statistical learning; data mining, inference and prediction. Springer Series in Statistics, 533, USA.
- İMKB (İstanbul Menkul Kıymetler Borsası). 2006. Web sitesi. www.imkb.gov.tr. Erişim Tarihi: 20.03.2006
- Karan, M.B. ve Karacabey, A.A. 2003. Türkiye'de sermaye piyasasının mali sistem içindeki yeri, sorunları ve geleceği. Sermaye Piyasası Kurulu, Ankara.
- Kauffman, A. and Gupta, M.M. 1988. Fuzzy mathematical models in engineering and management science. Elsevier Science Publishers B.V., 338, North-Holland.
- Kaufmann, L. and Rousseeuw, P.J. 1990. Finding groups in data: an introduction to cluster analysis. John Wiley and Sons Inc, 368, USA
- KDNUGGETS. 2005. Web sitesi. www.kdnuggets.com. Erişim Tarihi: 19.09.2005
- Klir, G.J. and Yuan, B. 1995. Fuzzy sets and fuzzy logic theory and applications. Prentice Hall, 574, USA.
- Kovalerchuk, B. and Vityaev E. 2002. Data mining in finance: advances in relational and hybrid methods. Kluwer Academic Publishers, 308, USA.
- Koyuncugil, A.S. 1999. Stokastik hedef programlamaya bulanık algoritma yaklaşımı ve yatırım problemine uygulanması. Yüksek Lisans Tezi (basılmamış). Ankara Üniversitesi, Ankara.
- Lai, Y. and Hwang, C. 1992. Fuzzy mathematical programming. Springer-Verlag, 301, Germany.

- Moss, L.T. and Atre, S. 2003. Business intelligence roadmap: the complete project lifecycle for decision-support applications. Addison-Wesley Publishing, 576, USA.
- Negoita, M.G. 2000. Data mining for knowledge discovery (Book review). SIGKDD Explorations, 1 (2); 118-119.
- NYSE (New York Stock Exchange). 2005. Web sitesi. www.nyse.com. Erişim Tarihi: 15.10.2005
- Özbay, R. 1990. Hisse senedi fiyatlarında yükseliş ve çöküşler: borsalarda spekülasyon ve manipülasyon. Sermaye Piyasası Kurulu, 94, Ankara.
- Piatetsky-Shapiro, G. 1991. Knowledge discovery in real databases: a workshop report. AI Magazine, 11(5); 68-70.
- Rao, C.R. 2001. Statistics: reflections on the past and visions for the future. Communication in Statistics-Theory and Methods, 30(11); 2235-2257.
- Reinhold, D. and Wolfgang, G. 2000. Classification and information processing at the turn of the millennium. Springer, 492, Germany.
- Ribic, C.A. and Miller, T.W. 1998. Evaluation of alternative model selection criteria in the analysis of unimodal response curves using CART. Journal of Applied Statistics; 25(5), 685-698.
- Roiger, R. J. and Geatz, M. 2002. Data mining: a tutorial based primer. Addison-Wesley Publishing, 408, USA.
- SAS. 2005. Web sitesi. www.sas.com. Erişim Tarihi: 18.06.2005
- SET. 2006. Web sitesi. http://www.set.or.th/en/operation/supervision/surveillance_p1.html. Erişim tarihi: 18.09.2006
- SPK (Sermaye Piyasası Kurulu). 2006. Web sitesi. www.spk.gov.tr. Erişim Tarihi: 26.09.2005
- SPLUS. 2001. SPLUS 6 for windows guide to statistics, volume II. Insightful Corporation, 622, USA.
- SPSS. 2002. Answer Tree 3.1 user's guide. SPSS Inc., 240, USA.
- Terano, T., Asai, K. and Sugeno, M. 1991. Fuzzy systems theory and its applications, Academic Press Inc., 268, San Diego.

- Tezcanlı, M.V. 1993. Insider trading and market manipulation in stock markets: an empirical investigation and development of a surveillance model. Ph.D. thesis (unpublished). Marmara Üniversitesi, İstanbul.
- Thearling, K. 2005. Web sitesi: www.thearling.com. Erişim Tarihi: 18.06.2005
- Theis, W. and Weihs, C. 1999. Clustering techniques for the detection of business cycles, Technical Report 40/1999, SFB 475, Universitat Dortmund, Germany
- Tuncel, Ö. 1996. Bulanık doğrusal programlama. Yüksek Lisans Tezi (basılmamış). Hacettepe Üniversitesi, Ankara.
- TWO CROWS. 1999. Introduction to data mining and knowledge discovery. Two Crows Corporation, 36, USA
- Wang, L. X. 1997. A course in fuzzy systems and control. Prentice Hall PTR, Upper Saddle River, NJ, 424, USA.
- Witten, I.H. and Frank, E. 2000. Data mining. Morgan Kaufmann Publishers, 371, San Francisco, California.
- Ye, N. 2003. The handbook of data mining. Lawrence Erlbaum Associates, Publishers, 689, London.
- Zadeh, L. 1965. Fuzzy sets. Information and control, 8; 338-353.
- Zaine, O.R. 1999. Principles of KDD. Ph. D. Thesis (Unpublished). University of Alberta, Department of Computing Sciences.

EK 1 Şüpheli 218 gözlemin fiyat değişim oranı (fdo_i), üyelik dereceleri ($\mu_{1,i}$ ve $\mu_{2,i}$), en

büyük aidiyet derecesi (w_i) ve bulanık fiyat değişim oranı ($bfdo_i$) çizelgesi

i	fdo_i	$\mu_{1,i}$	$\mu_{2,i}$	w_i	$bfdo_i$
1	33,3333	0,9956	0,0044	0,9956	33,1871
2	33,3333	0,9956	0,0044	0,9956	33,1871
3	33,3333	0,9956	0,0044	0,9956	33,1871
4	33,3333	0,9956	0,0044	0,9956	33,1871
5	33,3333	0,9956	0,0044	0,9956	33,1871
6	33,3333	0,9956	0,0044	0,9956	33,1871
7	33,3333	0,9956	0,0044	0,9956	33,1871
8	33,3333	0,9956	0,0044	0,9956	33,1871
9	33,3333	0,9956	0,0044	0,9956	33,1871
10	33,3333	0,9956	0,0044	0,9956	33,1871
11	33,3333	0,9956	0,0044	0,9956	33,1871
12	33,3333	0,9956	0,0044	0,9956	33,1871
13	33,3333	0,9956	0,0044	0,9956	33,1871
14	33,3333	0,9956	0,0044	0,9956	33,1871
15	33,3333	0,9956	0,0044	0,9956	33,1871
16	33,3333	0,9956	0,0044	0,9956	33,1871
17	33,3333	0,9956	0,0044	0,9956	33,1871
18	33,3333	0,9956	0,0044	0,9956	33,1871
19	33,3333	0,9956	0,0044	0,9956	33,1871
20	33,3333	0,9956	0,0044	0,9956	33,1871
21	33,3333	0,9956	0,0044	0,9956	33,1871
22	33,3333	0,9956	0,0044	0,9956	33,1871
23	33,3333	0,9956	0,0044	0,9956	33,1871
24	33,3333	0,9956	0,0044	0,9956	33,1871
25	33,3333	0,9956	0,0044	0,9956	33,1871
26	33,3333	0,9956	0,0044	0,9956	33,1871
27	33,3333	0,9956	0,0044	0,9956	33,1871
28	33,3333	0,9956	0,0044	0,9956	33,1871
29	33,3333	0,9956	0,0044	0,9956	33,1871
30	33,3333	0,9956	0,0044	0,9956	33,1871
31	33,3333	0,9956	0,0044	0,9956	33,1871
32	33,3333	0,9956	0,0044	0,9956	33,1871
33	33,3333	0,9956	0,0044	0,9956	33,1871
34	33,3333	0,9956	0,0044	0,9956	33,1871
35	33,3333	0,9956	0,0044	0,9956	33,1871
36	33,3333	0,9956	0,0044	0,9956	33,1871
37	33,3333	0,9956	0,0044	0,9956	33,1871
38	33,3333	0,9956	0,0044	0,9956	33,1871
39	33,3333	0,9956	0,0044	0,9956	33,1871
40	33,3333	0,9956	0,0044	0,9956	33,1871

EK 1 (devam)

i	fdo_i	$\mu_{1,i}$	$\mu_{2,i}$	w_i	$bfdo_i$
41	33,3333	0,9956	0,0044	0,9956	33,1871
42	33,3333	0,9956	0,0044	0,9956	33,1871
43	33,3333	0,9956	0,0044	0,9956	33,1871
44	33,3333	0,9956	0,0044	0,9956	33,1871
45	33,3333	0,9956	0,0044	0,9956	33,1871
46	33,3333	0,9956	0,0044	0,9956	33,1871
47	33,3333	0,9956	0,0044	0,9956	33,1871
48	33,3333	0,9956	0,0044	0,9956	33,1871
49	33,3333	0,9956	0,0044	0,9956	33,1871
50	33,3333	0,9956	0,0044	0,9956	33,1871
51	33,3333	0,9956	0,0044	0,9956	33,1871
52	33,3333	0,9956	0,0044	0,9956	33,1871
53	33,3333	0,9956	0,0044	0,9956	33,1871
54	33,3333	0,9956	0,0044	0,9956	33,1871
55	33,3333	0,9956	0,0044	0,9956	33,1871
56	31,5789	0,9866	0,0134	0,9866	31,1555
57	31,5789	0,9866	0,0134	0,9866	31,1555
58	31,5789	0,9866	0,0134	0,9866	31,1555
59	31,5789	0,9866	0,0134	0,9866	31,1555
60	31,5789	0,9866	0,0134	0,9866	31,1555
61	31,5789	0,9866	0,0134	0,9866	31,1555
62	31,5789	0,9866	0,0134	0,9866	31,1555
63	31,5789	0,9866	0,0134	0,9866	31,1555
64	31,5789	0,9866	0,0134	0,9866	31,1555
65	31,5789	0,9866	0,0134	0,9866	31,1555
66	31,5789	0,9866	0,0134	0,9866	31,1555
67	31,5789	0,9866	0,0134	0,9866	31,1555
68	31,5789	0,9866	0,0134	0,9866	31,1555
69	31,5789	0,9866	0,0134	0,9866	31,1555
70	31,5789	0,9866	0,0134	0,9866	31,1555
71	31,5789	0,9866	0,0134	0,9866	31,1555
72	31,5789	0,9866	0,0134	0,9866	31,1555
73	31,5789	0,9866	0,0134	0,9866	31,1555
74	31,5789	0,9866	0,0134	0,9866	31,1555
75	31,5789	0,9866	0,0134	0,9866	31,1555
76	31,5789	0,9866	0,0134	0,9866	31,1555
77	31,5789	0,9866	0,0134	0,9866	31,1555
78	31,5789	0,9866	0,0134	0,9866	31,1555
79	31,5789	0,9866	0,0134	0,9866	31,1555
80	31,5789	0,9866	0,0134	0,9866	31,1555
81	31,5789	0,9866	0,0134	0,9866	31,1555
82	31,5789	0,9866	0,0134	0,9866	31,1555
83	31,5789	0,9866	0,0134	0,9866	31,1555

EK 1 (devam)

i	fdo_i	$\mu_{1,i}$	$\mu_{2,i}$	w_i	$bfdo_i$
84	31,5789	0,9866	0,0134	0,9866	31,1555
85	31,5789	0,9866	0,0134	0,9866	31,1555
86	31,5789	0,9866	0,0134	0,9866	31,1555
87	31,5789	0,9866	0,0134	0,9866	31,1555
88	31,5789	0,9866	0,0134	0,9866	31,1555
89	31,5789	0,9866	0,0134	0,9866	31,1555
90	31,5789	0,9866	0,0134	0,9866	31,1555
91	31,5789	0,9866	0,0134	0,9866	31,1555
92	31,5789	0,9866	0,0134	0,9866	31,1555
93	31,5789	0,9866	0,0134	0,9866	31,1555
94	31,5789	0,9866	0,0134	0,9866	31,1555
95	35,1351	0,9742	0,0258	0,9742	34,2288
96	35,1351	0,9742	0,0258	0,9742	34,2288
97	35,1351	0,9742	0,0258	0,9742	34,2288
98	35,1351	0,9742	0,0258	0,9742	34,2288
99	35,1351	0,9742	0,0258	0,9742	34,2288
100	35,1351	0,9742	0,0258	0,9742	34,2288
101	35,1351	0,9742	0,0258	0,9742	34,2288
102	35,1351	0,9742	0,0258	0,9742	34,2288
103	35,1351	0,9742	0,0258	0,9742	34,2288
104	35,1351	0,9742	0,0258	0,9742	34,2288
105	35,1351	0,9742	0,0258	0,9742	34,2288
106	35,1351	0,9742	0,0258	0,9742	34,2288
107	35,1351	0,9742	0,0258	0,9742	34,2288
108	35,1351	0,9742	0,0258	0,9742	34,2288
109	35,1351	0,9742	0,0258	0,9742	34,2288
110	-23,0000	0,0242	0,9758	0,9758	-22,4432
111	-26,0000	0,0147	0,9853	0,9853	-25,6170
112	-26,0000	0,0147	0,9853	0,9853	-25,6170
113	-26,0000	0,0147	0,9853	0,9853	-25,6170
114	-26,0000	0,0147	0,9853	0,9853	-25,6170
115	-26,0000	0,0147	0,9853	0,9853	-25,6170
116	-26,0000	0,0147	0,9853	0,9853	-25,6170
117	-26,0000	0,0147	0,9853	0,9853	-25,6170
118	-26,0000	0,0147	0,9853	0,9853	-25,6170
119	-26,0000	0,0147	0,9853	0,9853	-25,6170
120	-26,0000	0,0147	0,9853	0,9853	-25,6170
121	-26,0000	0,0147	0,9853	0,9853	-25,6170
122	-26,0000	0,0147	0,9853	0,9853	-25,6170
123	-26,0000	0,0147	0,9853	0,9853	-25,6170
124	-26,0000	0,0147	0,9853	0,9853	-25,6170
125	-26,0000	0,0147	0,9853	0,9853	-25,6170
126	-26,0000	0,0147	0,9853	0,9853	-25,6170

EK 1 (devam)

i	fdo_i	$\mu_{1,i}$	$\mu_{2,i}$	w_i	bfd_o_i
127	-26,0000	0,0147	0,9853	0,9853	-25,6170
128	-24,0000	0,0069	0,9931	0,9931	-23,8346
129	-24,0000	0,0069	0,9931	0,9931	-23,8346
130	-24,0000	0,0069	0,9931	0,9931	-23,8346
131	-24,0000	0,0069	0,9931	0,9931	-23,8346
132	-24,0000	0,0069	0,9931	0,9931	-23,8346
133	-24,0000	0,0069	0,9931	0,9931	-23,8346
134	-24,0000	0,0069	0,9931	0,9931	-23,8346
135	-24,0000	0,0069	0,9931	0,9931	-23,8346
136	-24,0000	0,0069	0,9931	0,9931	-23,8346
137	-24,0000	0,0069	0,9931	0,9931	-23,8346
138	-24,0000	0,0069	0,9931	0,9931	-23,8346
139	-24,0000	0,0069	0,9931	0,9931	-23,8346
140	-24,0000	0,0069	0,9931	0,9931	-23,8346
141	-24,0000	0,0069	0,9931	0,9931	-23,8346
142	-24,0000	0,0069	0,9931	0,9931	-23,8346
143	-24,0000	0,0069	0,9931	0,9931	-23,8346
144	-24,0000	0,0069	0,9931	0,9931	-23,8346
145	-24,0000	0,0069	0,9931	0,9931	-23,8346
146	-24,0000	0,0069	0,9931	0,9931	-23,8346
147	-24,0000	0,0069	0,9931	0,9931	-23,8346
148	-24,0000	0,0069	0,9931	0,9931	-23,8346
149	-24,0000	0,0069	0,9931	0,9931	-23,8346
150	-24,0000	0,0069	0,9931	0,9931	-23,8346
151	-24,0000	0,0069	0,9931	0,9931	-23,8346
152	-24,0000	0,0069	0,9931	0,9931	-23,8346
153	-24,0000	0,0069	0,9931	0,9931	-23,8346
154	-24,0000	0,0069	0,9931	0,9931	-23,8346
155	-24,0000	0,0069	0,9931	0,9931	-23,8346
156	-24,0000	0,0069	0,9931	0,9931	-23,8346
157	-24,0000	0,0069	0,9931	0,9931	-23,8346
158	-24,0000	0,0069	0,9931	0,9931	-23,8346
159	-24,0000	0,0069	0,9931	0,9931	-23,8346
160	-24,0000	0,0069	0,9931	0,9931	-23,8346
161	-24,0000	0,0069	0,9931	0,9931	-23,8346
162	-24,0000	0,0069	0,9931	0,9931	-23,8346
163	-24,0000	0,0069	0,9931	0,9931	-23,8346
164	-24,0000	0,0069	0,9931	0,9931	-23,8346
165	-24,0000	0,0069	0,9931	0,9931	-23,8346
166	-24,0000	0,0069	0,9931	0,9931	-23,8346
167	-24,0000	0,0069	0,9931	0,9931	-23,8346
168	-24,0000	0,0069	0,9931	0,9931	-23,8346
169	-24,0000	0,0069	0,9931	0,9931	-23,8346

EK 1 (devam)

i	fdo_i	$\mu_{1,i}$	$\mu_{2,i}$	w_i	$bfdo_i$
170	-25,0000	0,0031	0,9969	0,9969	-24,9223
171	-25,0000	0,0031	0,9969	0,9969	-24,9223
172	-25,0000	0,0031	0,9969	0,9969	-24,9223
173	-25,0000	0,0031	0,9969	0,9969	-24,9223
174	-25,0000	0,0031	0,9969	0,9969	-24,9223
175	-25,0000	0,0031	0,9969	0,9969	-24,9223
176	-25,0000	0,0031	0,9969	0,9969	-24,9223
177	-25,0000	0,0031	0,9969	0,9969	-24,9223
178	-25,0000	0,0031	0,9969	0,9969	-24,9223
179	-25,0000	0,0031	0,9969	0,9969	-24,9223
180	-25,0000	0,0031	0,9969	0,9969	-24,9223
181	-25,0000	0,0031	0,9969	0,9969	-24,9223
182	-25,0000	0,0031	0,9969	0,9969	-24,9223
183	-25,0000	0,0031	0,9969	0,9969	-24,9223
184	-25,0000	0,0031	0,9969	0,9969	-24,9223
185	-25,0000	0,0031	0,9969	0,9969	-24,9223
186	-25,0000	0,0031	0,9969	0,9969	-24,9223
187	-25,0000	0,0031	0,9969	0,9969	-24,9223
188	-25,0000	0,0031	0,9969	0,9969	-24,9223
189	-25,0000	0,0031	0,9969	0,9969	-24,9223
190	-25,0000	0,0031	0,9969	0,9969	-24,9223
191	-25,0000	0,0031	0,9969	0,9969	-24,9223
192	-25,0000	0,0031	0,9969	0,9969	-24,9223
193	-25,0000	0,0031	0,9969	0,9969	-24,9223
194	-25,0000	0,0031	0,9969	0,9969	-24,9223
195	-25,0000	0,0031	0,9969	0,9969	-24,9223
196	-25,0000	0,0031	0,9969	0,9969	-24,9223
197	-25,0000	0,0031	0,9969	0,9969	-24,9223
198	-25,0000	0,0031	0,9969	0,9969	-24,9223
199	-25,0000	0,0031	0,9969	0,9969	-24,9223
200	-25,0000	0,0031	0,9969	0,9969	-24,9223
201	-25,0000	0,0031	0,9969	0,9969	-24,9223
202	-25,0000	0,0031	0,9969	0,9969	-24,9223
203	-25,0000	0,0031	0,9969	0,9969	-24,9223
204	-25,0000	0,0031	0,9969	0,9969	-24,9223
205	-25,0000	0,0031	0,9969	0,9969	-24,9223
206	-25,0000	0,0031	0,9969	0,9969	-24,9223
207	-25,0000	0,0031	0,9969	0,9969	-24,9223
208	-25,0000	0,0031	0,9969	0,9969	-24,9223
209	-25,0000	0,0031	0,9969	0,9969	-24,9223
210	-25,0000	0,0031	0,9969	0,9969	-24,9223
211	-25,0000	0,0031	0,9969	0,9969	-24,9223
212	-25,0000	0,0031	0,9969	0,9969	-24,9223

EK 1 (devam)

i	fdo_i	$\mu_{1,i}$	$\mu_{2,i}$	w_i	$bfdo_i$
213	-25,0000	0,0031	0,9969	0,9969	-24,9223
214	-25,0000	0,0031	0,9969	0,9969	-24,9223
215	-25,0000	0,0031	0,9969	0,9969	-24,9223
216	-25,0000	0,0031	0,9969	0,9969	-24,9223
217	-25,0000	0,0031	0,9969	0,9969	-24,9223
218	-25,0000	0,0031	0,9969	0,9969	-24,9223

ÖZGEÇMİŞ

Adı Soyadı : Ali Serhan KOYUNCUGİL

Doğum Yeri : Gaziantep

Doğum Tarihi : 31.07.1973

Medeni Hali : Bekar

Yabancı Dili : İngilizce

Eğitim Durumu (Kurum ve Yıl)

Lise : Ankara Kocatepe Mimar Kemal Lisesi (1987-1990)

Lisans : Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü (1992-1996)

Yüksek Lisans: Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilimdalı (1996-1999)

Çalıştığı Kurum/Kurumlar ve Yıl

Devlet İstatistik Enstitüsü, 1998 – 2002

Sermaye Piyasası Kurulu, 2002 -

Yayımları (SCI ve diğer)

Yüksek Lisans Tezi

Stokastik Hedef Programlamaya Bulanık Algoritma Yaklaşımı ve Bir Yatırım Problemine Uygulanması

Hakemli Uluslararası Endeksli Dergi

Özgülbaş, N. and **Koyuncugil, A.S.** 2006. Using CHAID Decision Tree Algorithm for Financial Performance Measurement Scale. The Business Review, Cambridge, Vol:5, No:1

Derginin Tarandığı Endeksler: ABI/INFORM (ProQuest), EBSCO, CABELL'S and ULRICH'S Directories.

Uluslararası Konferans Bildirileri

Koyuncugil, A.S. and Özgülbaş, N. 2006. Financial Profiling of SME's: An Application by Data Mining. 26-29 June 2006, 2006 Applied Business Research Conference. Clute Institute for Academic Research, Florence, Italy.

Özgülbaş, N. and **Koyuncugil, A.S.** 2006. Determination of Success Limit of Capacity Utilization Rate and its Impact Factors via Data Mining. 2nd International Conference on Business, Management and Economics. University of Yasar, 15-18 June 2006, İzmir, Turkey.

Özgülbaş, N. and **Koyuncugil, A.S.** 2006. Application of Data Mining Method For Financial Profiling. 5th International Conference on Corporate Social Responsibility and Accounting Finance and Regulation, University of Trakya & London Metropolitan University, 1-4 May 2006, Edirne, Turkey.

Koyuncugil, A.S. and Apaydın, A. 2005. A Sampling Method Suggestion Based On Data Mining For Confidence Index Of Stock Exchange Market Investors. 55th Session of the ISI (International Statistical Institute), 10-14 April 2005, Sydney-Australia.

Apaydın, A. and **Koyuncugil, A.S.** 2001. Fuzzy Goal Programming Modelling Of Intelligent Character Recognition (ICR) System. 53rd Session of the ISI (International Statistical Institute) 22-29 August 2001, Seoul- Korea

Gebizlioğlu, Ö. L., Apaydın, A. and **Koyuncugil A.S.** 1999. Simulation Of A Macro Optimization Model For Census Of Population. 52nd Session of the ISI (International Statistical Institute), 10-18 August 1999, Helsinki-Finland.

Ulusal Dergi

Koyuncugil, A.S. 2004. Veri madenciliği veya bir başka deyişle akıllı algoritmalarla istatistik kullanımı. Emniyet Genel Müdürlüğü, Polis Dergisi, Bilişim Özel Sayısı, 37,

Koyuncugil, A.S. 2003. E-dönüşümde, doğru bilgi, şeffaflık ve demokrasiye erişimin yolu: İstatistik. Türkiye Bilişim Derneği, Bilişim Dergisi. Sayı: 86

Ulusal Konferans Bildirileri

Özgülbaş, N., Aksoy, B., Çolak, M., **Koyuncugil, A.S.** ve Ersoy, K. 2006. Hastanelerde kapasite ve finansal performans ilişkisi. IV. Ulusal Sağlık Kurumları ve Hastane Yönetimi Kongresi, 13-16 Nisan 2006, Fethiye, Türkiye.

Özgülbaş, N. ve **Koyuncugil, A.S.** 2005. Finansal Performans Düzeyinin Artırılmasında Bir Strateji Olarak Benchmarking Tekniğinin Uygulanması. 29-30 Eylül 2005, 9. Ulusal Finans Sempozyumu, Kapadokya, Nevşehir, TÜRKİYE.

Sözen, C., **Koyuncugil, A.S.** ve Özgülbaş, N. 2005. Bilgi Sistemleri Kurulumunun Maliyet-Etkililiği Üzerine Bir Çalışma: Hastane Otomasyon Yöntemlerinin Karşılaştırılması. Ulusal Hastane Yönetimi Kongresi, 29-30 Eylül 2005, Ankara, Türkiye.

Koyuncuđil, A.S. ve Apaydın, A. 2001. Büyük Ölçekli İstatistiki Çalışmalarda Kaynak ve Maliyet Planlaması ve Nüfus Sayımı Örneđi. 3. İstatistik Kongresi, 16-20 Nisan 2003, Antalya, Türkiye.

Apaydın, A. ve **Koyuncuđil, A.S.** 2001. Ekonomi Federasyonları ve Türkiye İin Modellemesi. 2. İstatistik Kongresi, 2-6 Mayıs 2001, Belek, Antalya, Türkiye.

Apaydın, A. ve **Koyuncuđil, A.S.** 1999. Nüfus Sayımı İin Bir Mikro Optimizasyon Modeli. XX. Ulusal Yöneylem ve Endüstri Mühendisliđi Kongresi, 8-9 Haziran 1999, Kara Harp Okulu, Ankara, Türkiye.

Apaydın, A. ve **Koyuncuđil, A.S.** 1999. Stokastik Hedef Programlamaya Bulanık Algoritma Yaklaşımıyla Bir Yatırım Probleminin Modellenmesi. 1. İstatistik Kongresi, 5-9 Mayıs 1999, Belek, Antalya, Türkiye.